CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 123, No. 2 February 2008

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

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 and notices are* 2 months *prior to publication.*

Announcement of the 2008 Election

In accordance with the provisions of the bylaws, the following Nominating Committee was appointed to prepare a slate for the election to take place on 19 June 2008:

William A. Yost, Chair

Michael Bailey

Charles Gaumond

Philip Marston

Michael Stinson

Maureen Stone

The bylaws of the Society require that the Executive Director publish in the *Journal*, at least 90 days prior to the election date, an announcement of the election and the Nominating Committee's nominations for the offices to be filled. Additional candidates for these offices may be provided by any Member or Fellow in good standing by letter received by the Executive Director not less than 60 days prior to the election date, and the name of any eligible candidate so proposed by 20 Members or Fellows shall be entered on the ballot. Biographical information about the candidates and statements of objectives of the candidates for President-Elect and Vice President-Elect will be mailed with the ballots.

CHARLES E. SCHMID
Executive Director

The Nominating Committee has submitted the following slate:

For President-Elect

Whitlow W. L. Au

David L. Bradley

For Vice President-Elect

James V. Candy

Diane Kewley-Port

 0001-4966/2008/123(2)/571/8/$23.00

Robin O. Cleveland

Peter H. Dahl

Brenda L. Lonsbury-Martin

James H. Miller

Clark S. Penrod

Andrea M. Simmons

Preliminary Notice: Acoustics'08 Paris

Acoustics'08 Paris is the Second Joint Meeting of the Acoustical Society of America and the European Acoustics Association encompassing the 155th meeting of the Acoustical Society of America, the 5th Forum Acusticum of the European Acoustics Association, the 9th Congress of the French Acoustical Society, and integrating Euronoise—the 7th European Conference on Noise Control, ECUA—the 9th European Conference on Underwater Acoustics, and the 60th Anniversary of the SFA.

The meeting will be held Sunday through Friday, 29 June to 4 July 2008 at the Palais des Congrès in Paris, France. Information about the meeting also appears on the meeting website at www.acoustics08-paris.org.

Technical Program

Seventeen regular sessions (sessions denoted by "00") and a large number (172) of structured (special) sessions will be organized jointly by ASA and EAA representatives. Forty-two structured sessions related to underwater acoustics (AO, AB, EA, SP, and UW) are organized jointly with ECUA and thirty-eight related to noise (NS and SA) are organized with Euronoise.

The technical program will consist of both lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 30 June–4 July. The conference language is English.

Structured (Special) Sessions

AA00—**Architectural Acoustics**—Regular (not structured)
AA01—Acoustics of opera houses–Joint with: MU, PA, NS
AA02—Archeological acoustics
AA03—Comparison of US and European standards in building/room acoustics–Joint with: ASACOS
AA04—Coupled volume acoustics
AA05—Acoustics and electroacoustics of small rooms–Joint with: EA
AA06—Low frequency absorption: Mechanisms, measurement methods and application–Joint with: NS
AA07—New measurement parameters in performing arts spaces–Joint with: MU
AA08—Student design competition
AA09—New frontiers in room acoustical modeling
AA10—Airborne and impact sound insulation–Joint with: NS
AA11—Acoustics of open-plan spaces–Joint with: NS
AA12—Speech segregation in rooms–Joint with: PP
AA13—Acoustics and privacy in healthcare facilities I: Emerging policy around the world–Joint with: NS
AA14—Acoustics and privacy in healthcare facilities II: Emerging research around the world–Joint with: NS
AA15—Surround sound acoustics–Joint with: MU
AA16—Prediction methods in building acoustics

AA17—Measuring methods and uncertainty in building acoustics
AA18—Acoustics of concert halls

AB00—**Animal Bioacoustics**—Regular (not structured)
Continued in ecua section, AB01 to AB07

BB00—**Biomedical Ultrasound/Bioresponse to Vibration**—Regular (not structured)
BB01—High-intensity focused ultrasound
BB02—High-intensity focused ultrasound metrology and standards–Joint with: EA
BB03—Shock waves in medicine–Joint with: PA
BB04—Theoretical and computational models of ultrasonic propagation in bones
BB05—Ultrasonic characterization of bone
BB06—Quantitative ultrasound methods for diagnosis and therapy
BB07—Light and sound for medical imaging and therapy–Joint with: PA
BB08—Transducers for medical imaging and therapy–Joint with: EA
BB09—Ultrasound contrast agents for imaging–Joint with: PA
BB10—Ultrasound contrast agents for therapy
BB11—Biomedical applications of acoustic radiation force

EA00—**Engineering Acoustics**—Regular (not structured)
EA01—Hearing aid engineering–Joint with: PP
EA02—Microphone array signal processing–Joint with: SP
EA03—Silicon microphones
EA04—Ultrasonic acoustic MEMS
EA05—Transducers and signal processing for the oil and gas industry–Joint with: SP
Continued in ecua section, EA06 to EA08

ED00—**Education in Acoustics**—Regular (not structured)
ED01—Interactive science and performance of jazz for high school students–Joint with: MU
ED02—Take 5's
ED03—Acoustics in the public school science classrooms
ED04—Acoustics education software

MU00—**Musical Acoustics**—Regular (not structured)
MU01—Acoustics and psychoacoustics of pipe organs–Joint with: PP
MU02—Brass instrument acoustics–Joint with: PA
MU03—Interdisciplinary research on the science of singing: A tribute to Johan Sundberg–Joint with: SC
MU04—Virtual musical instruments
MU05—Singing voice and source-filter interaction–Joint with: SC
MU06—Interaction between instrument and instrumentalist
MU07—Control of natural and synthetic musical sounds
MU08—Signal representations and models of musical sounds–Joint with: SP
MU09—Acoustic measurements on wind instruments
MU10—Edge tone and flue pipes

PA00—**Physical Acoustics**—Regular (not structured)
PA01—Acoustic landmine detection–Joint with: SP
PA02—Acoustic probes of planetary environments
PA03—Acoustically activated bubble dynamics and applications–Joint with: AO, BB
PA04—Ultrasonics under extreme conditions
PA05—Infrasound
PA06—Sonic, ultrasonic, and megasonic cleaning
PA07—Nonlinear acoustics of unconsolidated granular media
PA08—Nonlinear acoustics of consolidated materials and nondestructive testing
PA09—Time reversal acoustics for nonlinear imaging
PA10—Nonlinear acoustics in earthquake processes and other earth processes
PA11—Outdoor sound propagation and uncertainties
PA12—Phononic crystals
PA13—Photoacoustics
PA14—Acoustics of porous media
PA15—Thermoacoustics
PA16—Ultrafast acoustics
PA17—Diffraction of waves on periodical structures: Acoustic, ultrasonic and acousto-optical diffraction phenomena
PA18—Combustion noise and thermo-acoustics

PP00—**Psychological and Physiological Acoustics**—Regular (not structured)
PP01—Role of temporal fine structure in speech and nonspeech perception for normal & hearing-impaired people
PP02—Cross-spectral auditory integration: Physiological, psychophysical, and clinical evidence
PP03—Auditory perception and signal processing by prostheses
PP04—Cochlear implants: Going beyond the envelope
PP05—Applications of psychoacoustics–Joint with: ASACOS
PP06—Binaural perception by hearing-aid wearers
PP07—Acoustic features and speech perception–Joint with: SC
PP08—Integrated approaches to auditory scene analysis
PP09—Jens Blauert and his contributions–Joint with: AA, NS, SP
PP10—Auditory perception of sound source properties
PP11—Loudness, from controlled stimuli to environmental sounds
PP12—Otoacoustic emissions, from cochlear modeling to experimental techniques, and back
PP13—Computational auralization

SC00—**Speech Communication**—Regular (not structured)
SC01—Articulatory modeling and control of speech and singing organs
SC02—Multi-modal speech technology
SC03—How do physical and motor knowledge matter to speech perception?
SC04—Acoustics of speech production: Aeroacoustics and phonation
SC05—Speech recognition in noisy environments
SC06—Measurement of sociophonetic variation in speech
SC07—Neurobiology of speech perception
SC08—Speaker identification by machine
SC09—Cross-language speech perception and production
SC10—Speech prosody and how it relates to segmental aspects of speech

SP00—**Signal Processing in Acoustics**—Regular (not structured)
SP01—Overview of time reversal in acoustics–Joint with: PA, BB, UW
SP02—Time reversal methods for array imaging and signal processing–Joint with: BB, UW
SP03—Biomedical applications of time reversal–Joint with: PA

Continued in ecua section, SP04 to SP05

Joint with European Conference on Underwater Acoustics
UW00—Underwater Acoustics—Regular (not structured)
UW01—Impact of environmental variability on mid-frequency sonar performance–Joint with: AO
UW02—Impact of internal waves on shallow water propagation–Joint with: AO
UW03—Geoacoustic sediment modeling
UW04—Seabed and sea surface interaction measurements and modeling–Joint with: AO
UW05—Sound propagation in 3-dimensional environments
UW06—Acoustic vector fields and sensor processing–Joint with: SP
UW07—Broadband underwater communications–Joint with: SP
UW08—Acoustic data fusion
UW09—High frequency variability
UW10—Sonar system and transducer calibration methodology
UW11—Fifty years of progress in sonar acoustic research: The role of NURC/SACLANTCEN
UW12—Scattering from objects near boundaries
UW13—Noise suppression, robust direction of arrival, and target strength estimation
UW14—Determination of acoustic properties of materials for sonar applications
UW15—Tank experiments
UW16—Low-frequency and high-frequency synthetic aperture sonar
UW17—Numerical methods in underwater scattering
UW18—Sensor coalition
UW19—High-frequency scattering
UW20—Synthetic aperture sonar and radar convergences
UW21—Automatic target recognition, sensors and algorithms
UW22—Nonlinear acoustic methods in searching for buried objects

UW23—Auralization of sonar signals

AB00—**Animal Bioacoustics**—Regular (not structured)
AB01—Sound production and reception in amphibious marine mammals
AB02—Animal sonar systems–Joint with: UW, AO, SP
AB03—Auditory brainstem response and behavior correlation–Joint with: PP
AB04—Animal bioacoustic censusing
AB05—Anthropogenic noise effects on animals–Joint with: NS
AB06—Odontocete acoustics
AB07—Numerical modeling: Animal sound production and reception

AO00—**Acoustical Oceanography**—Regular (not structured)
AO01—Marine ecosystem acoustics
AO02—Acoustic characterization of sea floor habitats
AO03—Acoustical oceanography of polar environments
AO04—Geoacoustic characterization of the ocean bottom and geoacoustic inversion–Joint with: UW, SP
AO05—Rapid environmental assessment–Joint with: UW
AO06—Adjoint modeling for geoacoustic inversion–Joint with: SP
AO07—Passive acoustic tomography–Joint with: SP

EA00—**Engineering Acoustics**—Regular (not structured)
EA06—Sonar transducer design and modeling–Joint with: UW
EA07—Sensor technologies for autonomous acoustic sensing systems–Joint with: UW, SP
EA08—Acoustics in marine archeology–Joint with: UW, AO

SP00—**Signal Processing in Acoustics**—Regular (not structured)
SP04—Bayesian signal processing–Joint with: AO
SP05—Model-based signal processing–Joint with: AO

Joint with the European Conference on Noise Control

NS00—**Noise**—Regular (not structured)
NS01—Classroom acoustics–Joint with: ASACOS, AA
NS02—Comparing noise regulations and codes in USA and Europe–Joint with: ASACOS
NS03—Cultural variations in sound/noise assessment
NS04—Measurement of occupational noise exposure–Joint with: ASACOS
NS05—Noise from wind power projects
NS06—Soundscape & community noise
NS07—Aeroacoustics–Joint with: PA
NS08—Tire-road noise from the road perspective
NS09—Prominent discrete tones–Joint with: ASACOS
NS10—Sound quality tools and applications–Joint with: ASACOS
NS11—Airframe noise measurement, prediction, and control–Joint with: SA, SP
NS12—Sound and vibration from explosions in air–Joint with: SA, PA
NS13—Environmental noise mapping
NS14—Session in honor of Henning von Gierke–Joint with: BB, ASACOS
NS15—Noise, vibration and acoustics for medical and research facilities and their occupants–Joint with: AA
NS16—Sleep disturbances and other health effects–Joint with: ASACOS
NS17—Vibration perception
NS18—Physical and psychophysical evaluation of vehicle exterior noise
NS19—Fan noise and low-Mach number rotating blade noise
NS20—Car acoustics
NS21—Action planning and global solutions for urban noise
NS22—Acoustic performance of energy efficient building products
NS23—Soundscape in the heritage of urban and natural areas
NS24—Railway noise and vibration
NS25—Potential to reduce tire/road noise
NS26—Noise mapping techniques and uncertainties
NS27—Noise, structure borne noise from building technical equipment, and ground borne noise from railways
NS28—EU projects for aircraft noise reduction
NS29—Source identification in radiation and scattering
NS30—Time-domain modeling methods in acoustics

SA00—**Structural Acoustics and Vibration**—Regular (not structured)
SA01—Vibration and radiation from complex structural systems
SA02—Source characterization in structure borne noise problems
SA03—Ground vehicle noise and vibration–Joint with: ASACOS
SA04—Acoustic imaging in confined space
SA05—Distributed active noise and vibration control
SA06—Active noise control: New strategies and innovative concepts
SA07—Efficient boundary element methods
SA08—Fluid-structure interaction

Other Technical Events

Plenary Lectures

Four plenary lectures are planned: two at the opening ceremony in the "Grand Amphitheater" on Monday morning, 30 June, and two others during the ASA/EAA/SFA plenary session on Wednesday afternoon, 2 July. For additional information on plenary lectures, please check at the following link: http://www.acoustics08-paris.org/program/plenary/

Technical Tours

Several technical tours are planned during the conference, including:

- IRCAM (Institut de Recherche et Coordination Acoustique/Musique): demonstrations in musical acoustics, sound spatialization, sound design, synthesis methods and sound processing, www.ircam.fr/?L=1
- Cité de la Musique: music museum and temporary exhibitions, www.cite-musique.fr/anglais/accueil.html
- Musée des Arts et Métiers: museum of science and technology including Blériot airplane, Lavoisier laboratory, www.arts-et-metiers.net/?lang=ang
- UGC theatre in Paris La Défense: a theatre that was built using innovative technical solutions for sound insulation and room acoustic treatment.
- ONERA/CEPR: the French Aerospace Laboratory, Thruster Test Center: anechoic wind tunnel, RACE tests facility (fans and turbojets), http://www.onera.fr/english
- CSTB (French Center for Building Technologies): acoustic test facilities for building materials (windows, panels…), http://international.cstb.fr
- LAM: visit to the Musical Acoustics Laboratory: musical acoustics and demonstrations, http://www.lam.jussieu.fr

Additional information about technical tours and the registration procedure is given at the following link: http://www.acoustics08-paris.org/program/technical-tours/

Exhibition

The meeting will be highlighted by an important exhibition (77 booths of 9 sqm). The exhibition will cover all areas of acoustics (instruments, equipment, software, services…).

The exhibition hall is conveniently located near the registration area and meeting rooms. In order to improve the contact between participants and exhibitors, both morning and afternoon coffee breaks as well as the opening reception will be held in the exhibition area.

The exhibition will open on Monday, 30 June, and will close Friday, 4 July (installation on Sunday, 29 June). A 5% discount will be given to Companies or Institutions that are sustaining members of ASA, SFA or any national acoustical society of the EAA. Additional information is available at the following link: http://www.acoustics08-paris.org/exhibition

Alternatively, information can also be obtained by contacting the conference secretariat: Armelle Guilloux, Acoustics'08 Paris c/o SFA Société Française d'Acoustique, 23 avenue Brunetière, 75017 Paris, France, phone: +33 637883979; fax: +33 148889060, email: acoustics08@laposte.net

Gallery of Acoustics

The Technical Committee on Signal Processing in Acoustics will sponsor the 10th Gallery of Acoustics at Acoustics'08 Paris. The Gallery provides a means by which the natural beauty, and the aesthetic and artistic appeal of acoustic phenomena can be shared and appreciated. Entries are invited and must be submitted by 3 March 2008. See the following website for full details: ⟨http://www.acoustics08-paris.org/technical-information/gallery/

Student Information

Special attention will be given to young acousticians at Acoustics'08 Paris by all of the organizing societies—ASA, EAA, SFA—as well as by the Conference Organizing Committee. Student representatives are part of the organizing committees. All means will be used to attract as many students as possible.

Travel Grants for Students

Travel grants will be provided to students and young scientists presenting papers to partially defray transportation expenses to the Conference. Students and young presenters who want to apply for travel grants must do so online, during the abstract submission process. Before applying, please refer to the conference website to verify the rules and criteria for each society.

http://www.acoustics08-paris.org/students/travel-grants

Student Dormitories

A limited number of rooms will be reserved for students. Priority will be given to students until 8 February 2008, then open to all participants. The rooms will be allocated on a "first-come first-served" basis. Standard rooms (single and twins) have individual bathrooms but no air-conditioning. For dormitory reservations, please go to the following link: http://www.acoustics08-paris.org/students/dormitories

Student Luncheon

A student luncheon will be held on Thursday, 3 July. This luncheon is free but students who wish to attend MUST PRE-REGISTER for the luncheon when they register for the conference:

http://www.acoustics08-paris.org/registration/

Deadline for pre-registration is Friday, 23 May.

Web Registration Procedures

The registration procedure is used to provide a Registration Identifier (confirmation number and a password)

To register, access the conference website directly at: http://www.acoustics08-paris.org/registration/

Alternatively, you can log on to the ASA Home Page (http://asa.aip.org), or the EAA Home Page (http://www.european-acoustics.net/Events/2008/ForumAcusticum-ASA-SFA-Meeting and click on "Acoustics'08 Paris" meeting and then on "Registration Submission" from the conference main page in the left column.

1. An item list appears below "Registration submission." Click on "Register now" subtitle.
2. On the next page, select the link corresponding to your situation (Regular registration, SFA member registration and Exhibitor registration); personal data form will be automatically filled out for SFA members.
3. The next page is used for online registration. Please enter your family name, first name, postal address, zip code, city, country and email address in the corresponding boxes. The email address will be used for sending you all the conference information. It must be confirmed (double checked) during the registration process.
4. Please note that your postal address will be used for conference invoice (receipt) unless you enter another address in the box "Address for the invoice."
5. In the "Registration options," select your registration status (e.g.: member of the ASA, the SFA or any National European Society of the EAA) in the scrolling menu "Registration as." The corresponding registration cost appears in the right column.
6. Click on "Validate" at the bottom of the screen.
7. You will then be asked to provide a recent photograph (jpeg). A photo database will be issued and made available during the Conference. It will be of great help, in particular, for students wishing to meet "seniors." It will also be useful to participants from around the world to meet one another for discussions and at social events.
8. At the bottom of this page, you will find your Registration Identifier (confirmation number (CXXXXXXXXX-XXX) and a 6-character Password). This is the acknowledgment that your registration has been entered into the database. A message acknowledging your registration and containing your identifier will be automatically sent to your email address. Please retain your registration identifier data as they will enable you to change your photograph or correct your personal information. In order to modify your registration options or to buy extra items, please contact the conference secretariat by email (acoustics08@laposte.net).
9. At this stage, you can either proceed immediately to submit payment, or return later on.
10. If you wish to pay for your registration, click on the "use the following link…" hyperlink following the sentence "You wish to pay for this registration." Again, you can return to this step later on, after acceptance, for instance.

Special Meeting Features

Conference Proceedings

The Proceedings of the Acoustics'08 Paris conference will be published and distributed at the conference to all participants. It will consist of three CDs (or DVDs): one general, one Euronoise, and one ECUA. The three CDs are included in the registration fees. There will also be paper proceedings available for the ECUA part of the conference which will be sold separately at production cost.

In addition to the abstracts, authors are given the opportunity to publish a full paper (6 pages) in the Conference Proceedings. Authors must write their papers in accordance with the templates provided on the conference website and then upload them into the conference database. Detailed instructions on the paper submission procedure and template are given at the following link: http://www.acoustics08-paris.org/technical-information/instructions/paper-template/

Deadline for paper submission: Wednesday, 30 April.

Social Events

Several social events are planned at Acoustics'08 Paris. Some events are open to all participants, and participation in others require a reservation.

Events open to all participants:

- Opening Ceremony: Monday morning, 30 June
- Exhibitors' Reception: Monday noon, 30 June
- Conference Reception: Tuesday evening, 1 July
- Closing Ceremony: Friday afternoon, 4 July

Events available upon reservation or invitation:

- Conference banquet: upon registration, Wednesday evening, 2 July, Fee: 70€
- Women in Acoustics luncheon: Tuesday, 1 July. Upon registration. Fee: €10 (students €5)
- Students' luncheon: Thursday, 3 July, free for all students but requires prior reservation.

For all events requiring reservations, registration must be made online before Friday, 23 May. After 23 May, reservations can only be accepted on a space-available basis.

Awards Ceremony

The ASA/EAA/SFA Plenary Session will be held on Wednesday afternoon, 2 July. At this Ceremony, the Societies' awards will be presented and newly-elected Fellows will be recognized.

Sponsoring

Companies or institutions wishing to sponsor Acoustics'08 Paris are warmly welcomed. The sponsoring rules can be found at the following link: http://www.acoustics08-paris.org/sponsoring/

Transportation, City Information and Hotel Accommodations

Air Transportation

Two main airports serve Paris: Paris-Charles de Gaulle International Airport (Airport Code CDG) and Paris-Orly International Airport (Airport Code ORY). They are served by most international airlines. Please keep in mind your departure airport and terminal since ground transportation has two drop off points. Further information is available on the following link: http://www.aeroportsdeparis.fr/Adp/en-GB/Passagers

Ground Transportation

From Paris-Charles de Gaulle International Airport

Paris-Charles de Gaulle International Airport is approximately 18.5 miles (30 km) from the Palais des Congrès (Porte Maillot).

Taxis: A taxi ride costs around €50.00 from the Paris-Charles de Gaulle international airport to Paris city center or to Palais des Congrès. A supplement of approximately 15% applies at night from 7:00 pm to 7:00 am and on Sundays and public holidays. Taxis are available upon exiting the baggage claim area at your arrival terminal: terminal 1 exit 20 at arrivals level, terminals 2A and 2C exit 6, terminals 2B and 2D exit 7, terminals 2E and 2F exit 1. Approximate travel time: 35 minutes.

Air France Shuttle bus: Direct shuttle service is available from the airport to Palais des Congrès for €13.00 (per person, one-way) or €18.00 (per person, round-trip). Follow the "Air France Shuttle" sign at the terminal. Line 2 shuttle stops at "Porte Maillot" 100 meters from the Conference entrance (boulevard Gouvion St-Cyr opposite the Méridien hotel). Shuttles are available from 5:50 am to 11:00 pm. Approximate travel time: 45 minutes.

Public transportation: Follow "RER" signs at the terminal. Take RER B towards "Robinson" or "Saint Rémy les Chevreuse" and exit at "St Michel Notre Dame." At "St Michel Notre Dame" station, follow "RER C" signs. Take RER C towards "Pontoise" or "Argenteuil" and exit at "Porte Maillot station." RER B is available from 5:00 am to 11:50 pm. Approximate travel time: 1 hour. Ticket machines are in the RER station. One-way ticket costs €8.20.

From Paris-Orly International Airport

Paris-Orly International Airport is approximately 14.3 miles (23 km) from the Palais des Congrès (Porte Maillot).

Taxis: A taxi ride costs around €50.00 from the Paris-Orly international airport to Paris city center or to Palais des Congrès. A supplement of approximately 15% applies at night from 7:00 pm to 7:00 am and on Sundays and public holidays. Taxis are available at terminal Paris-Orly Sud exit L and terminal Paris-Orly Ouest exit I at the arrivals level. Approximate travel time: 30 minutes.

Air France Shuttle: Shuttle service is available from the airport to Invalides for €8.00 (per person, one-way) or €12.00 (per person, round-trip). Follow the "Air France Shuttle" sign at the terminal. Line 1 shuttle stops at "Invalides." Then buy a metro ticket (cost €1.30) and take RER C towards "Pontoise" or "Argenteuil" and exit at "Porte Maillot station." Shuttles are available from 6:00 am to 11:00 pm. Approximate travel time: 45 minutes.

Public transportation: Follow "Orlyval" signs at the terminal. Take Orlyval metro towards "Antony" and exit at "Antony" station. Take RER B towards "Aéroport Charles de Gaulle" or "Mitry Claye" and exit at "St Michel Notre Dame." At "St Michel Notre Dame" station, follow "RER C" signs. Take RER C towards "Pontoise" or "Argenteuil" and exit at "Porte Maillot" station. "Orlyval" is available from 6:00 am to 11:00 pm. Approximate travel time: 1 hour. Ticket machines are in the RER station. One-way ticket costs €9.30.

Car Rental: Car rental agencies have offices in both airports.

Parking: Underground parking (1,500 spots) is available with direct access to the Palais des Congrès. 24h parking rate for cars is €27.00.

Paris City Information

Weather—Paris in July is usually sunny with average high and low temperatures of 59 and 77 degrees F (15 and 25 degrees C) during the day. Showers may occur with an average precipitation of 2.3 inches (59 mm).

Local Time—Standard time zone (GMT+1 hr) and daylight saving time (+1 hr) give a time zone offset of GMT +2 hours.

Currency and foreign exchange—The local currency is EURO (€). Currency can be exchanged at some banks and preferably at "Bureau de change" that can be found all around the "Palais de Congrès," in large department stores, main railway stations, airports and near tourist attractions. Please note: although the exchange rate is fixed, commission rates are not. They have to be clearly displayed.

Banking and credit cards—Banks are open generally from 9:00 am to 5:00 pm with an optional break at lunch time, from Monday to Friday or from Tuesday to Saturday. Automatic Teller Machines (cash dispensers) can be found almost everywhere and accept most international cards (Visa, Eurocard, and Mastercard). In shops and restaurants, mainly Visa, Eurocard, and Mastercard are accepted. American Express cards are also accepted in several places (but not all, please check signs at the entrance).

Electricity—The power supply is 220 V, 50 Hz. Round European style two-pin plugs are used. Appliances designed to operate on 110/120 V require a voltage converter and a plug adapter.

Restaurants—Paris offers a large variety of restaurants and cafés. Prices for a full menu (3 courses) range from about 10 to 40 euros. Typical lunch time is 12:30 pm and dinner 8:00 pm.

Useful links—For the latest information about the city of Paris, please visit the following links:

- Paris Convention and Visitors Bureau website: http://en.parisinfo.com
- General info about Paris: http://www.paris.org, http://www.v1.paris.fr/en/
- Pariscope magazine, what's up in Paris (in French): http://www.pariscope.fr
- Tickets for all shows, concerts, and events in Paris: http://www.ticketnet.fr/shop/en/accueil.asp

Hotel Reservation Information

Four modes of hotel reservation are available to the participants at Acoustics'08 Paris:

- Concorde Lafayette Hotel
- Reservation through a commissioned Travel Agency
- Dormitories (in priority for students)
- Personal reservation

Concorde Lafayette Hotel

A block of guest rooms at discounted rates has been reserved for meeting participants at the Concorde Lafayette Hotel. Early reservations are strongly recommended. Note that the special Acoustics'08 Paris rates are not guaranteed after Wednesday, 30 April 2008.

You must mention Acoustics'08 Paris when making your reservations to obtain the special ASA meeting rates.

http://www.concorde-lafayette.com/en

Hotel Concorde la Fayette
3, Place du Général Koenig
75017 Paris, France
+33 (0)1 40 68 50 68 (T)
+33 (0)1 40 68 50 43 (F)
1-800-888-4747 (US reservations)

Concord Lafayette Hotel is located in the same building as the Palais des Congrès, where the Conference takes place and has an easy access for handicapped persons. It is conveniently located close to the Champs-Elysées (2.5 km), between "La Défense" business district (4.5 km) and "Triangle d'Or" shopping area. The hotel boasts some of the most spectacular views of Paris and the Eiffel Tower (4 km) from its 33 floors of guest rooms and suites and features an exceptional convention space, linked to the Palais des Congrès exhibition center. With 950 rooms overlooking the city, two bars and a restaurant, as well as unrivaled meeting space, the hotel is a true Parisian landmark. For reservations, please use the following link:

http://www.acoustics08-paris.org/venue/accommodation/concorde-lafayette/

Rates (including all taxes and breakfast):
Single: € 190/night
Double: € 200/night
Reservation cut-off date: Wednesday, 30 April 2008

Conference Travel Agency

The organizers have commissioned a Travel Agency to handle hotel reservations. It will offer all price ranges for hotels. To book through this Agency, please connect to the following link: http://www.acoustics08-paris.org/venue/accommodation/travel-agency/

Dormitories

Dormitories are reserved, with priority for students. After 8 February, the remaining rooms will be offered to other participants. Details are described in the Students' section on the Acoustics'08 Paris website.

Pre and Post Conference Tours

A set of tours will be organized by the Conference Travel Agency before and after the conference. Tour descriptions will be posted and reservations can be made at the following link: http://www.acoustics08-paris.org/venue/tours

Registration Information

Registration and payment must be made online at the following link: http://www.acoustics08-paris.org/registration/

On-site registration will be possible but advance registration is highly recommended. The on-site registration desk will open on Sunday, 29 June (2:00 p.m. to 8:00 p.m.), on the Main floor of Palais des Congrès and will be available from 8:00 a.m. to 4:00 p.m. during the entire week. Registration fees are given in the table below.

All amounts are given in € Euros. No tax to be added (the SFA is not subject to VAT (sales) tax). Credit card is the preferred mode of payment. The secure registration system accepts the following cards: Visa, MasterCard, and American Express

CATEGORY	PREREGISTRATION by 8 February 08	PREREGISTRATION by 8 March 08	PREREGISTRATION after 8 March 08
Members[a]	€450	€500	€550
Nonmembers	€500	€550	€600
Student Members[b]	€150	€200	€250
Student Nonmembers[b]	€200	€250	€300
Emeritus members[c]	€150	€200	€250
Accompanying Persons[d]	€150	€150	€200
Ecua proceedings	€15	€15	€25
Banquet tickets[e]	€70	€70	€85
Women in Acoustics luncheon[e]	€10	€10	€20
Women in Acoustics luncheon (students)	€5	€5	€10

[a]Members of ASA, EAA, SFA and IEEE/OES. Membership number will be required.
[b]Active students or young investigators who obtained their Diploma (Master, PhD) less than a year ago.
[c]ASA and EAA Emeritus Members
[d]Accompanying Person registration entitles participation in the Accompanying Persons Program, but does not include access to conference rooms (except Opening and Awards Ceremonies).
[e]Reservations MUST be made before Friday, 23 May.

Cancellation Policy

A processing fee of €50 will be charged for a cancellation before 8 April 2008. No reimbursement will be made for a cancellation received after 8 April 2008.

Abstracts and papers of cancelled corresponding authors will not be published or scheduled for presentation.

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.

2008

29 June–4 July — Joint Meeting of the Acoustical Society of America, European AcousticsAssociation and the Acoustical Society of France, Paris, France) [Acoustical Society ofAmerica, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

28 July–1 Aug — 9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609, Tel.: 860-572-0680; Web: www.icben.org. Email: icben2008@att.net

10–14 Nov — 156th Meeting of the Acoustical Society of America, Miami, FL [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: http://asa.aip.org].

2009

18–22 May — 157th Meeting of the Acoustical Society of America, Portland, OR [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: http://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)

Volumes 105-114, 1999-2003: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 616, Price: ASA members $50; Nonmembers $90 (paperbound)

ACOUSTICAL NEWS—INTERNATIONAL

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

Statistics-19th International Congress on Acoustics in Madrid 2007

Statistics of the 19th International Congress on Acoustics which was held in Madrid 2–7 September 2007 have now been released by the organizers.

The last column of the following compilation shows the number of registered participants from the various countries represented at the Congress. The other columns show the number of participants at previous Congresses held in Belgrade (1989), Beijing (1992), Trondheim (1995), Seattle (1998), Rome (2001), and Kyoto (2004). The table reflects the fact that there have been some political changes in the 18 years covered by these statistics.

In addition to the numbers of participants listed in the Madrid column there were 26 participants whose countries were "uncertain" according to the organizers. The number of papers presented at the Madrid ICA was 1295 which includes 219 posters.

Country	Belgrade	Beijing	Trondheim	Seattle	Rome	Kyoto	Madrid
Algeria	0	0	1	0	1	0	8
Argentina	0	0	0	0	5	0	4
Armenia		0	0	0	1	0	0
Australia	4	12	19	36	27	16	20
Austria	2	0	6	11	10	5	8
Belarus		0	0	0	2	0	0
Belgium	9	4	9	15	23	14	18
Brazil	1	3	3	12	18	4	12
Bulgaria	6	0	0	0	0	0	0
Canada	18	10	10	75	17	22	23
Chile	0	1	1	2	3	0	12
China	15	370	4	47	30	52	14
Colombia	0	0	0	1	0	0	0
Dem. Rep. Congo	0	0	0	0	0	0	3
Croatia		1	0	0	6	0	4
Cuba	0	0	0	0	0	0	1
Czech Republic			4	3	11	3	10
Czechoslovakia	7	0					
Denmark	19	9	28	17	25	20	21
Ecuador	0	0	0	0	0	0	1
Egypt	1	0	0	1	1	1	2
Estonia		0	4	2	3	1	2
Finland	4	2	10	7	14	10	15
France	67	35	63	95	99	39	153
Germany (FRG)	55	32	41	49	64	58	100
Germany (GDR)	3						
Greece	3	0	0	1	4	0	3
Guinea	0	0	0	0	0	0	1
Hong Kong	0	2	0	4	0	3	0
Hungary	22	0	11	1	6	1	5
Iceland	0	0	0	0	1	1	2
India	8	3	3	9	12	6	10
Indonesia	0	1	0	0	2	1	1
Iran	1	2	0	0	0	1	6
Ireland	0	0	0	3	3	0	1
Israel	0	3	0	3	3	2	5
Italy	13	8	16	28	160	21	67
Japan	87	106	96	173	237	723	188
Korea	0	21	4	33	25	69	30
Kuwait	0	0	0	0	1	0	0
Latvia		0	0	0	0	0	2
Lebanon	0	0	0	0	2	0	0
Lithuania		0	0	0	2	2	1
Malaysia	0	2	0	0	0	0	0
Mexico	0	0	0	8	2	1	9
Morocco	0	1	0	0	0	0	2
Netherlands	21	8	14	27	19	21	42
New Zealand	1	0	2	9	5	3	3
Nigeria	0	0	0	0	0	5	1
Norway	9	3	89	18	24	7	16
Peru	0	0	1	0	0	1	3
Poland	16	2	24	12	36	13	43
Portugal	1	1	3	0	9	5	15
Romania	1	0	1	0	6	0	3
Russia		8	13	27	32	11	23
Serbia						2	2
Singapore	0	3	1	3	2	2	1
Slovakia			1	1	3	2	0
Slovenia		0	0	0	3	0	2
South Africa	1	1	0	3	0	0	0
Spain	4	2	7	13	25	12	251
Sweden	16	10	40	25	35	18	27
Switzerland	4	0	7	7	21	2	9
Taiwan						10	6
Thailand	0	1	0	0	0	0	0
Tunisia	0	0	0	0	2	1	0
Turkey	0	0	1	1	7	1	3
UAE	0	0	0	0	0	1	1
Uganda	0	0	0	0	0	2	1
United Kingdom	27	20	35	77	55	41	102
Ukraine		0	0	2	1	0	6
Uruguay	0	0	0	0	2	0	1
USA	61	64	69	1066	176	89	116
USSR	31						
Uzbekistan		0	0	0	1	0	1
Venezuela	0	0	0	2	3	0	5
Yugoslavia	169	3	0	3	6		
Total	707	754	642	1932	1293	1332	1473
Accompanying persons	63	95	98	192	N/A	91	100

The European Acoustics Association elects new Board

The General Assembly of the EAA, meeting in Madrid during the recent ICA, elected a new Board of Officers for the term 2007–2010. The new Officers are:

President:	Luigi Maffei (Second University of Naples, Italy)
Vice Presidents:	Michael Vorländer (RWTH Technical University, Aachen. Germany) Peter Swensson (University of Science and Technology, Trondheim. Norway)
General Secretary:	Kristian Jambrosic (University of Zagreb, Croatia)
Treasurer:	J. Salvador Santiago.

Canada to host 2013 ICA

The 2013 International Congress on Acoustics will be a joint meeting of the Canadian Acoustical Association and the Acoustical Society of America which will be held in the Palais des congrès de Montréal June 2–7, 2013. Contact addresses will be announced later.

 0001-4966/2008/123(2)/579/1/$23.00

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 1N01, 2 Huntington Quadrangle, Melville, NY 11747-4502, e-mail: asa@aip.org*

Adaptive wave field synthesis [43.38.Md, 43.60.Tj, 43.50.Ki] —Philippe-Aubert Gauthier, *Groupe d'Acoustique de l'Université de Sherbrooke (GAUS), Department of mechanical engineering, Engineering Faculty, Université de Sherbrooke, Sherbrooke, Québec, Canada, September 2007 (Ph.D.).* Wave field synthesis (WFS) is a sound field reproduction technology which assumes that the reproduction environment is anechoic. A real reproduction space reduces the objective accuracy of WFS. This research involves the improvement of WFS performance by active noise control and adaptive filters. Simulations of sound field reproduction in a three-dimensional space show the physical possibility of progressive sound field reproduction in a closed space on the basis of optimal control of the reproduction errors at a sensor array. The proposed solution is adaptive wave field synthesis (AWFS). The originality of AWFS is the combination of the minimization of the reproduction error along with a penalty for any departure from the WFS solution. AWFS is theoretically analyzed on the basis of singular value decomposition. This suggests that the AWFS underlying mechanism is independent radiation mode control. Two adaptive algorithms for AWFS are developed. An experimental AWFS system is tested in different rooms (hemi-anechoic, laboratory, reverberant). It shows that AWFS performs better than WFS: AWFS significantly reduces the room effects on sound field reproduction. The algorithm based on independent radiation mode control contributes to the enlargement of the effective reproduction region. The thesis is written in French and in English.

Advisor: Alain Berry

Multifrequency ultrasound radiation force excitation and motion detection of harmonically vibrating scatterers [43.25.Qp, 43.80.Ev, 43.80.Vj] —Matthew W. Urban, *Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, Rochester, Minnesota 55905 urban.matthew@mayo.edu, September 2007 (Ph.D.).* Elasticity imaging is a medical imaging modality that creates images based on the mechanical response of tissue. In some techniques, ultrasound radiation force has been used to deform tissue and ultrasonic or other techniques detect the tissue displacement. The focus of this dissertation was development of methods utilizing multiple ultrasound frequencies simultaneously to produce multifrequency harmonic radiation force and development of harmonic motion detection techniques adapted for vibrometry with simultaneous radiation force excitation. Beamforming for multifrequency radiation force was developed analytically, numerically, and experimentally. Sample applications in the fields of vibrometry and vibro-acoustic breast imaging are shown. Numerical models for harmonic motion detection of a vibrating reflective target and a vibrating scattering medium were developed. Parametric analyses were performed to extensively analyze the error in the vibration amplitude and phase measurements. An experimental method using multifrequency radiation force and harmonic motion detection with a single transducer is described and experimental results for a reflective target and scattering medium are shown. This work produced techniques for excitation and measurement of small amplitude harmonic motion for performing vibro-acoustography and vibrometry in a multiplexed manner. Models of vibration responses were developed to further understanding of current and future applications in vibro-acoustography and vibrometry.

Advisor: James F. Greenleaf

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

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:

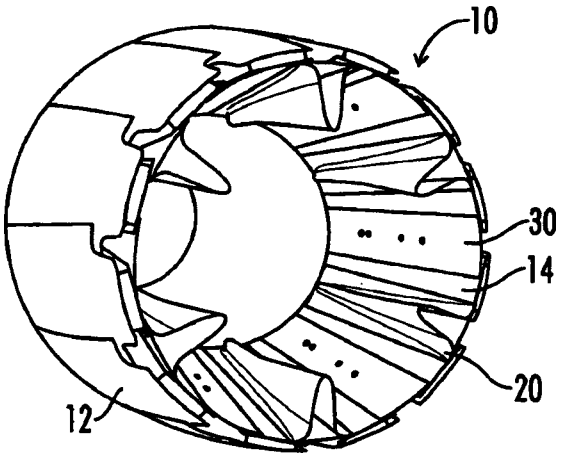
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026–2453*
ALIREZA DIBAZAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DIMITRI DONSKOY, *Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, New Jersey 07030*
GEOFFREY EDELMANN, *Naval Research Laboratory, Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,240,493

43.25.Vt METHOD AND DEVICE FOR REDUCING ENGINE NOISE

John M. Seiner, assignor to The University of Mississippi
10 July 2007 (Class 60/770); filed 30 November 2004

A fighter jet exhaust noise attenuator is claimed where semieliptical projections **20** into the exhaust stream forces the injection of static



atmosphere air into the high speed exhaust flow, reducing the shear gradient and reducing the Mach wave radiation.—AJC

7,263,031

43.28.Tc DISTANCE MEASURING DEVICE FOR ACOUSTICALLY MEASURING DISTANCE

Hughes Sanoner *et al.*, assignors to Solar Wide Industrial Limited
28 August 2007 (Class 367/99); filed 1 November 2005

The distance measuring device utilizes acoustic echo-location method measuring time delay and attenuation of a signal reflected from a target. Additionally, air temperature and humidity are measured and these measurements are used to estimate signal attenuation of the reflected signal. The temperature measurement is also used to compute the sound speed in air: $v = 331.45 \times \mathrm{Sqr}((T + 273.16)/273.16)$ m/s, where T is the temperature in degrees Celsius. The sound speed correction for temperature fluctuation and the comparison between the measured and predicted attenuation are used to improve accuracy and reliability of the device.—DMD

7,263,373

43.28.Tc SOUND-BASED PROXIMITY DETECTOR

Sven Mattisson, assignor to Telefonaktiebolaget L M Ericsson (publ)
28 August 2007 (Class 455/456.3); filed in Sweden 28 December 2000

The patent describes a proximity detector for use in mobile telephones enabling automatic adjustment of the loudspeaker sound level in the phone depending on the distance between the phone and user. The phone's loudspeaker emits a controlled signal and the phone's microphone receives the emitted signal directly propagated from the speaker and reflected from a user. The CPU of the phone correlates the directly transmitted and the reflected control signals and estimates the difference between the known distance from the speaker to the microphone and the distance along the reflected path.—DMD

7,257,049

43.30.Tg METHOD OF SEISMIC SURVEYING, A MARINE VIBRATOR ARRANGEMENT, AND A METHOD OF CALCULATING THE DEPTHS OF SEISMIC SOURCES

Robert Laws and Stephen Patrick Morice, assignors to W'etsern Geco L.I.
14 August 2007 (Class 367/23); filed in United Kingdom 2 September 1999

When surveying the bottom of the ocean, surface reflections can inter-

 0001-4966/2008/123(2)/583/16/$23.00

fere with the return signal. A strangely worded methodology to reduce the ghost paths from marine vibrators is described.—GFE

7,254,092

43.30.Vh METHOD AND SYSTEM FOR SWIMMER DENIAL

Frederick R. DiNapoli, assignor to Raytheon Company
7 August 2007 (Class 367/138); filed 31 March 2005

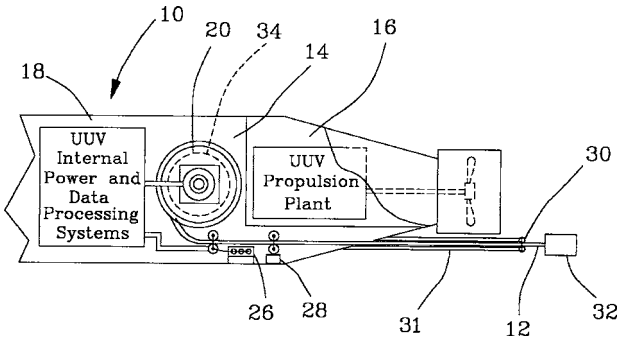
The use of time-reversal focusing to ensonify an intruder with high intensity sound for the purpose of deterrence or death is described.—GFE

7,252,046

43.30.Yj APPARATUS FOR DEPLOYING AND RECOVERING A TOWED ACOUSTIC LINE ARRAY FROM AN UNMANNED UNDERSEA VEHICLE

Richard M. Ead and Robert L. Pendleton, assignors to The United States of America as represented by the Secretary of the Navy
7 August 2007 (Class 114/254); filed 8 December 2003

A drum for deploying a towed line array from an unmanned undersea



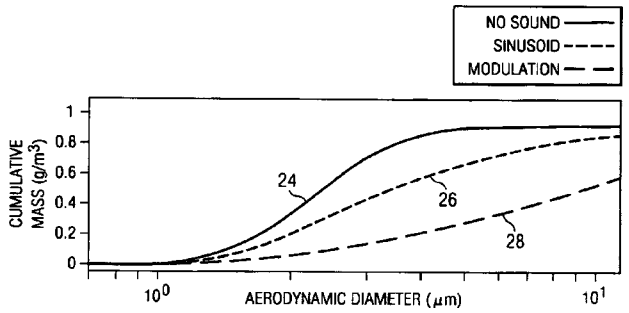
vehicle is described.—GFE

7,238,223

43.35.Ty ACOUSTICAL STIMULATION OF VAPOR DIFFUSION SYSTEM AND METHOD

G. Douglas Meegan, Jr., assignor to Board of the Regents, The University of Texas System
3 July 2007 (Class 95/29); filed 3 November 2003

An acoustic device to enhance the removal of mercury vapors in the exhaust from a lignite fueled power plant is claimed. A 150 dB SPL acoustic field is projected into that exhaust stream into which a sorbent powder,



activated charcoal of some 15 μm particle size, is also injected. Optimum mercury removal occurs when the sound audio frequency is around 500 Hz and whose amplitude is modulated **28**. (A total of 195 claims are presented!)—AJC

7,270,386

43.35.Zc LIQUID-DETECTING DEVICE AND LIQUID CONTAINER WITH THE SAME

Tomoaki Takahashi *et al.*, assignors to Seiko Epson Corporation
18 September 2007 (Class 347/7); filed in Japan 10 February 2003

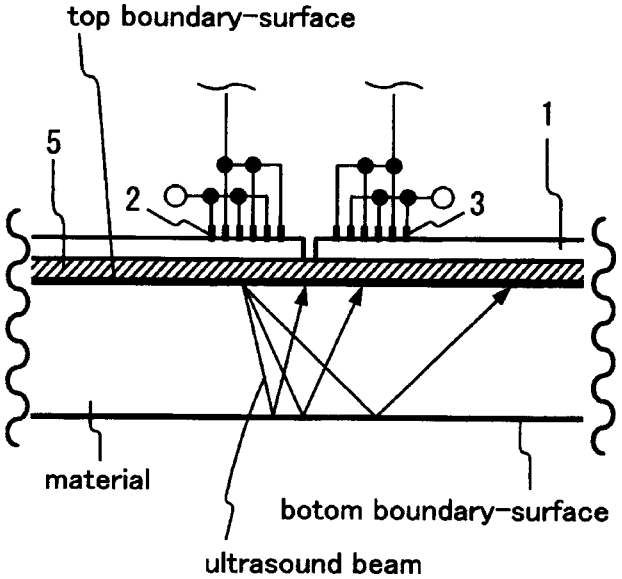
A liquid level detector is described that consists of a hole in the container wall with a piezoelectric disk covering it. Not surprisingly, the resonant frequency of this disk depends on the height of the liquid covering it. The device is being applied to measure levels of ink in printer heads, and it is claimed to be relatively insensitive to changes in viscosity and density of the inks. It is not very well written and difficult to read, even for a patent.—JAH

7,262,542

43.38.Ar ULTRASOUND RADIATION DEVICE INTO A MATERIAL

Kohji Toda, Yokosuka, Japan
28 August 2007 (Class 310/313 B); filed 7 November 2005

The authors describe a series of ultrasonic "leaky wave" transducers suitable for radiating energy into a substrate medium by mechanical contact with the base of the transducer. This type of transducer, it is said, might be useful in medical ultrasound probing. Referring to the figure, the idea is to use interdigitated electrodes **2** to launch a surface acoustic wave, and intercept the surface wave component with a hole or other absorbing barrier. The



receiver transducer **3** then picks up the bulk wave that leaks into the neighboring medium below. The authors claim that this allows the transducer to adapt its operating frequency to varying depths of material, but this is rarely desirable in such applications, so it is not clear what benefits this arrangement really confers to the user. The patent is brief and mostly devoted to diagrams of different implementations of this idea—some of them quite intriguing. Ultimately, reading the document leaves one wondering whether any of this works, as it is devoid of quantitative information—JAH

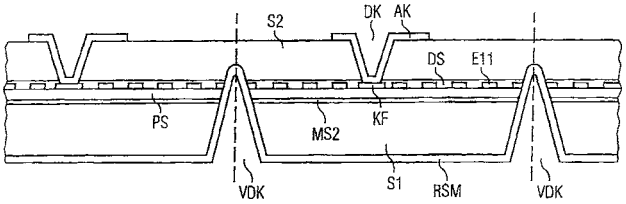
7,262,676

43.38.Ar ELECTROACOUSTIC COMPONENT AND METHOD FOR THE PRODUCTION THEREOF

Werner Ruile and Ulrike Rösler, assignor to EPCOS AG
28 August 2007 (Class 333/193); filed in Germany 4 June 2003

The authors disclose a guided interfacial wave resonator design that

can be fabricated from bonded silicon wafers with a piezoelectric layer sandwiched between them. The figure shows a cross section of such a device, where **S1** and **S2** are silicon wafers and **PS** is the piezoelectric intermediate layer. The patent goes through the fabricaton process in great detail,



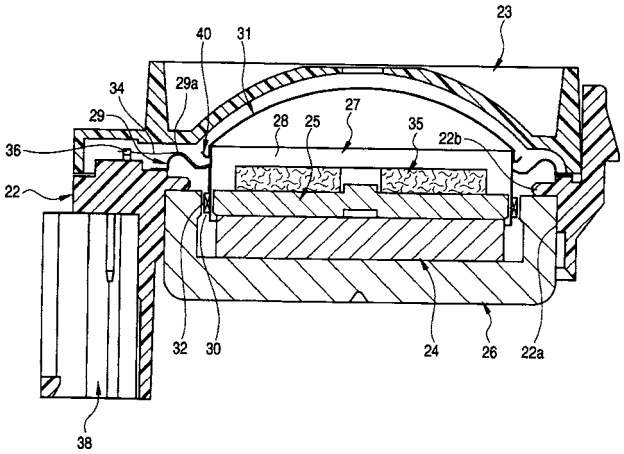
though no performance data are given. The benefits of this design are said to be compatibility with current CMOS fabrication processes and the relatively small step heights the electrical leads must cross to make electrical contact.—JAH

7,245,739

43.38.Dv DOME TYPE DIAPHRAGM AND LOUDSPEAKER APPARATUS

Tatsuya Suzuki, assignor to Pioneer Corporation
17 July 2007 (Class 381/430); filed in Japan 14 January 2004

Reinforcing portion **40** of dome **31** is curved, canted up, and/or folded back up instead of being finished in the horizontal plane. This increases the rigidity of the dome and reduces "transmission loss" while also reducing the



chances of portion **40** from contacting edge **29** and lead wires **34** and causing a short circuit.—NAS

7,251,196

43.38.Dv PASSIVE OPTICAL DETECTION OF UNDERWATER SOUND

Lynn T. Antonelli and Fletcher A. Blackmon, assignors to The United States of America as represented by the Secretary of the Navy
31 July 2007 (Class 367/149); filed 31 May 2005

An overbroad patent on acousto-optic sensing, that is the remote sensing of underwater acoustic pressure via laser interferometry, is described. The basic concept is the same as is used to eavesdrop on people's conversations by measuring the vibration on a window.—GFE

7,259,864

43.38.Dv OPTICAL UNDERWATER ACOUSTIC SENSOR

Lynn T. Antonelli *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
21 August 2007 (Class 356/502); filed 25 February 2005

Instead of a traditional hull mounted underwater sensor, an acousto-optic sensor is described that measures the vibration of the outer hull of a submarine from a laser system placed within the inner hull.—GFE

7,260,023

43.38.Dv REMOTE UNDERWATER LASER ACOUSTIC SOURCE

Theodore G. Jones *et al.*, assignors to United States of America as represented by the Secretary of the Navy
21 August 2007 (Class 367/149); filed 2 November 2005

Underwater sound is generated via pulsed laser ionization of the surface by this opto-acoustic device. A method of simulating an underwater source location using pulse compression is described.—GFE

7,268,503

43.38.Dv VIBRATION LINEAR ACTUATING DEVICE, METHOD OF DRIVING THE SAME DEVICE, AND PORTABLE INFORMATION APPARATUS USING THE SAME DEVICE

Hirokazu Yamasaki and Koji Kameda, assignors to Matsushita Electric Industrial Company, Limited
11 September 2007 (Class 318/114); filed in Japan 4 April 2002

The patent describes an electromagnetic shaker for use in mobile phones. It consists of a couple of permanent magnets, an energizing coil attracting and repelling the magnets, and a digitally controlled switching circuitry to synchronize the applied electrical current with mechanical motion of the magnets.—DMD

7,250,706

43.38.Fx ECHO SOUNDER TRANSDUCER

Hiroshi Shiba, assignor to NEC Corporation
31 July 2007 (Class 310/325); filed in Japan 1 July 2004

In order to increase low frequency sound pressure output without increasing transducer size, a technique is proposed to activate the bend vibration mode of a transducer.—GFE

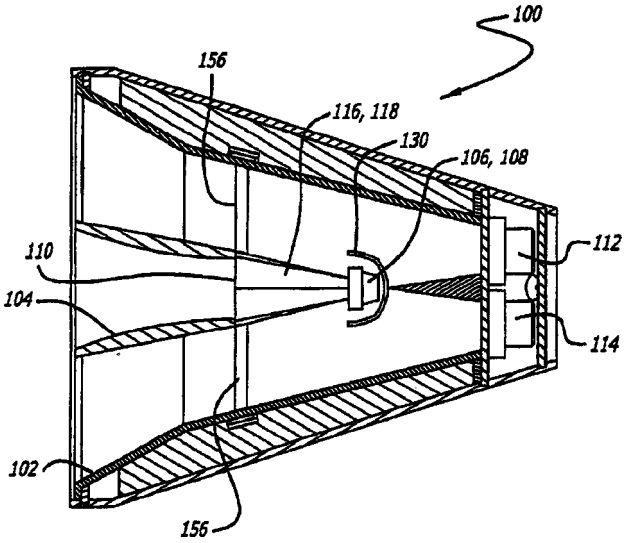
7,236,606

43.38.Ja SOUND SYSTEM HAVING A HF HORN COAXIALLY ALIGNED IN THE MOUTH OF A MIDRANGE HORN

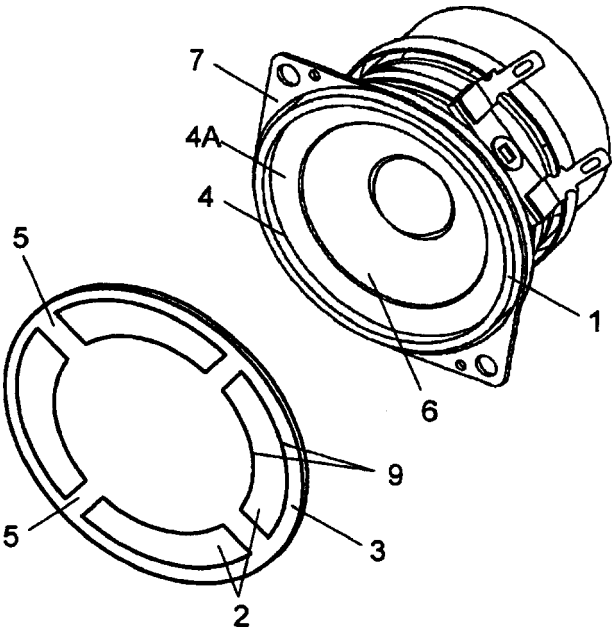
Bernard M. Werner, assignor to Harman International Industries, Incorporated
26 June 2007 (Class 381/342); filed 7 March 2002

Coaxial loudspeaker assembly **100** consists of high-frequency assembly composed of drivers **106**, **108**, shield **130**, couplers **116**, **118**, and high-frequency horn **104**. This is placed within midfrequency horn **102** by struts **156**. The patent asserts that the shadowing by the high frequency driver can

be minimized if the height and width of HF horn **104** is within a range "of between about 0.25 to about 0.4 as large as the height and width of the



midrange horn" **102**. Note the septum between the two midfrequency drivers that ends at the HF shield. The patent asserts that the configuration described minimizes any sectoral horn effects that may be seen in similar nested horn systems.—NAS



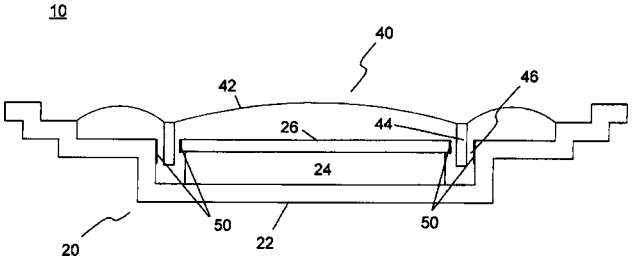
arrangement is preferable to other geometries having the same percentage of open area. The results are unexpected but believable for the configurations shown.—GLA

7,266,214

43.38.Ja AUDIO SPEAKER AND METHOD FOR ASSEMBLING AN AUDIO SPEAKER

Shiro Tsuda, assignor to Ferrotec Corporation
4 September 2007 (Class 381/415); filed 30 November 2004

The magnetic gap of a loudspeaker is sometimes filled with ferrofluid to add mechanical damping and improve heat dissipation. This patent describes a different application. A volatile magnetic fluid is used as an aid in



centering voice coil **44** during assembly, but the liquid then evaporates, leaving a thin lubricating layer **50** on the gap surfaces.—GLA

7,266,211

43.38.Ja SPEAKER GRILL

Kazuhiko Ikeuchi, assignor to Matsushita Electric Industrial Company, Limited
4 September 2007 (Class 381/391); filed in Japan 7 August 2003

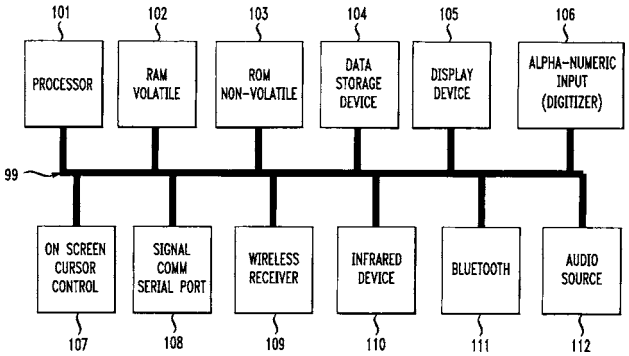
Protective speaker grill **3** is a solid plate with peripheral openings **2**. The combined area of the openings is roughly 50% of the speaker cone area. The patent includes 20 frequency response curves demonstrating that this

7,272,232

43.38.Lc SYSTEM AND METHOD FOR PRIORITIZING AND BALANCING SIMULTANEOUS AUDIO OUTPUTS IN A HANDHELD DEVICE

Jesse Donaldson and Lee R. Taylor, assignors to Palmsource, Incorporated
18 September 2007 (Class 381/55); filed 30 May 2001

Suppose that you are listening to music on your multifunction cellular telephone. To signal an incoming call, the music smoothly fades down and you hear a discrete ring tone with music in the background. If you like, the



background music can continue as you carry on a conversation. When the call is concluded the music returns to its original level. This patent describes a method for realizing the scenario previously described.—GLA

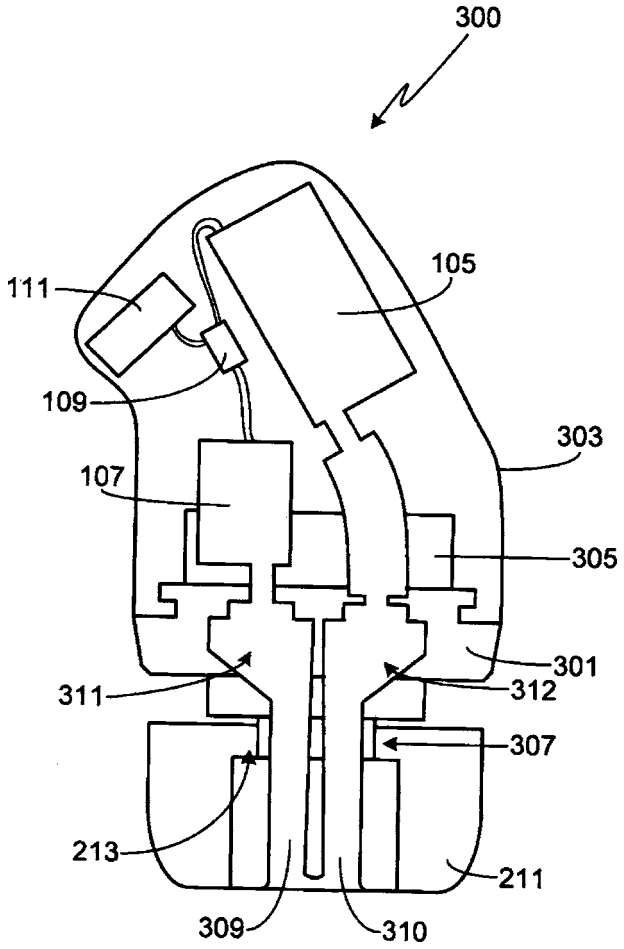
7,263,195

43.38.Si IN-EAR MONITOR WITH SHAPED DUAL BORE

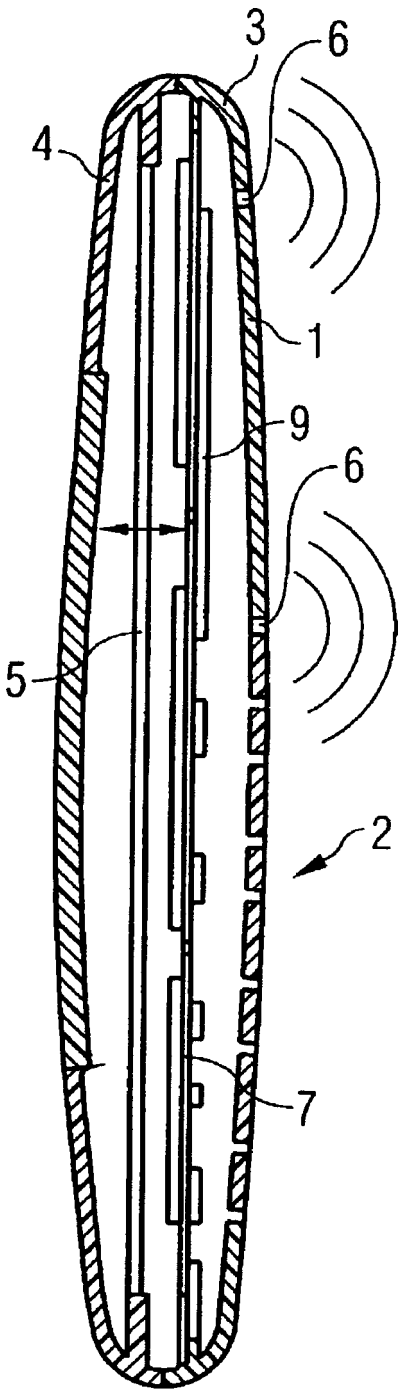
Jerry J. Harvey and Medford Alan Dyer, assignors to Ultimate Ears, LLC
28 August 2007 (Class 381/328); filed 4 February 2005

This is a continuation of earlier United States Patent 7,194,103 that

deals with the design of in-ear monitors, i.e., insertable earphones. Although some monitors are custom-fit, single-size generic models are more popular. A typical two-way in-ear monitor includes a low-frequency transducer **105**, a high-frequency transducer **107**, and a frequency dividing network **109**. In



this design, their respective acoustic outputs are conducted through transition chambers **311**, **312** and sound conducting tubes **309**, **310** all the way to the sound exit opening. By making the sound conducting tubes oval or D-shaped in cross section they can both be accomodated within the limited diameter of a generic earphone.—GLA



"…and a circuit board, wherein said transducer is physically integrated into the structure of the circuit board." This caveat is not mentioned in the abstract and seems to contradict much of the description of the invention. Such a disconnect between the abstract and the claims has become all too common. It probably is not intentional, but it is nonetheless misleading.—GLA

7,263,196

43.38.Si MOBILE COMMUNICATIONS TERMINAL WITH FLAT LOUDSPEAKER DISPOSED IN THE TERMINAL HOUSING

Roland Aubauer and Michael Hülskemper, assignors to Siemens Aktiengesellschaft
28 August 2007 (Class 381/332); filed in the European Patent Office 8 October 2001

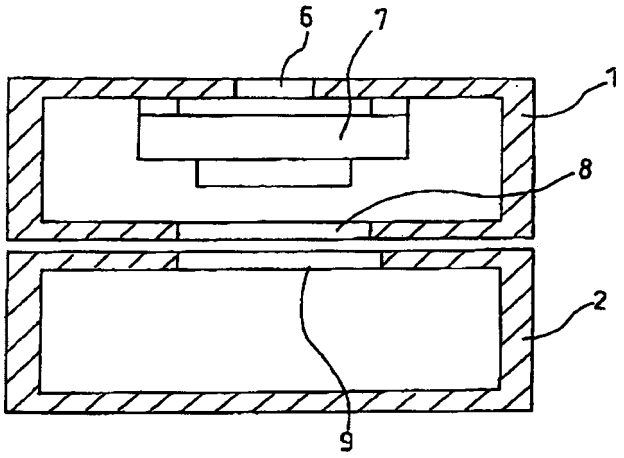
In this design, shallow loudspeaker **9** is located inside the housing of a cellular phone. Sound emerges from air passages **6**, providing greater space for a display device. Since this brief description applies equally well to a substantial body of prior art, one may wonder exactly what has been patented here. A clue is found at the end of the only independent patent claim:

7,263,391

43.38.Si ELECTRONIC EQUIPMENT

Shinya Sugiyama and Emi Suzuki, assignors to Matsushita Electric Industrial Company, Limited
28 August 2007 (Class 455/575.3); filed in Japan 18 April 2002

As cellular phones become smaller, there is less interior space for a loudspeaker and acoustical performance suffers as a result. In the



arrangement shown here, loudspeaker **7** is mounted in half-shell **1**, which has a front sound opening **6** and a rear opening **8**. The other half-shell **2** has a corresponding opening **9** to provide additional rear air volume when the case is closed.—GLA

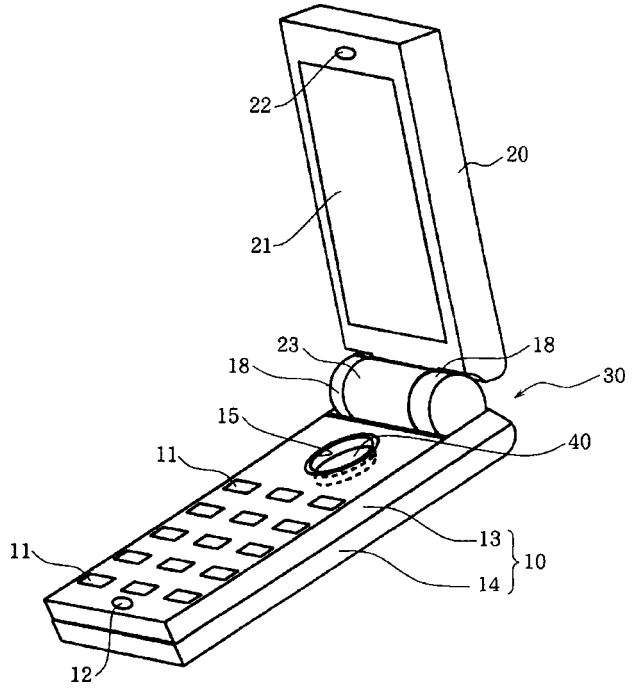
7,272,422

43.38.Si PORTABLE ELECTRONIC DEVICE

Yo Nagasawa and Michiaki Koizumi, assignors to Sanyo Electric Company, Limited

18 September 2007 (Class 455/575.1); filed in Japan 24 March 2003

This patent presents yet another solution to the problem of projecting audible sound from a cellular telephone, whether the case is open or closed.



A rotating shutter, operated by opening and closing the hinged case, directs sound through an appropriate opening in either the front or rear surface.—GLA

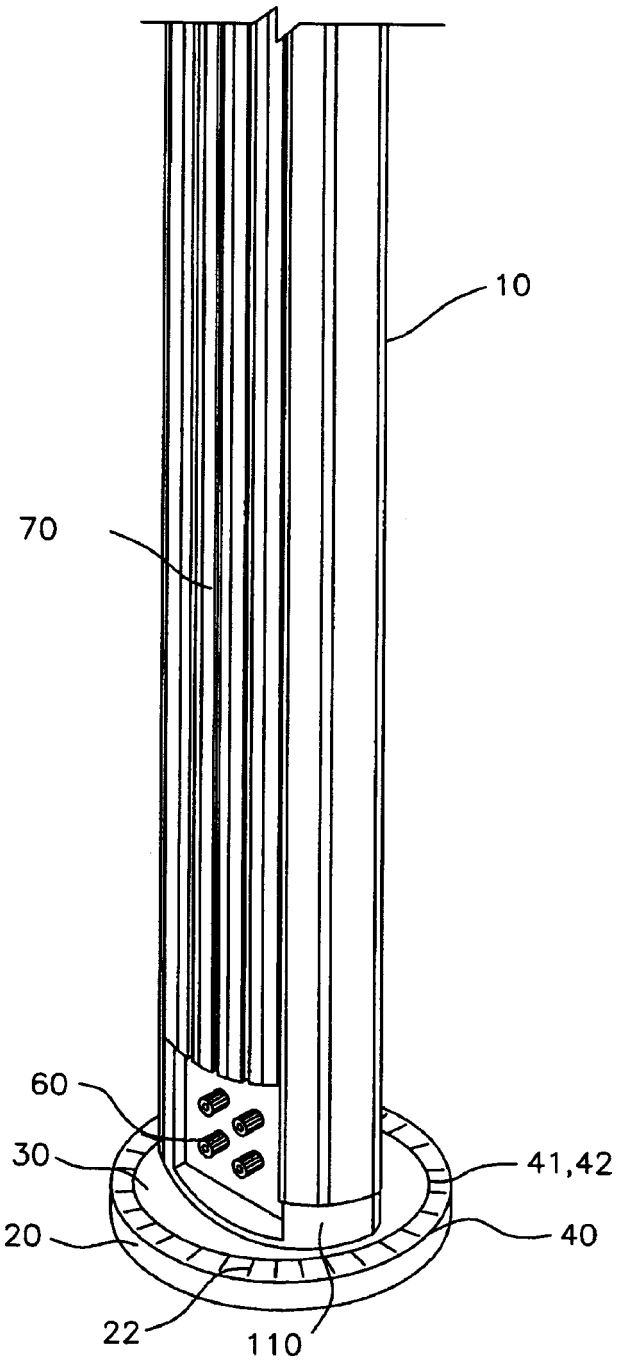
7,237,648

43.38.Tj SURROUND SOUND POSITIONING TOWER SYSTEM AND METHOD

Noel Lee and Demian Martin, assignors to Monster Cable Products, Incorporated

3 July 2007 (Class 181/199); filed 7 January 2004

By using a special base plate **30**, a tape measure, a rope, and defining a reference line between any two surround sound loudspeaker systems **10**, a means is disclosed for precisely aligning at least one system, and realigning of the said system if it becomes "disoriented, e.g., by a seismic event." It is good to know that subsequent to such a disorienting event, the system can be precisely and repeatably "disposed" to a uniform distance and alignment from the listener.—NAS

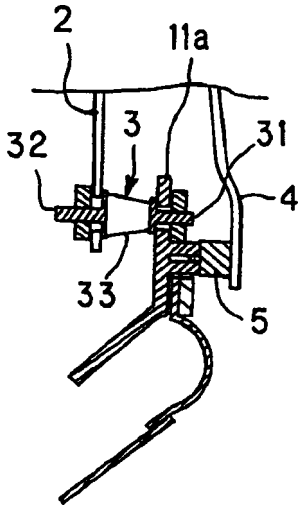


7,239,715

43.38.Ja SPEAKER APPARATUS TO BE MOUNTED ON A VEHICLE

Koji Maekawa *et al.*, assignors to Pioneer Corporation
3 July 2007 (Class 381/389); filed in Japan 25 June 2003

To prevent unwanted sound from being radiated from an automobile door frame **11a**, resilient mount **3** is claimed that replaces solid hold-down



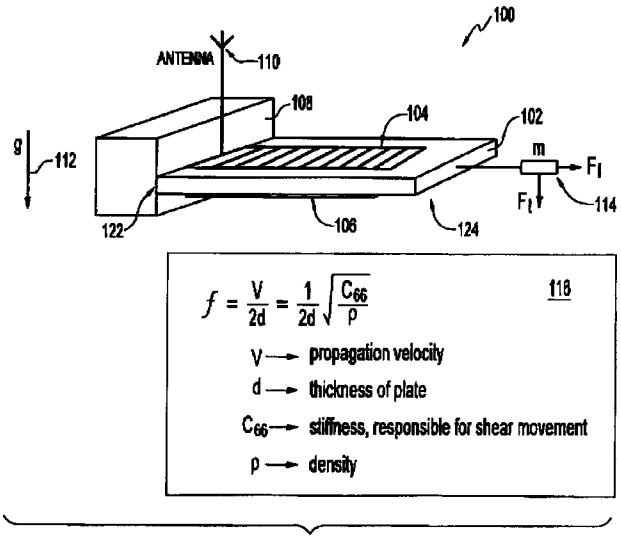
screws for an automobile door mounted loudspeaker **4** to prevent audio frequency vibrations from entering door frame **11a**.—AJC

7,243,544

43.38.Rh PASSIVE AND WIRELESS ACOUSTIC WAVE ACCELEROMETER

James Z. T. Liu and Aziz Rahman, assignors to Honeywell International, Incorporated
17 July 2007 (Class 73/514.28); filed 16 June 2005

Accelerometer sensor **100** is claimed where cantilevered SAW substrate **124** has mass *m* attached to its outer end **102**. Acceleration **112** causes a force Ft on *m*, in turn causing substrate **124** to bend, creating surface strains that results in a change of phase of the surface acoustic wave propagated on the top side **104** versus the wave propagated on bottom side **106**.



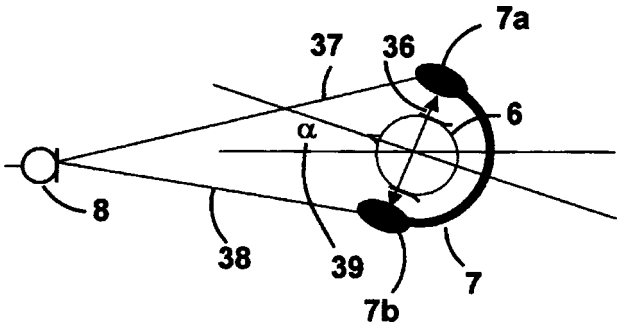
Antenna **110** can receive RF energy from an external interrogator (not shown) that serves to excite the SAWs and to read out the shifted phases, processing those into acceleration data.—AJC

7,272,073

43.38.Vk METHOD AND DEVICE FOR GENERATING INFORMATION RELATING TO THE RELATIVE POSITION OF A SET OF AT LEAST THREE ACOUSTIC TRANSDUCERS

Renato Pellegrini and Matthias Rosenthal, assignors to SonicEmotion AG
18 September 2007 (Class 367/124); filed in Switzerland 27 May 2002

This patent relates to virtual reality systems and the like. The diagram is a top view of a listener **6** wearing headphones **7**. Left and right audio signals to the headphones also include ultrasonic signals encoded to provide identifiable left and right tracking information. The airborne signals (from headphones?) are picked up by one or more microphones **8**. Assuming that



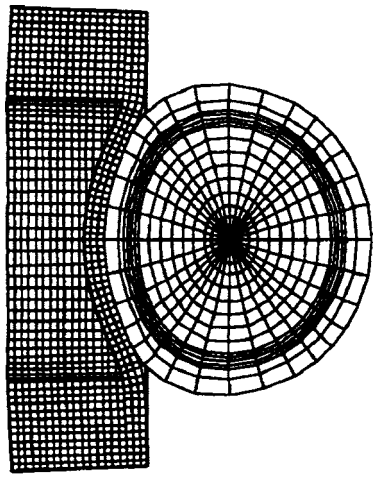
the encoded ultrasonic signals can be recovered accurately, it is apparent that path lengths from microphone **8** to left and right headphones **7a** and **7b** can be calculated, determining both the location of the listener and his head rotation.—GLA

7,241,230

43.40.At GOLF CLUB HEAD AND METHOD OF MAKING THE SAME

Masaya Tsunoda, assignor to SRI Sports Limited
10 July 2007 (Class 473/324); filed in Japan 6 August 2002

The patent discusses in a pretty straightforward manner the characteristics of golf club driver head design and golf balls to determine a means of improving the rebound performance of the club to increase the transfer of



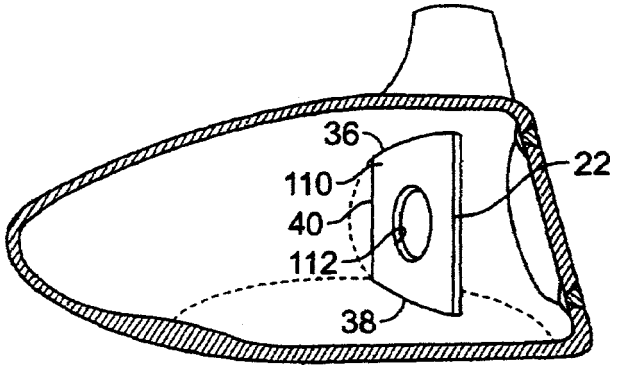
energy to the ball. Several spring mass models of a golf ball are presented. A means of measuring the impact response of the club using the impact hammer method is also disclosed.—NAS

7,247,103

43.40.At GOLF CLUB HEAD PROVIDING ENHANCED ACOUSTICS

Todd P. Beach *et al.*, assignors to Taylor Made Golf Company, Incorporated
24 July 2007 (Class 473/324); filed 5 June 2006

Some improvements in club head design that have improved the coefficient of restitution and the overall performance of the club have also changed the sonic and haptic characteristics of the club. Many golfers dislike the aural and tactile feedback characteristics that accompany these



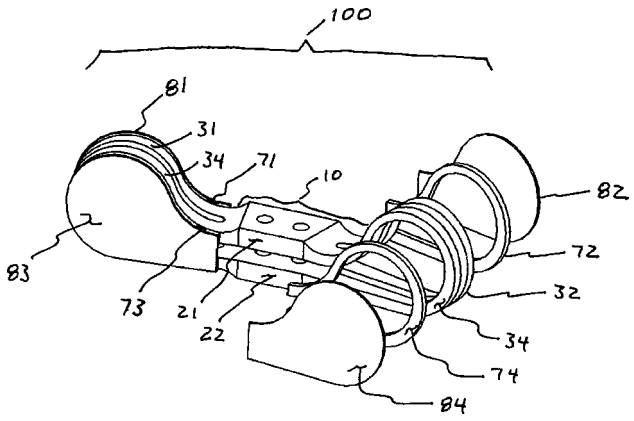
improvements, as their game may not have improved as much. To improve the golfer's aural and tactile experience, stiffener **22**, which can take various shapes, is attached to the interior of the club head on the shaft side.—NAS

7,249,756

43.40.Tm LOW-PROFILE, MULTIAXIS, HIGHLY PASSIVELY DAMPED, VIBRATION ISOLATION MOUNT

Paul S. Wilke and Conor D. Johnson, assignors to CSA Engineering, Incorporated
31 July 2007 (Class 267/152); filed 1 February 2006

This vibration isolator, a group of which is intended to be used between a cylindrical vibration source and a payload, for example, has loop-shaped resilient elements **32** and **34** that connect the support attachment **22**



to the payload attachment **21**. Loop-shaped elements **72** and **74** of viscoelastic material are sandwiched between the resilient elements and covers **82** and **84** to provide damping.—EEU

7,250,224

43.40.Tm COATING SYSTEM AND METHOD FOR VIBRATIONAL DAMPING OF GAS TURBINE ENGINE AIRFOILS

Ramgopal Darolia *et al.*, assignors to General Electric Company
31 July 2007 (Class 428/673); filed 12 October 2004

The surface of a turbine airfoil is plasma-cleaned, and then a metallic coating is applied to the airfoil. Thereafter a ceramic coating is applied atop the metallic coating. The metallic coating, which may have a composition that contains silver or tin, apparently acts like the viscoelastic layer in a constrained layer damping arrangement.—EEU

7,255,196

43.40.Tm WINDSHIELD AND SOUND-BARRIER FOR SEISMIC SENSORS

William B. Coney and Peter A. Krumhansl, assignors to BBN Technologies Corporation
14 August 2007 (Class 181/122); filed 14 November 2003

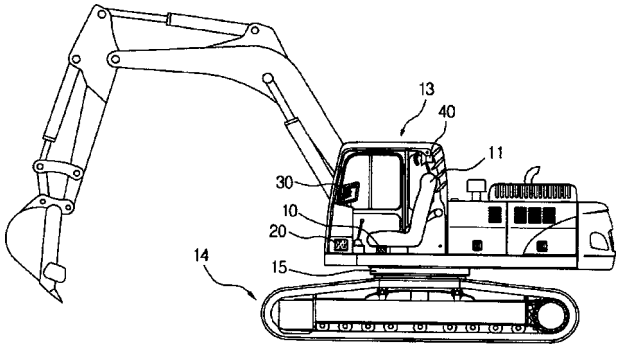
To reduce ambient (wind) noise a dome is placed over seismic sensors.—GFE

7,236,868

43.40.Yq APPARATUS FOR MONITORING OPERATOR VIBRATION FOR EARTH MOVING MACHINERY WITH OPERATION CAB

Jong Min Kang, assignor to Volvo Construction Equipment Holding Sweden AB
26 June 2007 (Class 701/50); filed in Republic of Korea 11 August 2005

Three-axis accelerometer **10** is placed in the lower part of seat **11**, which is in cab **13**. The accelerometer is connected to equipment that conditions and processes the accelerometer signal, which then determines if the vibration exceeds the exposure limit value (ELV), and sets off an alarm



when the ELV is exceeded. The processing system also calculates the ELV based on mean work time per day, the maximum vibration, and the vibration dose value.—NAS

7,255,008

43.40.Yq VIBRATION-TESTING SYSTEM

Takehiro Fukushima *et al.*, assignors to IMV Corporation
14 August 2007 (Class 73/664); filed 6 April 2006

In this shaker system, separate electrodynamic drivers produce vertical motion and horizontal motion in two perpendicular directions. In general,

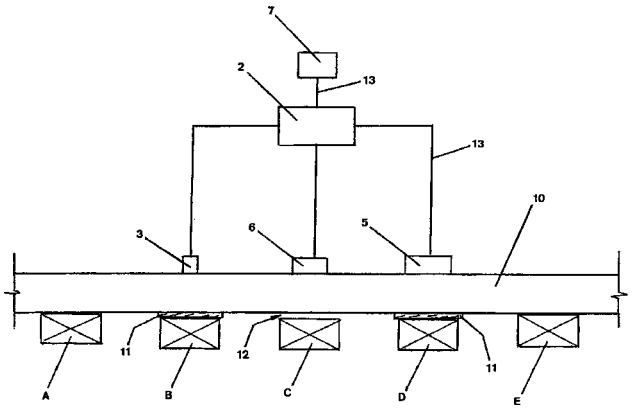
application of horizontal forces to the table that supports the test article may cause unwanted rotation of that table about a horizontal axis. Here this rotational motion is sensed and automatic adjustment of the vertical location of the axes of action of the horizontal shakers is provided so as to suppress this rotation.—EEU

7,263,886

43.40.Yq APPARATUS FOR AND METHODS OF STRESS TESTING METAL COMPONENTS

Brent Felix Jury, New Plymouth, New Zealand
4 September 2007 (Class 73/579); filed in New Zealand 20 September 2002

This patent describes an acoustical test method for detecting incipient stress fatigue in railway lines and pipelines. As shown, a section of metal rail is isolated from tie **C** by inserting temporary shims **11** on adjacent ties. The rail is excited by adjustable shaker **3** and the resulting vibrations are



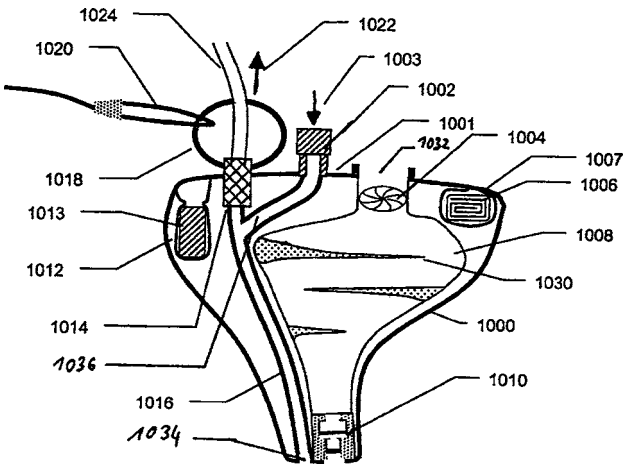
sensed by accelerometer **5**. Temperature is recorded by sensor **6**. The resulting information is then plotted by a computer program that also controls the test sequence. The results can be used to determine whether the line is under tension or compression.—GLA

7,240,765

43.50.Hg CUSTOMIZED HEARING PROTECTION EARPLUG WITH AN ACOUSTIC FILTER AND METHOD FOR MANUFACTURING THE SAME

Christian Berg and Hilmar Meier, assignors to Phonak AG
10 July 2007 (Class 181/135); filed 25 August 2004

This patent describes a "Swiss army knife" of a passive hearing protector. The made-to-measure hard shell encloses **1010** an acoustic filter, **1008** a chamber to enhance speech frequencies, **1030** a structure to provide tuning, and **1004** a mechanical peak clipper. The HPD attenuation can be bypassed for communication by opening a button **1002** and the effective noise reduction can be measured through removable tube **1024**.



Finally, there is an RFID tag **1006** to track compliance of users and a metal object **1013** to identify the HPD as a foreign object in industries such as food processing.—JE

7,245,735

43.50.Hg EARMUFF STRUCTURE FOR HEADSET OR EAR PROTECTOR

David Han, Taipei, Taiwan
17 July 2007 (Class 381/371); filed 2 April 2004

Two molded shells comprise the inner and outer halves of an earmuff. The outer shell contains an integrally molded shelf providing a support for affixing a circuit board with loudspeaker.—JE

7,248,706

43.50.Hg AUDIO-SIGNAL DETECTING DEVICE

Ching Kuo Chuang and Wugu Hsiang, Taipei County, Taiwan
24 July 2007 (Class 381/72); filed 29 October 2003

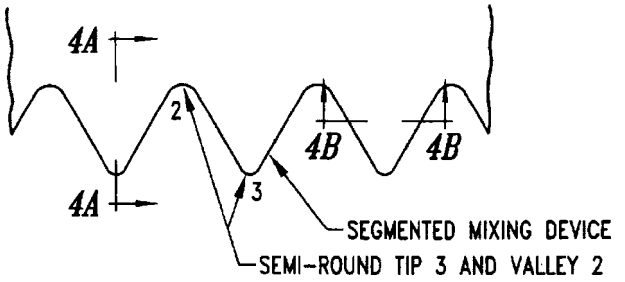
This communication system to be installed in a muff-type hearing protector incorporates circuitry to interrupt output of signals exceeding a threshold by grounding the input to the amplifier, driving the output transducer.—JE

6,786,037

43.50.Nm SEGMENTED MIXING DEVICE HAVING CHEVRONS FOR EXHAUST NOISE REDUCTION IN JET ENGINES

Ronald L. Balzer, assignor to The Boeing Company
7 September 2004 (Class 60/204); filed 23 January 2003

A jet engine exhaust nozzle with a serrated trailing edge is designed to promote mixing, thus reducing radiated noise. In contrast to similar schemes in which the chevrons have sharp corners, these chevrons have carefully



chosen radii of curvature that enhance mixing without the addition of high frequency noise.—KPS

6,813,877

43.50.Nm GAS TURBINE ENGINE EXHAUST NOZZLE HAVING A NOISE ATTENUATION DEVICE DRIVEN BY SHAPE MEMORY MATERIAL ACTUATORS

Nigel T. Birch and John R. Webster, assignors to Rolls-Royce plc
9 November 2004 (Class 60/226.1); filed in United Kingdom 3 March 2001

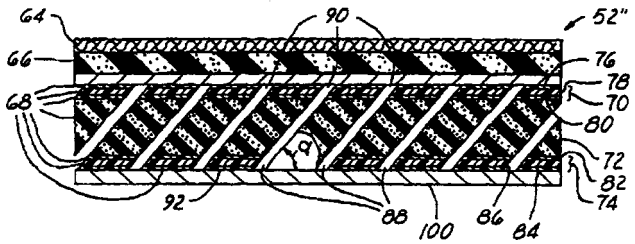
A method to reduce noise from aircraft jet exhaust flow consists of tabs arranged on the trailing edge of the jet nozzle. These devices are intended to promote mixing and thus reduce radiated noise. In contrast to fixed-geometry chevrons (e.g., United States Patent 6,786,037), these are designed to be deployed only when needed, when the aircraft is in the vicinity of an airport. Stowing the device during the majority of the flight time minimizes the performance penalty to the aircraft that is incurred by the forced mixing of the jet exhaust. The actuation mechanism utilizes shape memory material (e.g., Nitinol) that controls the deflection of the tabs into the exhaust flow. The means of actuation may be either an applied electric field or naturally occurring thermal conditions.—KPS

6,789,646

43.50.Gf TUNABLE SOUND ABSORBING AND AIR FILTERING ATTENUATING DEVICE

Shuo Wang *et al.*, assignors to Lear Corporation
14 September 2004 (Class 181/293); filed 11 October 2002

A material system intended to absorb sound as an automobile headliner consists of numerous layers. The inner, visible fabric layer, **64**, is adjacent to an absorption layer, **66**, followed by multiple perforated layers, **68**, which are permeable and lightweight. These layers include structural layers, **70** and **74**, and substrate layers, **72**. The latter are formed by stiff open-cell foam.



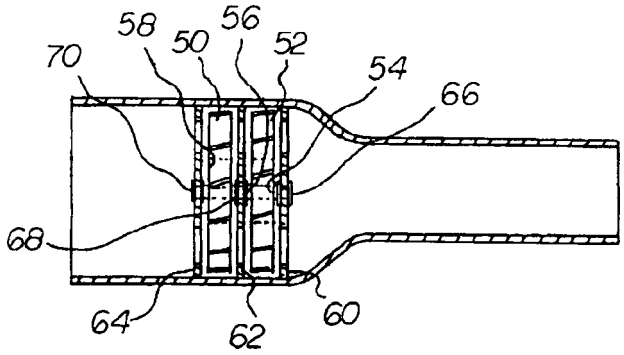
The perforations in these various layers are aligned to form resonator tubes, **88**. The resonance frequency of the tubes is determined by the tube dimensions and the angle of inclination.—KPS

6,810,992

43.50.Gf SOUND PRODUCING VEHICLE EXHAUST SYSTEM

Mario Lombardo, Treasure Island, Florida
2 November 2004 (Class 181/227); filed 19 September 2002

Judging from the title alone, one might question the need for such a device. It is attached to the tailpipe of an automobile and is designed to



create a "turbinelike sound." In essence, it is a siren (rotating plates with holes) driven by the flow of the exhaust gases.—KPS

7,182,087

43.50.Hg DUAL POSITION HEARING PROTECTION DEVICE

Robert E. Marsh, Kansas City, Missouri
27 February 2007 (Class 128/867); filed 18 October 2002

A mechanism to incorporate into a hearing protector that in one position occludes the sound path into the external ear and in the other, provides free transmission of sound through the HPD.—JE

7,185,734

43.50.Hg HEARING PROTECTION EARPLUG, USE OF SUCH AN EARPLUG AND METHOD FOR MANUFACTURING SUCH AN EARPLUG

Christoph Widmer and Christian Berg, assignors to Phonak AG
6 March 2007 (Class 181/135); filed 25 August 2004

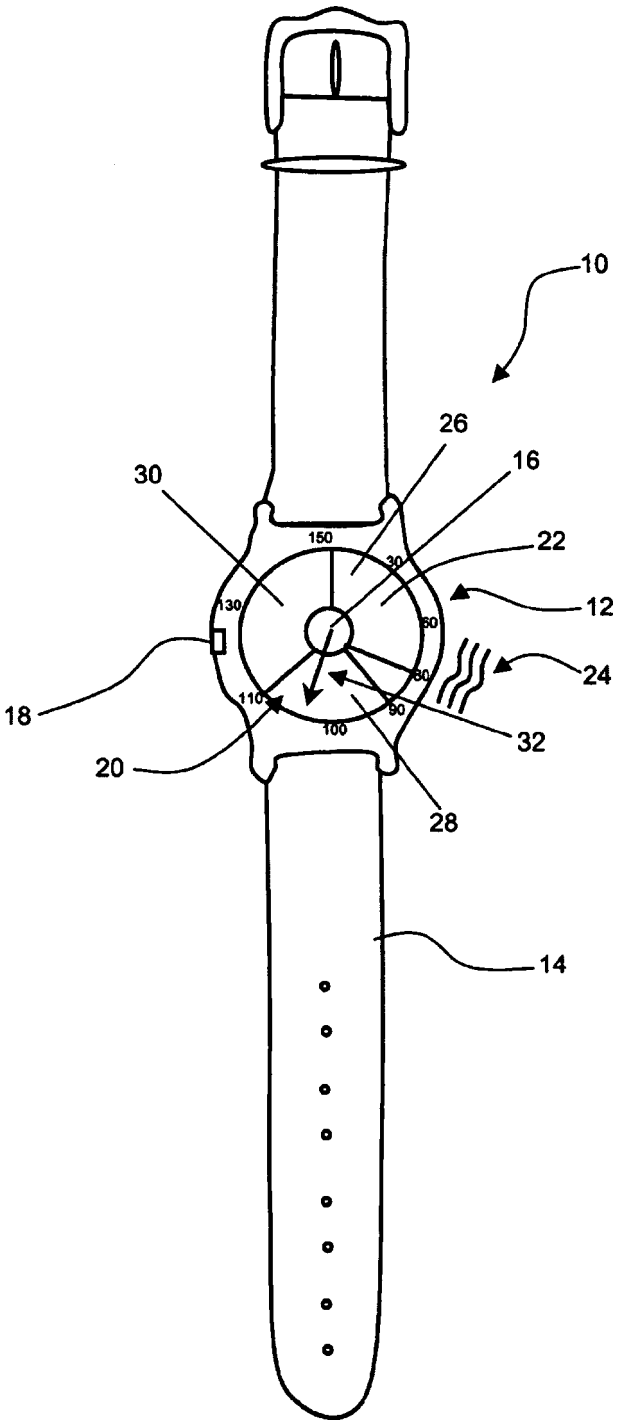
This is essentially the HPD in United States Patent 7,240,765 with the incorporation of an active device to replace the passive components of the previous patent.—JE

7,199,720

43.50.Yw HEARING PROTECTION WARNING DEVICE

Michael Shapiro, Incline Village, Nevada
3 April 2007 (Class 340/573.1); filed 19 November 2004

A wearable warning device that measures sound levels and provides a



visual and tactile warning. The tactile signal can change with the level of the sound.—JE

7,249,653

43.55.Ti ACOUSTIC ATTENUATION MATERIALS

Ping Sheng *et al.*, assignors to RSM Technologies Limited
31 July 2007 (Class 181/290); filed 3 May 2004

Acoustic barrier panels according to this patent consist of a relatively soft material sandwiched between two relatively stiff outer layers, with masses included near midthickness of the soft material. The masses, which may be in the shape of spheres, disks, or wire mesh, lead to high attenuation in the limited frequency region where there occur resonances of the elastically supported masses. Selection of the masses and of the soft material permits tuning of this frequency region.—EEU

7,266,930

43.55.Ti CONSTRUCTION BLOCK

Myles A. Fisher, assignor to US Block Windows, Incorporated
11 September 2007 (Class 52/306); filed 3 November 2003

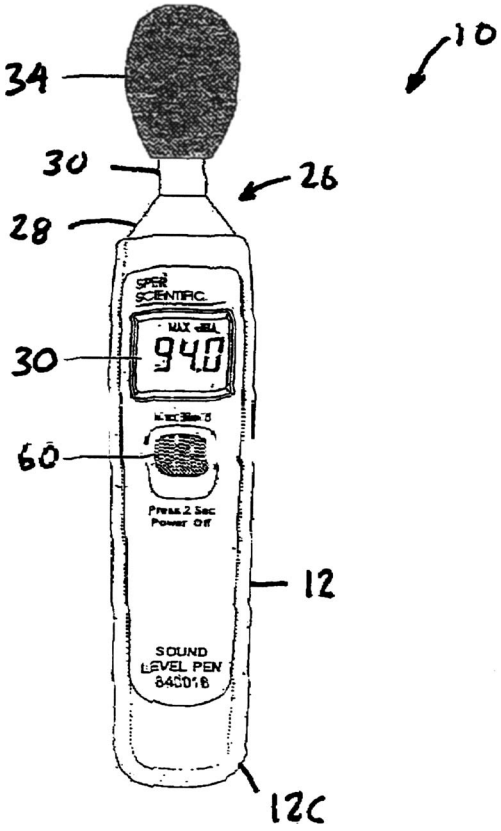
This construction block is made of glass or plastic. Thermal insulation and sound isolation are enhanced with the addition of an interior baffle or septum.—CJR

7,270,006

43.58.Fm SINGLE BUTTON OPERATING SOUND LEVEL METER AND METHOD THEREFOR

Devin Sper, assignor to Sper Scientific Limited
18 September 2007 (Class 73/646); filed 25 July 2005

A handheld sound level meter is controlled by a single button **60** that



steps through several operational functions.—GLA

7,260,227

43.60.Ac METHOD AND DEVICE FOR MEASURING SOUND WAVE PROPAGATION TIME BETWEEN LOUDSPEAKER AND MICROPHONE

Daisuke Higashihara *et al.*, assignors to Etani Electronics Company, Limited
21 August 2007 (Class 381/56); filed in Japan 9 December 2002

Ludicrously, a patent that describes using a broadband pulse to improve the correlation of travel time estimation has been granted.—GFE

7,251,336

43.60.Dh ACOUSTIC TALKER LOCALIZATION

Maziar Amiri *et al.*, assignors to Mitel Corporation
31 July 2007 (Class 381/92); filed in United Kingdom 30 June 2000

A device to enhance the localization of a speaker during a conference call is put forth that utilizes spectral conditioning with an activity detector in order to beamform on the true speaker (and not a reflection or coherent noise).—GFE

7,254,241

43.60.Dh SYSTEM AND PROCESS FOR ROBUST SOUND SOURCE LOCALIZATION

Yong Rui and Dinei Florencio, assignors to Microsoft Corporation
7 August 2007 (Class 381/92); filed 26 July 2005

Microsoft puts forth a method to improve teleconferencing via a microphone array using a one step time-of-delay-arrival system and steered beams.—GFE

7,251,345

43.60.Jn FAST FOURIER TRANSFORM CORRELATION TRACKING ALGORITHM WITH BACKGROUND CORRECTION

Ruey-Yuan Han, assignor to Lockhead Martin Corporation
31 July 2007 (Class 382/103); filed 10 June 2005

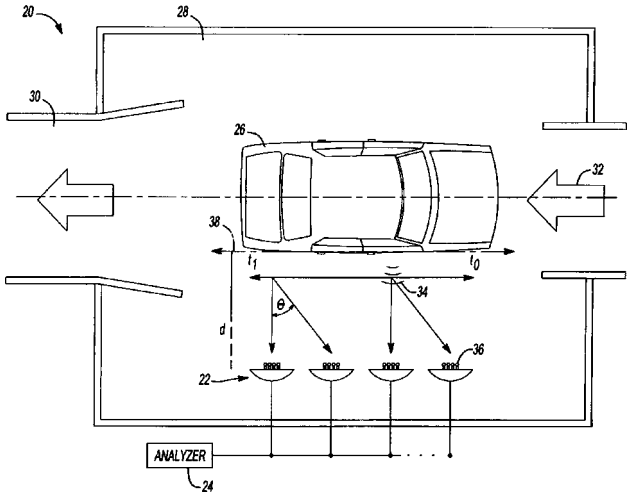
A real-time method of tracking targets is explained. A mean-squared error surface is calculated, via a fast Fourier transform with a 2D sinc function, whose minimum corresponds to the target location.—GFE

7,240,544

43.60.Rw AERODYNAMIC NOISE SOURCE MEASUREMENT SYSTEM FOR A MOTOR VEHICLE

Jean P. Mallebay–Vacqueur and Mitchell Puskarz, assignors to DaimlerChrysler Corporation
10 July 2007 (Class 73/147); filed 22 December 2004

A moving vehicle aerodynamic noise measurement system **20** is claimed where surface flow noise **34** emitted from a test vehicle **26** in a wind tunnel **30** is received by multiple microphone arrays **36** backed by reflectors **22** outside that air stream **32**. Signal processing systems **24** to calculate and display both the sound spectrum of that noise and an image of the surface noise emitter locations are also claimed.—AJC

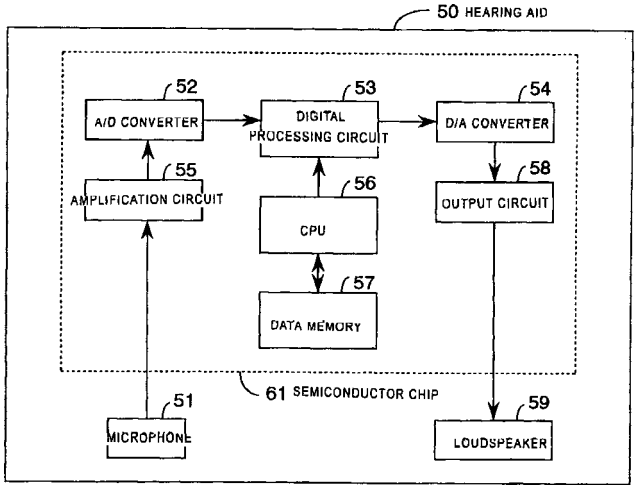


7,262,992

43.66.Ts HEARING AID

Akihide Shibata *et al.*, assignors to Sharp Kabushiki Kaisha
28 August 2007 (Class 365/185.05); filed in Japan 16 May 2003

To reduce hearing aid size and cost, both EEPROM memory and logic circuitry for signal processing are included on a single chip. The memory cells have a reduced size and hence a reduced cost due to the incorporation



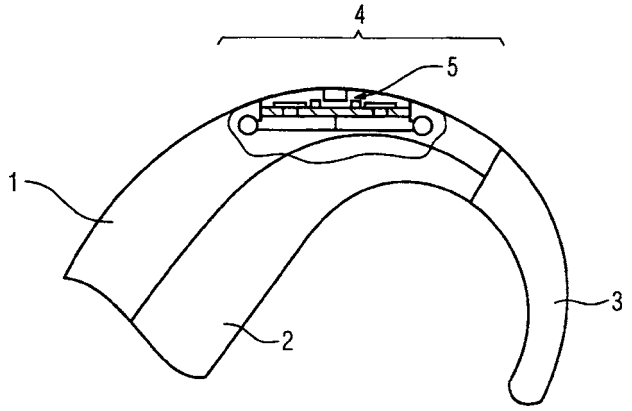
of a thin gate insulating film. Other benefits are said to include the ability to store two bits of information in each memory cell, lower memory power consumption, and shorter time for rewriting the hearing aid parameters into memory.—DAP

7,263,194

43.66.Ts HEARING DEVICE

Torsten Niederdränk and Christian Weistenhöfer, assignors to Siemens Audiologische Technik GmbH
28 August 2007 (Class 381/324); filed in Germany 18 September 2003

Hearing aid microphones are sometimes packaged in a separate housing from the main portion of a behind-the-ear hearing aid to help prevent vibratory-induced feedback oscillation. This type of design results in larger hearing aids. To reduce the overall hearing aid size, silicon microphones,



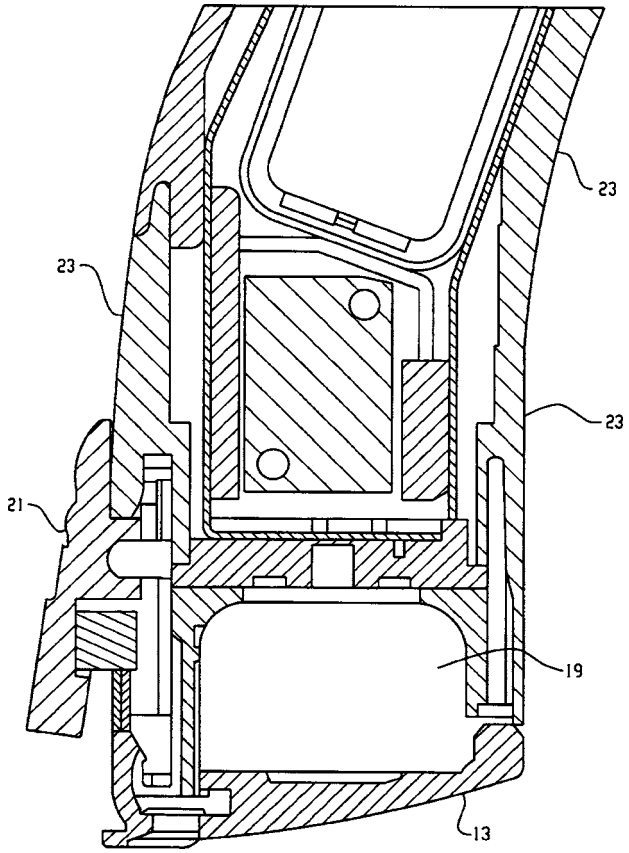
which have less sensitivity to body vibrations, are utilized. These microphones may be packaged together with the hearing aid signal processing circuitry on a single board.—DAP

7,267,847

43.66.Ts HYDROPHOBIC COATING OF INDIVIDUAL COMPONENTS OF HEARING AID DEVICES

Erdal Karamuk, assignor to Phonak AG
11 September 2007 (Class 427/569); filed 30 December 2003

A hydrophobic coating is applied to seal small openings within hearing aid housings, such as battery compartments, from moisture ingress which



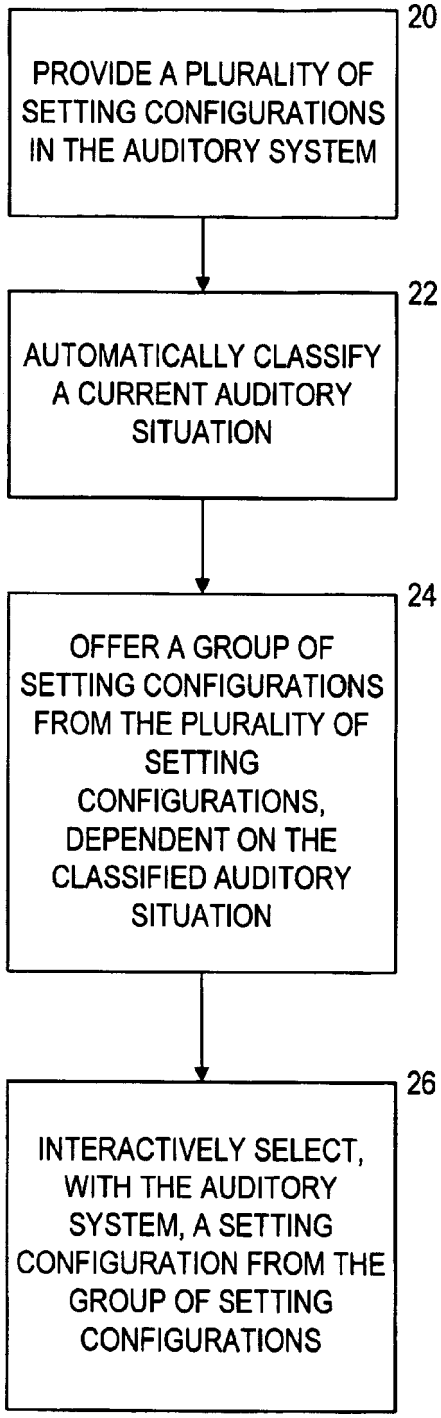
may produce corrosion. The coating allows air exchange with the environment while preventing liquids from entering the hearing aids.—DAP

7,269,269

43.66.Ts METHOD TO ADJUST AN AUDITORY SYSTEM AND CORRESPONDING AUDITORY SYSTEM

Volkmar Hamacher and Matthias Wesselkamp, assignors to Siemens Audiologische Technik GmbH
11 September 2007 (Class 381/314); filed in Germany 27 February 2003

In order to optimize hearing aid settings for auditory situations encountered by the hearing aid wearer, acoustic environments are first automatically classified and then used to automatically select a group of appropriate hearing aid setting configurations for each auditory situation. A



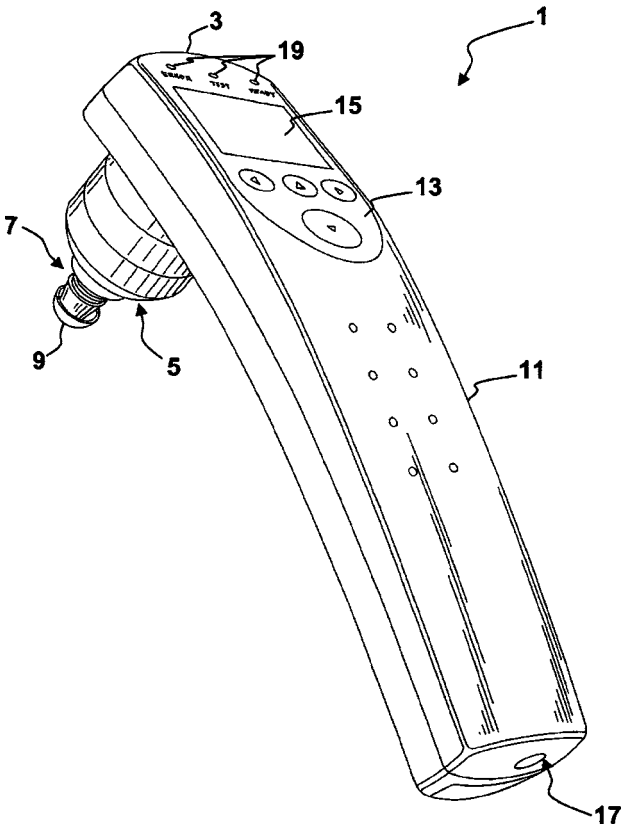
user-operated input device, such as a remote control, facilitates an interactive selection of a single hearing aid setting configuration from the selected group.—DAP

7,269,262

43.66.Yw HEARING TEST APPARATUS AND METHOD HAVING AUTOMATIC STARTING FUNCTIONALITY

Steven J. Iseberg *et al.*, assignors to Etymotic Research, Incorporated
11 September 2007 (Class 381/60); filed 18 May 2006

To determine if a testing probe used with an otoacoustic emission tester has been positioned properly, the system senses and provides feedback



to the user via tones pre-test conditions such as whether the probe position is temporally stable and whether an adequate acoustic seal in the ear canal has to the user via tones pretest conditions such as whether the probe position is been achieved. When at least one pretest condition has been met, the processor automatically starts a hearing test without any operator input.—DAP

7,260,213

43.72.-p TELEPHONE ECHO CANCELLER

Alexander Stenger, assignor to NXP B.V.
21 August 2007 (Class 379/406.07); filed in the European Patent Office 8 January 2003

Another echosuppression technique for telephony is put forth. This one specifically is designed to reduce nonlinear echo feedback.—GFE

7,260,231

43.72.-p MULTI-CHANNEL AUDIO PANEL

Donald Scott Wedge, Santa Cruz, California
21 August 2007 (Class 381/310); filed 26 May 1999

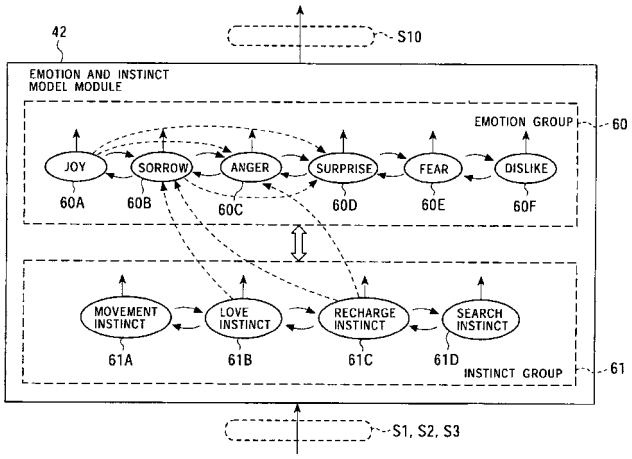
The system described provides comprehension queues to otherwise monochannel acoustic signals. For instance, you hear the voice of a person flying to your left in the left ear of your headphones.—GFE

7,251,606

43.72.Ja ROBOT DEVICE WITH CHANGING DIALOGUE AND CONTROL METHOD THEREFOR AND STORAGE MEDIUM

Rika Horinaka *et al.*, assignors to Sony Corporation
31 July 2007 (Class 704/272); filed in Japan 27 March 2001

This robotic pet toy device includes a speech synthesizer and software to construct sentences representing the internal status of the device. Since the device also includes a speech recognizer and a dialogue management system, the range of possible internal states may include not only the obvious issues of the state of battery charge and operating conditions, but may also include responses to commands, statements, or questions by the user, as



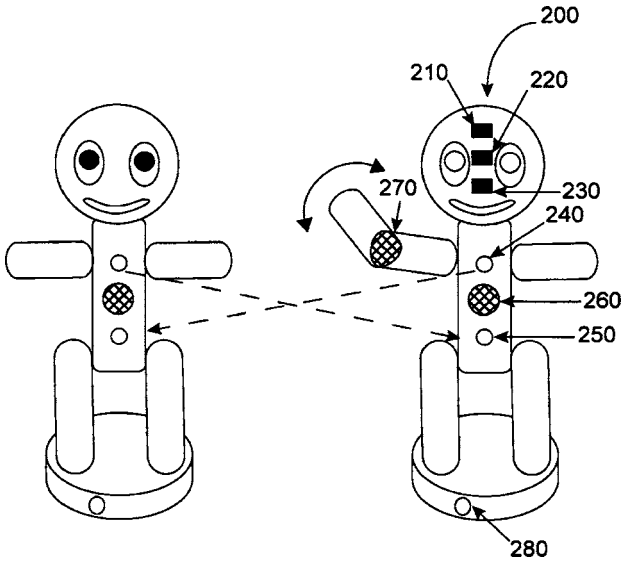
well as responses to actions taken by the user, such as petting, feeding, kicking, or otherwise interacting with the device. The device's internal state management system makes distinctions among the various possible causes for internal state changes, expressing the resulting states in terms more appropriate to human emotions than to the actual machine status as might be expressed in engineering terms.—DLR

7,252,572

43.72.Ja FIGURINES HAVING INTERACTIVE COMMUNICATION

Will Wright *et al.*, assignors to Stupid Fun Club, LLC
7 August 2007 (Class 446/484); filed 12 May 2004

This is a set of toy figurines, each of which may include a speech synthesizer. When several of the figurines are placed near each other, they interact by an infrared channel to form a network. Once a network has been formed, the figurines can "formulate behaviors based on attributes, requests, and actions of the other figurines within the network." Each figurine is said



to have a "personality" and a "current psychological state" based on stored tables containing information on past and current actions of itself and other figurines in the network. There is no mention of speech recognition capability.—DLR

7,254,539

43.72.Ne BIDIRECTIONAL NATURAL LANGUAGE SYSTEM FOR INTERFACING WITH MULTIPLE BACK-END APPLICATIONS

Martin D. Carberry *et al.*, assignors to International Business Machines Corporation
7 August 2007 (Class 704/257); filed 24 June 2002

This is a fairly elaborate network-based speech interaction system intended to perform a variety of natural language interactions for multiple clients on the network. Although the patent summary and claims sections describe the system in quite general terms, the detailed descriptions cover only a voice-operated airline reservation system. Although, in principle, the grammatical and semantic structures would be sufficiently general to allow servicing of other types of applications, this cannot really be determined from the present description.—DLR

7,260,519

43.72.Ne SYSTEMS AND METHODS FOR DYNAMICALLY DETERMINING THE ATTITUDE OF A NATURAL LANGUAGE SPEAKER

Livia Polanyi *et al.*, assignors to Fuji Xerox Company, Limited
21 August 2007 (Class 704/9); filed 13 March 2003

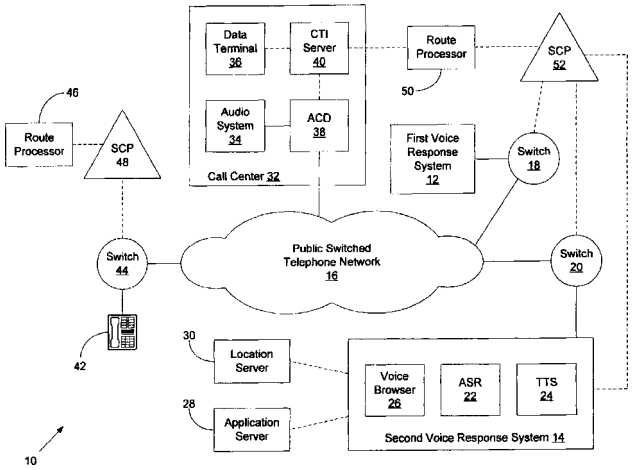
This speech recognition system is organized so as to be able to determine the attitude of a human speaker toward the subject being discussed. The methods for making this determination would include the analysis of the choice of words used by the speaker, as well as analysis of the prosodic contents of the speech signal. The patent includes extensive analysis of the choice of lexical items used in an utterance and what those choices might say about the attitude of the speaker. A shorter section describes the analysis of pitch accents and how these might relate to the speaker's intentions.—DLR

7,263,177

43.72.Ne METHOD AND SYSTEM FOR OPERATING INTERACTIVE VOICE RESPONSE SYSTEMS TANDEM

Thomas L. Paterik *et al.*, assignors to Sprint Spectrum L.P.
28 August 2007 (Class 379/88.18); filed 9 July 2004

A first voice response system receives a call, interacts with the caller, and when a predetermined condition is met, forwards the call and at least one call identifier number through a telecommunication system to a second



voice response system. Based on the call identifier(s), a second voice response system selects and interacts with at least one appropriate media resource, such as an automatic speech recognition engine and a text-to-speech converter.—DAP

7,263,483

43.72.Ne USB DICTATION DEVICE

David E. Pearah and Francis Chen, assignors to Dictaphone Corporation
28 August 2007 (Class 704/235); filed 28 April 2003

This patent introduces a handheld device for digitally recording audio input for the purpose of dictation. The device is equipped with superior audio quality microphone, USB interface with computer, and ergonomically convenient way to navigate through forms. The recorded audio is processed by a speech recognition engine for generating text.—AAD

7,272,558

43.72.Ne SPEECH RECOGNITION TRAINING METHOD FOR AUDIO AND VIDEO FILE INDEXING ON A SEARCH ENGINE

Pascal Soucy *et al.*, assignors to Coveo Solutions, Incorporated
18 September 2007 (Class 704/235); filed 23 March 2007

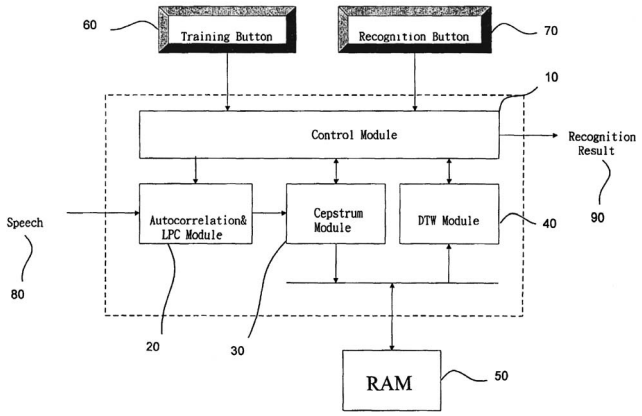
A search engine retrieves textual material from training documents contained in a source of training documents and indexes the content. A speech recognition profile is then trained using the indexed text. A search engine retrieves audio/video documents, extracts the content, and converts the associated audio into transcriptions using the trained speech recognition profile. A search engine indexes and saves the transcriptions. The number of sentences used for training is determined by comparing the number of summary sentences to a threshold.—DAP

7,266,496

43.72.Ne SPEECH RECOGNITION SYSTEM

Jhing-Fa Wang *et al.*, assignors to National Cheng–Kung University
4 September 2007 (Class 704/241); filed in Taiwan 25 December 2001

A modular speech recognition system for use in portable systems is implemented on a single application specific integrated circuit (ASIC) to enhance execution speed and lower cost. Modules include a control signal unit to sense whether a training or recognition button has been pressed, an



autocorrelation coefficient and linear predictive coefficient (LPC) unit, a cepstrum unit, and a dynamic timing warping (DTW) unit to output the recognition result. An external random access memory (RAM) stores the cepstrum of the training speech data.—DAP

7,261,694

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
28 August 2007 (Class 600/443); filed 10 November 2003

The deformation of tissue is described by a calculation of strain. The calculation is made from strain-rate estimates by using consecutive image frames that span an interval triggered, for example, by an R-wave in an ECG trace. The rate of strain is estimated from an analysis of Doppler shift caused by motion of the tissue.—RCW

7,261,695

43.80.Vj TRIGGER EXTRACTION FROM ULTRASOUND DOPPLER SIGNALS

Svein Brekke and Hans Garmann Torp, assignors to General Electric Company
28 August 2007 (Class 600/443); filed 9 March 2004

Ultrasound echo signals are processed to extract a trigger. The trigger is used to synchronize displayed images in a manner like that in which an ECG signal is used to synchronize displays of heart motion. The display may also use a variable delay relative to the trigger event.—RCW

7,263,616

43.80.Vj ULTRASOUND IMAGING SYSTEM HAVING COMPUTER VIRUS PROTECTION

Charles Cameron Brackett, assignor to GE Medical Systems Global Technology Company, LLC
28 August 2007 (Class 713/188); filed 22 September 2000

Files entering an ultrasound imaging system from a hard disk or network port are scanned to detect the possible presence of a virus. Also, each time a new process is started in the imaging system, the process is scanned for identifying information. If the possibility of a virus exists, the system operator is warned.—RCW

7,264,592

43.80.Vj SCANNING DEVICES FOR THREE-DIMENSIONAL ULTRASOUND MAMMOGRAPHY

Ramez E. N. Shehada, assignor to Alfred E. Mann Institute for Biomedical Engineering at the University of Southern California
4 September 2007 (Class 600/444); filed 27 June 2003

This scanner consists of a stationary chamber that holds fluid, a moveable chamber that is within the stationary chamber and also holds fluid, and a breast scanning apparatus. Transducer housing configurations in the scanner are described. Also described are filling and draining mechanisms, leakage protection, means to reduce turbulence, and control systems.—RCW

7,267,650

43.80.Vj ULTRASOUND DIRECTED GUIDING CATHETER SYSTEM AND METHOD

Mina Chow and William E. Webler, assignors to Cardiac Pacemakers, Incorporated
11 September 2007 (Class 600/467); filed 16 December 2002

This catheter system employs a flexible shaft with a bend at the distal end. An ultrasound transducer with a field of view directed toward the distal tip of the catheter is mounted near the bend. The catheter includes an open lumen into which a smaller catheter or guide wire can be inserted.—RCW

7,272,762

43.80.Vj METHOD AND APPARATUS FOR TESTING AN ULTRASOUND SYSTEM

Michael J. Horwath *et al.*, assignors to General Electric Company
18 September 2007 (Class 714/727); filed 16 June 2005

Circuit boards in this ultrasound system are tested by a controller that accesses profiles of board performance.—RCW

LETTERS TO THE EDITOR

This Letters section is for publishing (a) brief acoustical research or applied acoustical reports, (b) comments on articles or letters previously published in this Journal, and (c) a reply by the article author to criticism by the Letter author in (b). Extensive reports should be submitted as articles, not in a letter series. Letters are peer-reviewed on the same basis as articles, but usually require less review time before acceptance. Letters cannot exceed four printed pages (approximately 3000–4000 words) including figures, tables, references, and a required abstract of about 100 words.

Stresses and displacements for some Rayleigh-type surface acoustic waves propagating on an anisotropic half space (L)

Daniel Royer[a)]
Laboratoire Ondes et Acoustique, Université Paris 7, CNRS, ESPCI, 10 rue Vauquelin 75231 Paris, France

(Received 19 June 2007; revised 13 November 2007; accepted 13 November 2007)

Specific relations between mechanical displacements and stresses for Rayleigh-type surface acoustic waves propagating on an anisotropic half space are demonstrated. For 16 symmetry configurations belonging to the orthorhombic, tetragonal, hexagonal and cubic systems, involving only two displacement and stress components, it is shown that the ratio between the shear and normal stresses inside the propagation media is equal to the ratio between the normal and in-plane displacement components at the free surface. This result generalizes the previous one obtained in the case of an isotropic solid [W. Hassan and P. B. Nagy, J. Acoust. Soc. Am. **104**, 3107–3110 (1998)].
 [DOI: 10.1121/1.2821978]

PACS number(s): 43.35.Pt [LLT] Pages: 599–601

I. INTRODUCTION

Since their discovery in 1885, Rayleigh waves propagating on the free surface of an elastic half space have been extensively investigated and are discussed in many textbooks.[1–3] In an isotropic solid, only shear and normal stresses are produced by a Rayleigh wave and the elliptical particle trajectory is located in the sagittal plane. Although secular equations giving the speed of surface acoustic waves (SAWs) have been derived for general anisotropy,[4] explicit expressions for the stress and displacement fields cannot be obtained, excepted in particular cases. The difficulty is that the three displacement components have to satisfy three coupled equations of propagation and three boundary conditions at the traction free surface. Then, powerful numerical techniques have been developed.[5] The most advanced methods used the integral formalism developed by Barnett and Lothe.[6] Solving this problem becomes easier if the mechanical displacement having only two components must satisfy only two boundary conditions. As pointed out by Royer and Dieulesaint, these requirements, needed for the existence of a two-component surface acoustic wave, are satisfied by 16 configurations in crystals belonging to the orthorhombic, tetragonal, hexagonal and cubic symmetry systems.[7] Recently more complex cases, including piezoelectric SAW, have been solved.[8]

In a letter,[9] Hassan and Nagy showed that for an isotropic solid the ratio between the shear and normal stresses is independent of the depth and is identically the same as the ratio between the normal and in-plane displacement components at the free surface. In this letter, we demonstrate this property for two-component surface acoustic waves propagating on an anisotropic half space belonging to the previous 16 configurations.

II. MECHANICAL DISPLACEMENTS AND STRESSES

First, we establish that the ratio between the shear and normal stresses is independent of the penetration depth in the crystal. Second, we demonstrate that this ratio is equal to the ratio between the normal and in-plane displacement components, i.e., the aspect ratio of the elliptical trajectory, on the free surface.

In an elastic solid the stress tensor components T_{ij} are related to the particle displacement components u_i by the stiffness tensor c_{ijkl} referred to the Cartesian axes x_i $(i=1,2,3)$

$$T_{ij} = c_{ijkl}\frac{\partial u_l}{\partial x_k} \quad i,j,k,l = 1,2,3. \tag{1}$$

The crystal belonging to the orthorhombic symmetry system is cut along the symmetry plane $x_2=0$ and occupies the half-space $x_2>0$. A mechanical displacement of the form

$$u_i = a_i \exp(ikqx_2)\exp ik(x_1 - Vt) \quad \text{with} \quad \mathrm{Im}[q] > 0 \tag{2}$$

describes a wave which propagates with a phase velocity V and a wave number k along the direction x_1 and whose amplitude decreases with depth below the free surface $x_2=0$. Substituting this partial solution into the wave equation leads to the Christoffel equations $(i=1,2,3)$

a)Electronic mail: daniel.royer@espci.fr.

$$(\Gamma_{il} - \zeta\delta_{il})a_l = 0, \quad \text{with} \quad \zeta = \rho V^2 \tag{3}$$

and

$$\Gamma_{il} = c_{i11l} + (c_{i12l} + c_{i21l})q + c_{i22l}q^2. \tag{4}$$

Assuming that the conditions defined in Ref. 7 are fulfilled, the Christoffel equations factorize into two parts giving a shear horizontal wave polarized along x_3 and a Rayleigh-type surface acoustic wave polarized in the sagittal plane (x_1,x_2). For this latter wave, the propagation equations reduce to the system

$$\begin{bmatrix} c_{11} + c_{66}q^2 - \zeta & (c_{12} + c_{66})q \\ (c_{12} + c_{66})q & c_{66} + c_{22}q^2 - \zeta \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 0, \tag{5}$$

where $c_{11}=c_{1111}$, $c_{22}=c_{2222}$, $c_{12}=c_{1122}$ and $c_{66}=c_{1212}$.

The eigenvector $[a_1 \quad a_2]^t$ is composed of the amplitudes of the two displacement components u_1 and u_2. From Eq. (1), it can be shown that the Rayleigh wave creates only a shear stress T_{12}

$$T_{12} = ikc_{66}u_2 + c_{66}\frac{\partial u_1}{\partial x_2} \tag{6}$$

and a normal stress T_{22}

$$T_{22} = ikc_{12}u_2 + c_{22}\frac{\partial u_2}{\partial x_2}. \tag{7}$$

As shown in Ref. 7, the secular equation of system (5) is a quadratic equation in q^2. For the two acceptable solutions q_1 and q_2, which have positive imaginary parts, the eigenvectors can be chosen as $[1 \quad p_1]^t$ and $[1 \quad p_2]^t$, where p_1 and p_2 are defined as the ratio a_2/a_1 for the roots q_1 and q_2, respectively. The general solution is a linear combination of these two partial waves, propagating at the same phase velocity V

$$\begin{aligned} u_1 &= [A_1\exp(ikq_1x_2) + A_2\exp(ikq_2x_2)] \\ &\quad \times \exp ik(x_1 - Vt), \\ u_2 &= [A_1p_1\exp(ikq_1x_2) + A_2p_2\exp(ikq_2x_2)] \\ &\quad \times \exp ik(x_1 - Vt). \end{aligned} \tag{8}$$

Omitting the propagation factor $\exp ik(x_1 - Vt)$, the shear and normal stresses produced by the Rayleigh wave are given by Eqs. (6) and (7)

$$\begin{aligned} T_{12} &= ikc_{66}[(q_1 + p_1)A_1e^{iq_1kx_2} + (q_2 + p_2)A_2e^{iq_2kx_2}] \\ T_{22} &= ik[(c_{12} + c_{22}q_1p_1)A_1e^{iq_1kx_2} \\ &\quad + (c_{12} + c_{22}q_2p_2)A_2e^{iq_2kx_2}]. \end{aligned} \tag{9}$$

The weighting factors A_1 and A_2 and the velocity V are determined by the boundary conditions $T_{12}=0=T_{22}$ on the free surface $x_2=0$

$$\begin{aligned} (q_1 + p_1)A_1 &= -(q_2 + p_2)A_2, \\ (c_{12} + c_{22}q_2p_2)A_2 &= -(c_{12} + c_{22}q_1p_1)A_1. \end{aligned} \tag{10}$$

Then the two stresses have identical functional dependence on the depth x_2

$$T_{12} = -ikc_{66}(q_2 + p_2)A_2[e^{iq_1kx_2} - e^{iq_2kx_2}],$$

$$T_{22} = ik(c_{12} + c_{22}q_1p_1)A_1[e^{iq_1kx_2} - e^{iq_2kx_2}]. \tag{11}$$

In any plane parallel to the surface, the ratio between the shear and normal stresses is constant. This property is a consequence of the traction-free boundary conditions. Using the relation

$$c_{12} + c_{22}q_1p_1 = -c_{66}\frac{q_2 + p_2}{p_2}, \tag{12}$$

established in Ref. 7, the ratio between the shear and normal stresses is found to be

$$\frac{T_{12}(x_2)}{T_{22}(x_2)} = p_2\frac{A_2}{A_1}. \tag{13}$$

Taking into account the relation $p_1A_1^2=p_2A_2^2$, demonstrated in Ref. 7 [Eq. (24)], the ratio between the normal and transverse displacement components at the free surface is equal to

$$\frac{u_2(0)}{u_1(0)} = \frac{p_1A_1 + p_2A_2}{A_1 + A_2} = \frac{p_2A_2}{A_1}. \tag{14}$$

From Eqs. (13) and (14), we deduced that the stress ratio at any depth in the anisotropic solid $(x_2>0)$ is equal to the aspect ratio of the elliptical trajectory, on the free surface

$$\left.\frac{T_{12}}{T_{22}}\right|_{x_2>0} = \left.\frac{u_2}{u_1}\right|_{x_2=0}. \tag{15}$$

This result indicates that the equality between the stress ratio at any depth in the propagation media and the displacement component ratio at the free surface is not limited to isotropic solids. It generalizes the feature pointed out in Ref. 9 to specific configurations in anisotropic solids, supporting two-component Rayleigh waves for which the Christoffel equation is a biquadratic. It should be noted that the argument, based on the requirement that just below the surface the energy flow be parallel to the surface,[9] applies to these 16 anisotropic configurations. Since the rule is that the Christoffel equation is a quartic for two-component Rayleigh waves,[10] it should be interesting to investigate the general case. However, this property cannot be generalized to surface acoustic waves having three displacement components. In this case, three partial waves are required to satisfy the three stress-free boundary conditions at the surface of the crystal.

III. CONCLUSION

In this letter, we study mechanical stresses and displacements for Rayleigh-type surface acoustic waves propagating on the free surface of an anisotropic half space. We consider 16 symmetry configurations belonging to the orthorhombic, tetragonal, hexagonal and cubic systems for which the particle displacement is elliptically polarized in the sagittal plane. First, we demonstrate that shear and normal stresses have identical functional dependence versus the depth. This property is a consequence of the stress-free boundary conditions. Second, we show that, at any depth, the ratio between shear and normal stresses is equal to the ratio between the normal and in-plane displacement components at the free surface. This result generalizes the previous one obtained in the case of an isotropic solid.

[1]I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967), pp. 1–29.
[2]J. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1980), pp. 187–194.
[3]D. Royer and E. Dieulesaint, *Elastic Waves in Solids* (Springer, Berlin, 2000), Vol. **1**, pp. 276–290.
[4]T. C. T. Ting, "The polarization vector and secular equation for surface waves in an anisotropic elastic half space," Int. J. Solids Struct. **41**, 2065–2083 (2004).
[5]G. W. Farnell, Properties of Elastic Surface Waves, in *Physical Acoustics* (edited by W. P. Mason and R. N. Thurston) (Academic, New York, 1970), Vol. **6**, pp. 109–166.
[6]D. M. Barnett and J. Lothe, "Free surface (Rayleigh) waves in anisotropic elastic half spaces: The surface impedance model," Proc. R. Soc. London, Ser. A **402**, 135–152 (1985).
[7]D. Royer and E. Dieulesaint, "Rayleigh wave velocity and displacements in orthorhombic, tetragonal, hexagonal and cubic crystals," J. Acoust. Soc. Am. **76**, 1438–1444 (1984).
[8]B. Collet and M. Destrade, "Explicit secular equations for piezoacoustic surface waves: Rayleigh modes," J. Appl. Phys. **98**, 054903 (2005).
[9]W. Hassan and P. B. Nagy, "Simplified expressions for the displacements and stresses produced by Rayleigh waves," J. Acoust. Soc. Am. **104**, 3107–3110 (1998).
[10]M. Destrade and Y. B. Fu, "The speed of interfacial waves polarized in a symmetry plane," Int. J. Eng. Sci. **44**, 26–36 (2006).

Comment on "Mutual suppression in the 6 kHz region of sensitive chinchilla cochleae" [J. Acoust. Soc. Am. 121, 2805–2818 (2007)] (L)

M. A. Cheatham[a)]
Communication Sciences and Disorders, 2-240 Frances Searle Building, Northwestern University, Evanston, Illinois 60208

(Received 8 August 2007; revised 6 November 2007; accepted 6 November 2007)

Rhode [J. Acoust. Soc. Am. **121**, 2805–2818 (2007)] acknowledges that two-tone neural rate responses for low-side suppression differ from those measured in basilar membrane mechanics, making one question whether this aspect of suppression has a mechanical correlate. It is suggested here that signal coding between mechanical and neural processing stages may be responsible for the fact that the total rate response (but not the basilar membrane response) for low-frequency suppressors is smaller than that for the probe-alone condition. For example, the velocity dependence of inner hair cell (IHC) transduction, membrane/synaptic filtering and the sensitivity difference between ac and dc components of the IHC receptor potential all serve to reduce excitability for low-side suppressors at the single-unit level. Hence, basilar membrane mechanics may well be the source of low-side suppression measured in the auditory nerve.
 [DOI: 10.1121/1.2821414]

PACS number(s): 43.64.Kc, 43.64.Ld, 43.64.Nf [BLM] Pages: 602–605

A recent paper by Rhode (2007) acknowledges three differences between two-tone interactions observed in mechanical versus neural responses. His results obtained in sensitive chinchilla cochleae resolve differences relating to suppression thresholds on the tails of tuning curves (also resolved by Temchin *et al.*, 1997), as well as the rates of suppression growth. The third issue, which remains unresolved, relates to the fact that two-tone neural rate responses for low-side suppressors are lower than those for the probe alone. In contrast, "the maximum displacement of the basilar membrane for this condition is never smaller than the maximum of the unsuppressed response amplitude (Cooper, 1996; Geisler and Nuttall, 1997). This leaves undecided the mechanical correlate (if any) for rate suppression in the auditory nerve (Rhode, 2007, p. 2815)." Hence, the question remains: how can one observe decreases in the total rate response measured in the presence of suppressors well below probe frequency? Rhode suggests that a possible mechanism by which a low-frequency suppressor could be attenuated prior to activation of auditory-nerve dendrites may reflect the fact that inner hair cells (IHCs) are driven by the velocity of fluid displacement at low frequencies. Although this contention is attributed to Cheatham and Dallos (1999), the behavior was first suggested on the basis of cochlear microphonic recordings from guinea pig cochleae lacking basal outer hair cells (OHCs) due to aminoglycoside intoxication (Dallos *et al.*, 1972). The velocity dependence was subsequently confirmed by several investigators at the single IHC level (Sellick and Russell, 1980; Nuttall *et al.*, 1981; Dallos and Santos-Sacchi, 1983; Patuzzi and Yates, 1987). Unfortunately, the high-pass nature of the velocity dependence will only reduce responses to very low-frequency suppressors, making this explanation incomplete at best. Although Cheatham and Dallos (1999) acknowledged this aspect of IHC response properties in their paper dealing with *the phases of single-tone responses*, this mechanism was not suggested as a satisfactory explanation for the discrepancy between mechanical and neural measures of low-side two-tone suppression (2TS). We did, however, address this issue in other publications (Cheatham and Dallos, 1992; 1995). In order to clarify this matter, a summary of IHC results pertinent to this discussion is provided here.

Two-tone suppression was demonstrated by Sellick and Russell (1979) in basal-turn IHCs, thereby solidifying the idea that these interactions are not neural in origin. Their initial paper demonstrated that the inhibitory response area was included in the excitatory one when the ac component of the IHC receptor potential was measured (their Fig. 3a). In other words, two tones inside the excitatory response area mutually suppress one another. This result is shown here in Fig. 1(A), which represents iso-inhibitory areas for the ac receptor potential. The curve plotted with the dotted line shows the region of excitation obtained for the dc receptor potential, while the solid curve represents the ac iso-inhibitory response area recorded when the characteristic frequency (CF) probe (isolated symbol) is presented in combination with various suppressor tones. Unfortunately, it is impossible to study rate suppression in the auditory nerve if the suppressor increases rate when presented alone, i.e., when the suppressor is excitatory. Hence, suppression areas do not overlap the single-unit excitation region but merely surround it (Sachs and Kiang, 1968). This behavior was also demonstrated by Sellick and Russell (their Fig. 3b) and is represented here in Fig. 1(B) where iso-inhibitory areas are plotted for the dc component of the IHC response. Lack of overlap between excitation and suppression areas is consis-

a)Electronic mail: m-cheatham@northwestern.edu

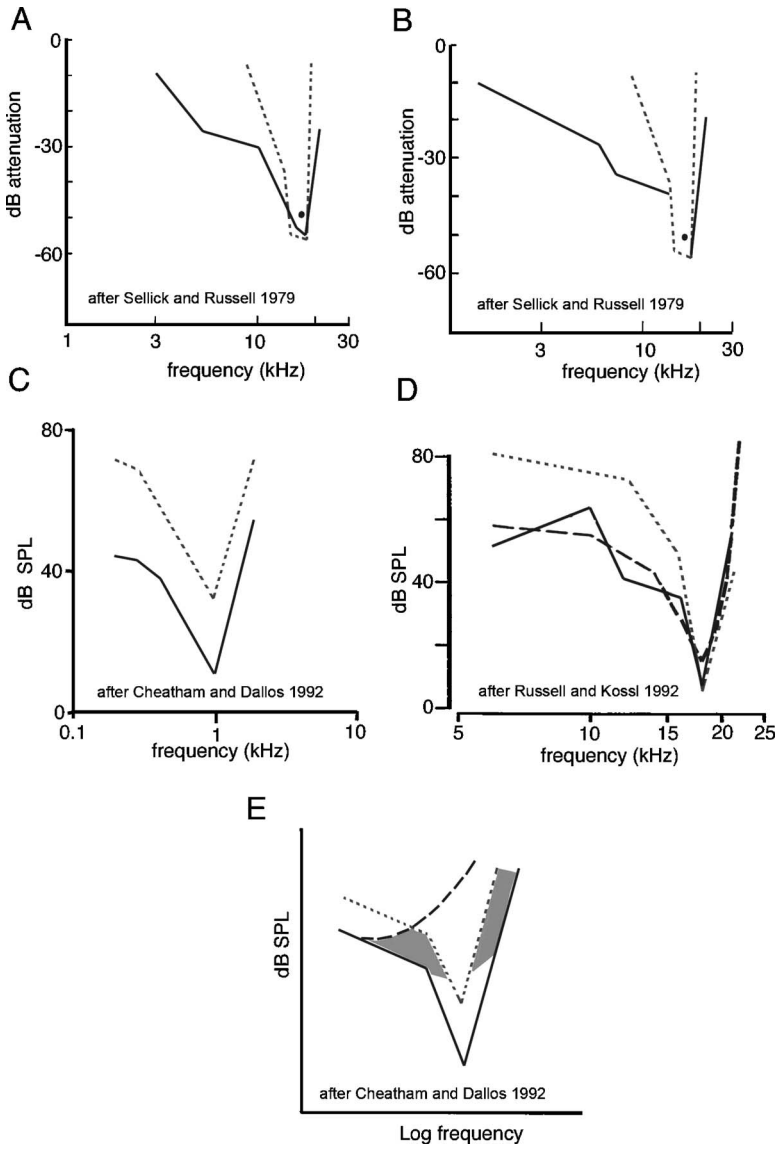


FIG. 1. Panel A provides a highly stylized version of Sellick and Russell's (1979) Fig. 3(a). The excitatory IHC response area, shown with dotted lines, reflects the dc receptor potential. The solid curve shows the iso-inhibitory ac response obtained for a CF probe, which is represented by the isolated circle. Panel B displays similar results, except that the iso-inhibitory area is measured for the dc receptor potential, again according to Sellick and Russell (Fig. 3b). The schematic in panel C represents data acquired in guinea pig from a third-turn IHC (Cheatham and Dallos, 1992, Fig. 8). The dc receptor potential for a 1 mV criterion response (dotted line) is plotted along with its companion ac receptor potential for a 1 mV peak criterion (solid line). Panel D provides similar material published by Russell and Kössl (1992, Fig. 5) from the base of the cochlea at 18 kHz, again in guinea pig. In addition to the IHC dc receptor potential (dotted line, 0.8 mV), the OHC ac response, recorded extracellularly and compensated for the time constant of the recording system (solid line, 0.5 mV), is appended along with the displacement (dashed line) response of the basilar membrane. Because the extracellular OHC response reflects the OHC's receptor current, it is not attenuated by filtering associated with the cell's basolateral membrane. Panel E provides iso-response functions for the IHC dc (dotted line) and ac receptor potentials (dashed line), as well as the compensated ac response (solid line). Shaded areas illustrate regions where rate suppression would be observed at the neural level for a "non-excitatory" suppressor, which would not generate an IHC receptor potential sufficient to induce transmitter release. This diagram is similar to those published by Cheatham and Dallos (1992, 1995).

tent with the idea that the behavior in IHCs is reflected at the single-unit level. Rate suppression produced by high-side suppressors has already been explained by assuming that the suppressor interferes with the probe tone's region of amplification (Geisler *et al.*, 1990) whose source is basal to the characteristic frequency (CF) location (Neely and Kim, 1983; Diependahl *et al.*, 1987; Geisler and Shan, 1990; Kolston *et al.*, 1990; Geisler, 1991; Zweig, 1991), which prevents high-side suppressors from increasing rate. In contrast, the nature of low-side suppression has resisted explanation (Kim, 1985). Our proposal is outlined below.

Several processes affect the voltages recorded from IHCs. As Rhode acknowledges, inputs below ~470 Hz (Sellick and Russell, 1980; Russell and Sellick, 1983; Dallos, 1984) are reduced by high-pass filtering associated with the velocity dependence of the IHC transducer. Reductions in the IHC's ac receptor potential also arise due to low-pass filtering by resistances and capacitances associated with the hair cell's basolateral membrane. Although the cutoff frequency of this first order, low-pass filter is estimated to be ~500 Hz (Russell and Sellick, 1978; 1983; Dallos, 1984; 1985; Palmer and Russell, 1986), the cutoff is probably artificially high because of the leak associated with all intracellular recordings, which reduces the input resistance of the cell (Santos-Sacchi, 1989). Further attenuation results from calcium-dependent processes associated with the synapse (−12 dB/oct above ~1100 Hz) (Palmer and Russell, 1986; Kidd and Weiss, 1990). Hence, it is ultimately the IHC's dc receptor potential that sets the single unit rate response at moderate and high stimulus frequencies (Russell and Sellick, 1978). These relationships complicate comparisons between suppression observed in basilar membrane mechanics and that recorded at the level of the auditory nerve. It is, therefore, important to document responses produced by the suppressor alone at the various stages of signal processing in the peripheral auditory system.

In addition to these attenuations, differences in sensitivity between ac and dc components of the IHC receptor potential must also be considered, as indicated here in Fig. 1(C). For a third-turn guinea-pig IHC, whose CF is 1000 Hz, one requires a ~20 dB greater input to produce a criterion dc (dotted line) versus ac (solid line) receptor potential (Dallos, 1985; Cheatham and Dallos, 1992). This sensitivity difference very likely relates to the position of the set point along

the cell's transduction function. When more channels are open at rest, the set point is located where the function has a steeper slope, with the result that the cell responds linearly to low-level inputs. In other words, a higher input level is required to generate a dc potential. A reflection of this sensitivity difference between ac and dc components of the IHC response is seen in single units where phase locking can be detected 10–20 dB below rate threshold (Johnson, 1980). At the base of the cochlea, where the degree of nonlinearity is enhanced, there is also a threshold disparity for ac and dc components but mainly for inputs well below characteristic frequency (CF), as shown in Fig. 1(D). This difference was demonstrated by Russell and Kössl (1992) who reported that the tails of IHC tuning curves for the dc receptor potential (dotted line) are ~20 dB less sensitive than those for the OHC's ac receptor potential (solid line). The latter, however, are similar to basilar membrane displacement iso-response tuning curves (dashed lines). It should also be emphasized that the sensitivity difference between ac and dc receptor potentials is noted for all inputs to IHCs with low CFs (Fig. 1(C)) but primarily for low-frequency inputs to IHCs with high CFs (Fig. 1(D)). The latter pattern is probably related to the strong CF-dependent nonlinearity, which characterizes high-frequency responses. In contrast, the nonlinearity in the apex, although CF dependent, is less robust. In other words, it takes a ~20 dB higher level to produce a dc versus an ac component at CF; while in the base, a dc is produced at near-threshold levels.

These three factors affecting IHC voltages, the velocity dependence of cilia stimulation, membrane/synaptic filtering and ac/dc sensitivity differences, impact the degree to which a low-side suppressor is excitatory at the single-unit level, as represented in Fig. 1(E). In this schematic, the IHC ac receptor potential (dashed line) is measured only at low stimulus frequencies because of filtering by the cell's basolateral membrane. It is, therefore, not a good indicator of the degree to which a tone stimulates the IHC or of the signal magnitude present at a preceding mechanical nonlinearity. However, following adjustment for these reductions, the compensated ac function (solid line) provides an estimate of the ac receptor current. It is assumed that this compensated ac function reflects the lower boundary of the region where interactions are possible between two stimulus inputs at the level of the basilar membrane. Inputs near this boundary, however, do not produce an IHC receptor potential sufficient to induce transmitter release. Thus, shaded areas represent regions where two inputs can interact but where the suppressor produces no neural excitation. This description is analogous to that of Geisler and Shan (1990) who suggested that OHC receptor currents become nonlinear at suppressor levels that do not generate significant dc receptor potentials in IHCs.

Figure 1(E) also indicates that shading does not extend into the CF region because the suppressor itself would probably be reduced by a CF probe due to mutual suppression (Hind *et al.*, 1967; Kim *et al.*, 1980; Delgutte, 1990; Geisler and Shan, 1990; Cheatham and Dallos, 1992; Nuttall and Dolan, 1993). Therefore, the effect of a suppressor can be decreased or eliminated when it is near CF and when it is presented at a level commensurate with that of a CF probe. This behavior was in fact demonstrated by Rhode (2007, Fig. 2) in basilar membrane responses when suppressor frequency was near that of the CF probe. Because single-unit tuning curves with moderate and high CFs reflect the general character of the dc receptor potential produced by high-CF IHCs (Russell and Sellick, 1978), shaded regions approximate the 2TS areas that emerge above and below CF at the neural level for "non-excitatory" suppressors. It is also known that auditory nerve fibers exhibit broader low-side suppression areas as CF increases (Arthur *et al.*, 1971; Harris, 1979; Costalupes *et al.*, 1987). We propose that low-side suppression areas extend down to a low-frequency limit determined by the intersection of the high-frequency slope of the basolateral membrane filter and the tail of the single unit tuning curve. As CF increases, wider low-side suppression areas will develop along the tail.

Last, the reason for the lack of shading at the very lowest suppressor frequencies requires comment. In this frequency region, below the boundary representing the membrane filter, the unfiltered ac receptor potential may itself produce a dc receptor potential as a result of rectification by the synapse (Allen, 1983). Thus, a very low-frequency suppressor would produce an increase in discharge rate and would no longer be considered non-excitatory. Rate decreases, however, have been observed at the single unit level for non-excitatory suppressors with frequencies below the cutoff of the basolateral membrane filter (Schmiedt, 1982; Fahey and Allen, 1985; Cai and Geisler, 1996; Temchin *et al.*, 1997). As we stated earlier (Cheatham and Dallos, 1995), these effects may relate to the velocity dependence of the IHC's stereociliary deflection for low-frequency inputs. For example, the probe would be reduced due to interactions prior to IHC mechanoelectrical transduction. The very low-frequency suppressor, however, would not produce a neural response when presented alone because the IHC responds to basilar membrane velocity, i.e., the suppressor is attenuated due to high-pass filtering.

The scenario presented here suggests that basilar membrane mechanics could very well be the source of low-side suppression. This was also the conclusion reached by Temchin *et al.* (1997) who reported that the temporal patterns and thresholds of low-frequency suppression were similar in single unit and basilar membrane responses. In addition, rate suppression thresholds are known to be independent of spontaneous activity (Pang and Guinan, 1997; Temchin *et al.*, 1997), implying that the source of suppression is peripheral to the IHC synapse. Finally, modeling efforts (Geisler *et al.*, 1993) also suggest that the same mechanism can work for all suppression behaviors. It is, therefore, reasonable to propose that a low-side suppressor does not excite auditory-nerve fibers because of the velocity dependence of the IHC transducer, the membrane/synaptic filtering and the sensitivity differences between ac and dc components of the IHC receptor potential. As a result, the total two-tone rate response is smaller than that for the probe alone condition. In contrast, the overall response of the basilar membrane to paired tones is always greater than that for the probe alone when the suppressor is well below CF.

ACKNOWLEDGMENTS

Work supported by NIDCD Grant No. DC00089. The author thanks M. Ruggero and P. Dallos for comments.

Allen, J. B. (**1983**). "A hair cell model of neural response," in *Mechanics of Hearing*, edited by E. d. Boer and M. A. Viergever (Delft University Press, The Netherlands), pp. 193–202.

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). "Properties of two-tone inhibition in primary auditory neurons," J. Physiol. (London) **212**, 21–31.

Cai, Y., and Geisler, C. D. (**1996**). "Suppression in auditory-nerve fibers of cats using low-side suppressors. II. Effect of spontaneous rates," Hear. Res. **96**, 113–125.

Cheatham, M. A., and Dallos, P. (**1992**). "Two-tone suppression in inner hair cell responses: Correlates of rate suppression in the auditory nerve," Hear. Res. **60**, 1–12.

Cheatham, M. A., and Dallos, P. (**1995**). "Origins of rate versus synchrony suppression: An evaluation based on two-tone interactions observed in mammalian IHCs," in *Advances in Hearing Research*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus (World Scientific, Irsee, Bavaria), pp. 145–155.

Cheatham, M. A., and Dallos, P. (**1999**). "Response phase: A view from the inner hair cell," J. Acoust. Soc. Am. **105**, 799–810.

Cooper, N. P. (**1996**). "Two-tone suppression in cochlear mechanics," J. Acoust. Soc. Am. **99**, 3087–3098.

Costalupes, J. A., Rich, N. C., and Ruggero, M. A. (**1987**). "Effects of excitatory and non-excitatory suppressor tones on two-tone rate suppression in auditory nerve fibers," Hear. Res. **26**, 155–164.

Dallos, P. (**1984**). "Some electrical circuit properties of the organ of Corti. II. Analysis including reactive elements," Hear. Res. **14**, 281–291.

Dallos, P. (**1985**). "Response characteristics of mammalian cochlear hair cells," J. Neurosci. **5**, 1591–1608.

Dallos, P., Billone, M. C., Durrant, J. D., Wang, C., and Raynor, S. (**1972**). "Cochlear inner and outer hair cells: Functional differences," Science **177**, 356–358.

Dallos, P., and Santos-Sacchi, J. (**1983**). "AC receptor potentials from hair cells in the low-frequency region of the guinea pig cochlea," in *Mechanisms of Hearing*, edited by W. R. Webster and L. M. Aitkin (Monash University Press, Clayton, Australia), pp. 11–16.

Delgutte, B. (**1990**). "Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers," J. Acoust. Soc. Am. **87**, 791–809.

Diependahl, R. J., de Boer, E., and Viergever, M. A. (**1987**). "Cochlear power flux as an indicator of mechanical activity," J. Acoust. Soc. Am. **82**, 917–926.

Fahey, P. F., and Allen, J. B. (**1985**). "Nonlinear phenomena as observed in the ear canal and the auditory nerve," J. Acoust. Soc. Am. **77**, 599–612.

Geisler, C. D. (**1991**). "A cochlear model using feedback from motile outer hair cells," Hear. Res. **54**, 105–117.

Geisler, C. D., and Nuttall, A. L. (**1997**). "Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using "low-side" suppressors," J. Acoust. Soc. Am. **102**, 430–440.

Geisler, C. D., and Shan, X. (**1990**). "A model for cochlear vibrations based on feedback from motile outer hair cells," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, New York), pp. 86–95.

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.

Geisler, C. D., Bendre, A., and Liotopoulos, F. K. (**1993**). "Time-domain modeling of a nonlinear, active model of the cochlea," in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 330–337.

Harris, D. M. (**1979**). "Action potential suppression, tuning curves and thresholds: Comparison with single fiber data," Hear. Res. **1**, 133–154.

Hind, J. E., Anderson, D. J., Brugge, J. F., and Rose, J. E. (**1967**). "Coding of information pertaining to paired low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 794–816.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.

Kidd, R. C., and Weiss, T. F. (**1990**). "Mechanisms that degrade timing information in the cochlea," Hear. Res. **49**, 181–207.

Kim, D. O. (**1985**). "A review of nonlinear and active cochlear models," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 239–249.

Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). "Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure," J. Acoust. Soc. Am. **67**, 1704–1721.

Kolston, P. J., de Boer, E., Viergever, M. A., and Smoorenburg, G. F. (**1990**). "What type of force does the cochlear amplifier produce?," J. Acoust. Soc. Am. **88**, 1794–1801.

Neely, S. T., and Kim, D. O. (**1983**). "An active cochlear model showing sharp tuning and high sensitivity," Hear. Res. **9**, 123–130.

Nuttall, A. L., Brown, M. C., Masta, R. I., and Lawrence, M. (**1981**). "Inner hair cell responses to the velocity of basilar membrane motion in the guinea pig," Brain Res. **211**, 171–174.

Nuttall, A. L., and Dolan, D. F. (**1993**). "Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig," J. Acoust. Soc. Am. **93**, 390–400.

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential on inner hair cells," Hear. Res. **24**, 1–15.

Pang, X. D., and Guinan, J. J., Jr. (**1997**). "Growth rate of simultaneous masking in cat auditory-nerve fibers: Relationship to the growth of basilar-membrane motion and the origin of two-tone suppression," J. Acoust. Soc. Am. **102**, 3564–3575.

Patuzzi, R. B., and Yates, G. K. (**1987**). "The low-frequency response of inner hair cells in the guinea pig cochlea: Implications for fluid coupling and resonance of the stereocilia," Hear. Res. **30**, 83–98.

Rhode, W. S. (**2007**). "Mutual suppression in the 6 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **121**, 2805–2818.

Russell, I. J., and Kössl, M. (**1992**). "Sensory transduction and frequency selectivity in the basal turn of the guinea-pig cochlea," Philos. Trans. R. Soc. London, Ser. B **336**, 317–324.

Russell, I. J., and Sellick, P. M. (**1978**). "Intracellular studies of hair cells in the guinea pig cochlea," J. Physiol. (London) **284**, 261–290.

Russell, I. J., and Sellick, P. M. (**1983**). "Low-frequency characteristics of intracellularly recorded receptor-potentials in guinea pig cochlear hair cells," J. Physiol. (London) **338**, 179–206.

Sachs, M. B., and Kiang, N. Y.-S. (**1968**). "Two-tone inhibition in auditory nerve fibers," J. Acoust. Soc. Am. **43**, 1120–1128.

Santos-Sacchi, J. (**1989**). "Asymmetry in voltage-dependent movements of isolated outer hair cells from the organ of Corti," J. Neurosci. **9**, 2954–2962.

Schmiedt, R. A. (**1982**). "Boundaries of two-tone rate suppression of cochlear-nerve activity," Hear. Res. **7**, 335–351.

Sellick, P. M., and Russell, I. J. (**1979**). "Two-tone suppression in cochlear hair cells," Hear. Res. **1**, 227–236.

Sellick, P. M., and Russell, I. J. (**1980**). "Responses of inner hair cells to basilar membrane velocity during low frequency auditory stimulation in the guinea pig cochlea," Hear. Res. **2**, 439–446.

Temchin, A. N., Rich, N. C., and Ruggero, M. A. (**1997**). "Low-frequency suppression of auditory nerve responses to characteristic frequency tones," Hear. Res. **113**, 29–56.

Zweig, G. (**1991**). "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am. **89**, 1229–1254.

Speech recognition in noise as a function of highpass-filter cutoff frequency for people with and without low-frequency cochlear dead regions (L)

Vinay[a)]
All India Institute of Speech and Hearing, Manasagangothri, Mysore-570006, India

Thomas Baer and Brian C. J. Moore
Department of Experimental Psychology, Cambridge University, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 10 October 2007; revised 20 November 2007; accepted 20 November 2007)

Regions in the cochlea with very few functioning inner hair cells and/or neurons are called "dead regions" (DRs). Previously, we measured the recognition of highpass-filtered nonsense syllables as a function of filter cutoff frequency for hearing-impaired people with and without low-frequency (apical) DRs [J. Acoust. Soc. Am. 122, 542–553 (2007)]. DRs were diagnosed using the TEN(HL) test, and psychophysical tuning curves were used to define the edge frequency (f_e) more precisely. Stimuli were amplified differently for each ear, using the "Cambridge formula." The present study was similar, but the speech was presented in speech-shaped noise at a signal-to-noise ratio of 3 dB. For subjects with low-frequency hearing loss but without DRs, scores were high (65–80%) for low cutoff frequencies and worsened with increasing cutoff frequency above about 430 Hz. For subjects with low-frequency DRs, performance was poor (20–40%) for the lowest cutoff frequency, improved with increasing cutoff frequency up to about $0.56f_e$, and then worsened. As for speech in quiet, these results indicate that people with low-frequency DRs are able to make effective use of frequency components that fall in the range $0.56f_e$ to f_e, but that frequency components below $0.56f_e$ have deleterious effects. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2823497]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.71.Gv [MW] Pages: 606–609

I. INTRODUCTION

A region in the cochlea where the inner hair cells (IHCs) and/or neurons are functioning so poorly that a tone producing peak basilar-membrane vibration in that region is detected by "off-place" or "off-frequency" listening is called a "dead region" (DR) (Moore, 2004). The extent of a DR can be defined in terms of the characteristic frequencies of the IHCs and/or neurons immediately adjacent to the DR (Moore, 2001). A DR that starts at frequency f_e and extends upwards is called a high-frequency DR, while a DR that starts at frequency f_e and extends downwards is called a low-frequency DR. The value of f_e is referred to as the "edge frequency." It is also possible for a person to have one or more restricted DRs, each of which has a lower and upper edge frequency. The concept of off-place listening provides the basis for psychoacoustic tests for diagnosing DRs, such as psychophysical tuning curves (PTCs) (Thornton and Abbas, 1980; Kluk and Moore, 2005; 2006) and the TEN test (Moore *et al.*, 2000; 2004). Recent studies using the TEN(HL) test (Moore *et al.*, 2004) suggest that, among people with sensorineural hearing loss, there is a greater than 50% prevalence of a DR at a specific frequency when the hearing loss at that frequency is 70 dB or more (Aazh and Moore, 2007; Vinay and Moore, 2007a). However, it has been argued that the presence or absence of a DR cannot be determined reliably from the audiogram alone (Halpin *et al.*, 1994; Moore, 2001; Vinay and Moore, 2007a).

The present study is an extension of a previous study (Vinay and Moore, 2007b) examining the effects of highpass filtering on the recognition of speech in quiet for people with low-frequency hearing loss with and without low-frequency (apical) DRs. Previous relevant studies are reviewed in that paper. Here, we briefly summarize what was done in our previous study. All subjects had sensorineural hearing loss extending to low frequencies. The presence or absence of DRs was determined using the TEN(HL) test (Moore *et al.*, 2004), and PTCs were used to define the value of f_e more precisely for ears with DRs. The speech stimuli were vowel-consonant-vowel (VCV) nonsense syllables. The stimuli were subjected to the frequency-dependent amplification prescribed by the "Cambridge formula" (Moore and Glasberg, 1998), which is roughly equivalent to a half-gain rule for mid-range frequencies. The goal of the amplification was to restore audibility as far as possible, while avoiding excessive loudness. This was done separately for each ear, and all testing was monaural, using headphone presentation. Subjects were tested using broadband speech (upper frequency limit 7500 Hz) and speech that was highpass filtered with various cutoff frequencies.

For subjects with low-frequency hearing loss but without DRs, scores were high (about 78% on average) for low cutoff frequencies, remained approximately constant for cutoff frequencies up to 862 Hz, and then worsened with increasing cutoff frequency. For subjects with low-frequency

[a)]Author to whom correspondence should be addressed. Electronic mail: shrivinyasa@gmail.com.

TABLE I. Mean speech recognition scores (percentage correct) for two subjects with low-frequency DRs and two without DRs for different cutoff frequencies (Hz) and SNRs (dB).

Subject	Cutoff frequency (Hz)	SNR (dB)								
		−6	−3	0	3	6	9	15	30	No noise
S4 LE: DR	610		41	46	49		59			61
f_e=1100 Hz	1220	43	48	51	52	56		58		62
	2440	27	32	35	38	42			46	45
S8 RE: DR	610	35	43	44	48	52		56		58
f_e=1100 Hz	1220	32	33	32	37		38			45
	2440	19	25	27	30		35		43	40
S20 RE:	610	68	70	71	70	76		79		83
No DR	1220	57	59	60		65			71	73
	2440	24	30	29	35		40			42
S22 RE:	610	56	59	59	62	67		70		73
No DR	1220	49	54	57	63		65			69
	2440	17	27	30	33	35		40		41

DRs, performance was typically poor for the lowest cutoff frequency (100 Hz), improved as the cutoff frequency was increased to about $0.57f_e$, and worsened with further increases. The results were interpreted as indicating that people with low-frequency DRs are able to make effective use of frequency components that fall in the range $0.57f_e$ to f_e, but that frequency components below $0.57f_e$ have deleterious effects.

Vinay and Moore (2007b) discussed the implications of their results for the fitting of hearing aids. They stated that "If our results can be generalized to the perception of speech in everyday life…frequency components with frequencies down to about $0.57f_e$ should be amplified, but components with frequencies lower than that should not be amplified." However, their results were all obtained for speech in quiet, whereas in everyday life background noise is often present. It was not clear whether the same pattern of results would be obtained when background noise was present. This study was similar to that of Vinay and Moore (2007b), except that speech recognition was measured in the presence of a speech-shaped noise whose level was chosen so as to produce a moderate reduction in intelligibility relative to that measured in quiet.

II. METHOD

A. Subjects

The 28 subjects tested by Vinay and Moore (2007b) were also tested here. Details of the diagnosis of the hearing loss as sensorineural are given in Vinay and Moore (2007b). All subjects had reasonably flat audiograms with hearing loss at low frequencies (averaged over the frequencies 500, 750 and 1000 Hz) of 40 dB or more. Audiometric thresholds for the ears tested are presented in Vinay and Moore (2007b). That paper also describes in detail how DRs were diagnosed, and how the values of f_e were determined for subjects with DRs. Sixteen subjects were diagnosed as having a low-frequency DR in one or both ears. An ear was selected for further testing if the value of f_e was above 750 Hz (it was anticipated that a DR with f_e at 750 Hz or below would have little effect on speech intelligibility), and if the DR appeared to be continuous rather than "patchy." In some cases only one ear was tested due to the subject only being available for a limited time. Twelve subjects were diagnosed as not having a DR. Two were available for only a limited time and were tested using one ear only. The remainder were tested using both ears.

B. Stimuli and conditions

Speech intelligibility was measured using vowel-consonant-vowel (VCV) nonsense syllables, spoken by a man. The stimuli and presentation method were the same as described by Vinay and Moore (2007b). Ten lists were recorded onto compact disk (CD), each with the stimuli in a different order. A continuous noise with the same long-term average spectrum as the speech was recorded on the second channel of the CD. All stimuli were given frequency-selective amplification as described by Vinay and Moore (2007b) before presentation via one earpiece of a set of Sennheiser HD580 headphones.

A pilot experiment was conducted to determine an appropriate speech-to-noise ratio (SNR) to be used in the main experiment. The goal was to find an SNR that could be used with all subjects and that satisfied the following requirements: (1) The noise produced a moderate (approximately 10%) reduction of speech recognition relative to that measured in quiet; (2) "floor" and "ceiling" effects were minimal. Two subjects with DRs and two without DRs were tested in the pilot experiment. Each subject was tested using three highpass filter cutoff frequencies (610, 1220 and 2440 Hz) using several SNRs, chosen to encompass the SNR leading to a 10% reduction in speech recognition relative to that measured in quiet. The results are shown in Table I. On the basis of these results, an SNR of 3 dB was chosen for the main experiment. On average, this led to a 9.3% reduction in speech recognition relative to that measured in quiet. Although the SNR of 3 dB sometimes led to low scores, especially when the cutoff frequency was 2440 Hz, SNRs below 3 dB always led to worse performance, so floor effects were minimal.

In the main experiment, speech recognition was mea-

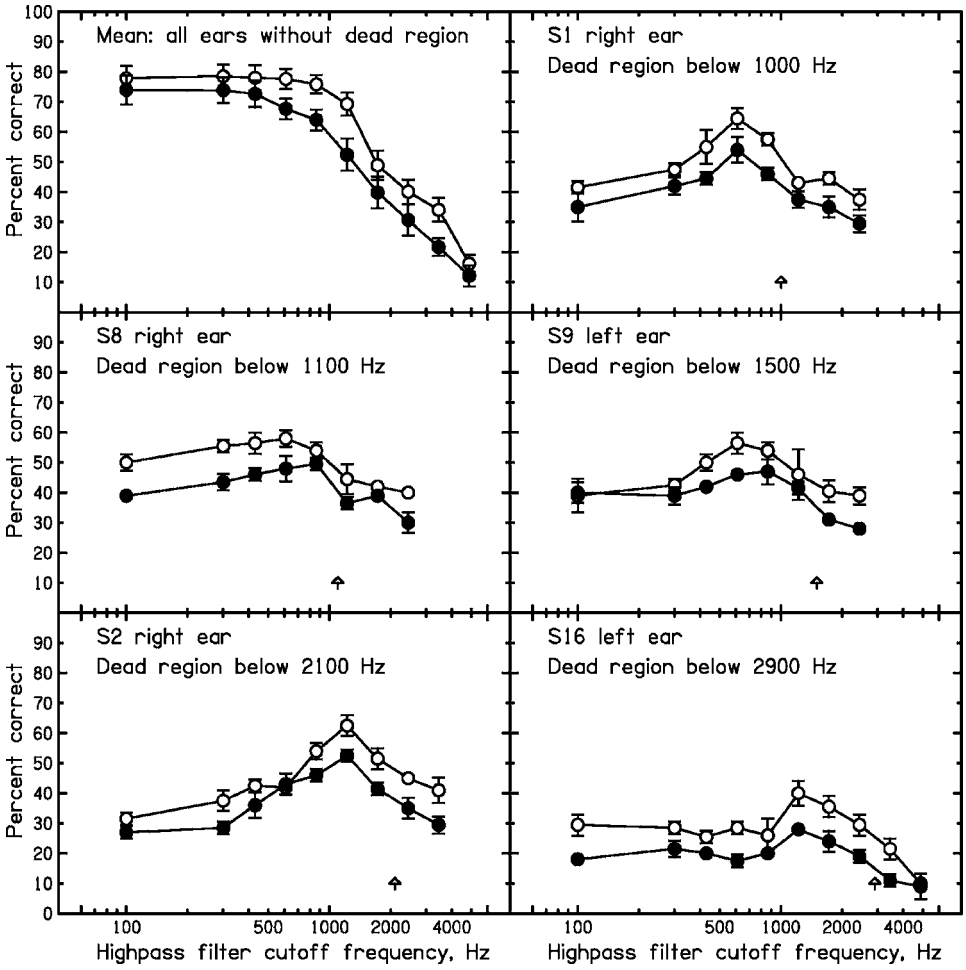


FIG. 1. Mean results of the VCV test across all ears without any DR (top-left panel) and for five ears of five subjects with DRs. The percent correct score is plotted as a function of high-pass filter cutoff frequency. Filled circles show results from the present study for speech in noise. Open circles reproduce results for speech in quiet from Vinay and Moore (2007b). Up-pointing arrows indicate the estimated edge frequency of the DR, f_e. For the individual ears, error bars indicate ± one standard deviation (SD) across repeated runs. For the mean data across ears, the error bars indicate ± one standard deviation (SD) across ears.

sured using the selected SNR of 3 dB as a function of high-pass filter cutoff frequency. Following amplification, stimuli were highpass filtered with one of the following cutoff frequencies: 100, 300, 430, 610, 862, 1220, 1725, 2440, 3450 and 4880 Hz. All other aspects of the stimuli and method were as described by Vinay and Moore (2007b). Testing was conducted in two sessions. Each cutoff frequency was used once in each session. The order of cutoff frequencies was randomized for the first session and the order was reversed for the second session to counterbalance effects of practice and fatigue. Scores presented are the means (and standard deviations) for each cutoff frequency across the two sessions.

III. RESULTS AND DISCUSSION

The pattern of results was similar across subjects without DRs. The filled circles in the top-left panel of Fig. 1 show the mean results across all 22 ears tested. For comparison, results for speech in quiet, obtained by Vinay and Moore (2007b), are shown as open circles. As expected, scores were lower for speech in noise than for speech in quiet. The pattern of results was similar for the two cases; at first, scores remained roughly constant with increasing cutoff frequency, and then they started to decline. However, scores started to decline at a lower cutoff frequency for speech in noise than for speech in quiet. The standard error of the mean scores was typically less than 1%, so taking a decline in score of 2% as being significant, scores for speech in quiet started to decline for a cutoff frequency of about 860 Hz, whereas scores for speech in noise started to decline for a cutoff frequency just above 430 Hz.

Results for the subjects with DRs were also similar in form to those found previously for speech in quiet. Results for five representative ears are shown in Fig. 1. These ears were selected so as to cover a range of values of f_e. For the lowest cutoff frequency (100 Hz), scores were rather low, typically between 20 and 40%. With increasing cutoff frequency, scores initially remained roughly constant, then improved, reached a maximum, and then decreased. The cutoff frequency giving the maximum score, f_{max}, varied across ears, and was related to the value of f_e. The Pearson correlation between f_{max} and f_e was 0.87 ($n=19$), which is statistically significant ($p<0.01$). The average ratio f_e/f_{max} was 1.78, with a standard deviation (SD) of 0.32. Put another way, the speech identification scores of the subjects with DRs were maximal when the cutoff frequency was $f_e/1.78$ or $0.56f_e$. The average ratio f_e/f_{max} is very similar to that obtained by Vinay and Moore (2007b) for speech in quiet.

Sequential information analysis (SINFA) (Wang and Bilger, 1973) was used to determine if the pattern of information about the phonetic features of voicing, manner and place (Miller and Nicely, 1955) was affected by the highpass filter cutoff frequency differently for the subjects with and without DRs. This was investigated using the stimulus-response matrices for subjects without DRs and for subjects

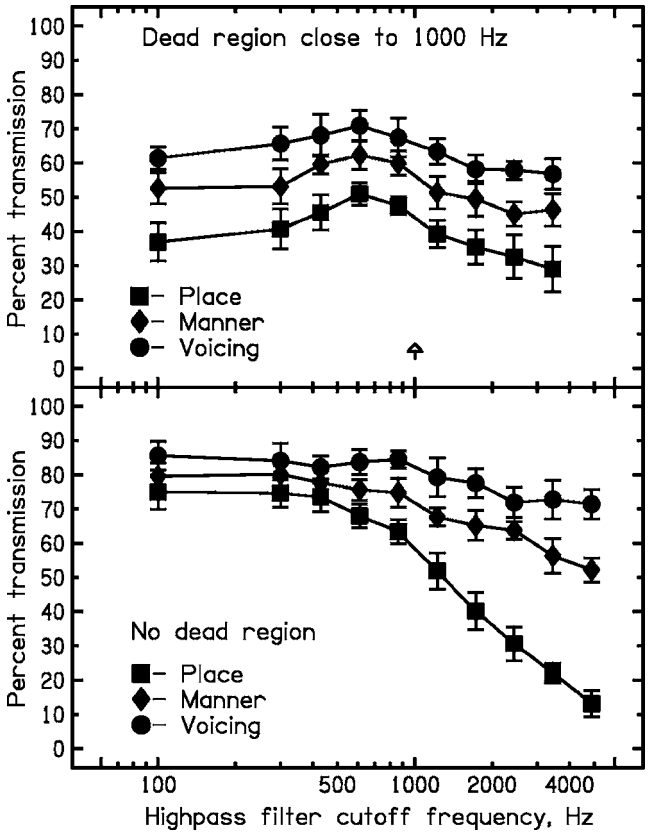


FIG. 2. Results of the SINFA analysis for subjects with DRs with f_e close to 1000 Hz (top) and for subjects without any DR (bottom). Error bars indicate±one standard deviation (SD) across ears.

with DRs with f_e close to 1000 Hz. The percentage of input information transmitted (IT) for a given feature was determined using the SINFA Analysis Suite "FIX" (Mike Johnson, Department of Phonetics and Linguistics, University College London). The analysis was conducted separately for each ear in the two groups, and then the results were averaged across the ears within each group.

The results are shown in Fig. 2. For the subjects without any DR (bottom), the IT was highest for voicing and lowest for place. With increasing cutoff frequency, the IT for voicing declined only slightly, the IT for manner declined somewhat more, and the IT for place declined dramatically, consistent with the idea that IT for place depends on spectral information over a reasonably wide frequency range. For the subjects with DRs (top), the ordering of IT values for the different features was the same as for the subjects without DRs. However, for low cutoff frequencies, there was a decrease in transmission of all features. It appears that providing speech information at frequencies below about $0.56f_e$ impairs the transmission of all features. Possible reasons for this were discussed by Vinay and Moore (2007b).

Overall, the results presented here for speech in noise are similar in form to those found previously for speech in quiet (Vinay and Moore, 2007b). The results support our tentative recommendation that, for people with low-frequency DRs with edge frequency f_e, amplification should be provided for frequencies down to about $0.56f_e$, but not for lower frequencies. However, further work is needed to establish whether similar results are obtained for sentence material, or under conditions where the face of the talker is visible.

ACKNOWLEDGMENTS

We thank the Director, All India Institute of Speech and Hearing, for providing the facilities to carry out this work. The work of the second and third authors was supported by the Medical Research Council (UK).

Aazh, H., and Moore, B. C. J. (**2007**). "Dead regions in the cochlea at 4 kHz in elderly adults: Relation to absolute threshold, steepness of audiogram, and pure tone average," J. Am. Acad. Audiol **18**, 96–107.

Halpin, C., Thornton, A., and Hasso, M. (**1994**). "Low-frequency sensorineural loss: Clinical evaluation and implications for hearing aid fitting," Ear Hear. **15**, 71–81.

Kluk, K., and Moore, B. C. J. (**2005**). "Factors affecting psychophysical tuning curves for hearing-impaired subjects," Hear. Res. **200**, 115–131.

Kluk, K., and Moore, B. C. J. (**2006**). "Detecting dead regions using psychophysical tuning curves: A comparison of simultaneous and forward masking," Int. J. Audiol. **45**, 463–476.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Moore, B. C. J., and Glasberg, B. R. (**1998**), "Use of a loudness model for hearing aid fitting," Br. J. Audiol. **32**, 317–335.

Moore, B. C. J. (**2001**). "Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids," Trends Amplif. **5**, 1–34.

Moore, B. C. J. (**2004**). "Dead regions in the cochlea: Conceptual foundations, diagnosis and clinical applications," Ear Hear. **25**, 98–116.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**2004**). "New version of the TEN test with calibrations in dB HL," Ear Hear. **25**, 478–487.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.

Thornton, A. R., and Abbas, P. J. (**1980**). "Low-frequency hearing loss: Perception of filtered speech, psychophysical tuning curves, and masking," J. Acoust. Soc. Am. **67**, 638–643.

Vinay, and Moore, B. C. J. (**2007a**). "Prevalence of dead regions in subjects with sensorineural hearing loss," Ear Hear. **28**, 231–241.

Vinay, and Moore, B. C. J. (**2007b**). "Speech recognition as a function of highpass filter cutoff frequency for subjects with and without low-frequency cochlear dead regions," J. Acoust. Soc. Am. **122**, 542–553.

Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

An improved acoustical wave propagator method and its application to a duct structure

S. Z. Peng[a] and L. Cheng
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

(Received 16 May 2007; accepted 12 November 2007)

The pseudospectral time-domain method has long been used to describe the acoustical wave propagation. However, due to the limitation and difficulties of the fast Fourier transform (FFT) in dealing with nonperiodic problems, the dispersion error is inevitable and the numerical accuracy greatly decreases after the waves arrive at the boundary. To resolve this problem, the Lagrange–Chebyshev interpolation polynomials were used to replace the previous FFT, which, however, brings in an additional restriction on the time step. In this paper, a mapped Chebyshev method is introduced, providing the dual benefit of preserving the spectral accuracy and overcoming the time step restriction at the same time. Three main issues are addressed to assess the proposed technique: (a) Spatial derivatives in the system operator and the boundary treatment; (b) parameter selections; and (c) the maximum time step in the temporal operator. Furthermore, a numerical example involving the time-domain evolution of wave propagation in a duct structure is carried out, with comparisons to those obtained by Euler method, the fourth-order Runge–Kutta method, and the exact analytical solution, to demonstrate the numerical performance of the proposed technique.
 [DOI: 10.1121/1.2821971]

PACS number(s): 43.20.Mv, 43.20.Bi [DSB] Pages: 610–621

I. INTRODUCTION

Partial differential equations (PDEs) describe a wide variety of physical processes such as molecular dynamics, heat conduction, flow, and sound propagation. Over the last decade, a great deal of effort has been devoted to the development of numerical methods for solving the time-dependent PDEs, in particular the time-dependent Schrödinger equation and wave equations.[1–8] Since the time-domain investigation provides insightful understanding on the governing physical phenomena, various numerical schemes have been developed in parallel in many fields by quantum chemists, quantum physicists, and acousticians with little across referencing. Typical numerical methods include the finite-difference time-domain, finite-element time-domain, and time-domain boundary-element method for modeling and simulating molecular encounters, calculating the dynamical properties of a quantum mechanical system including molecular dynamics, and for predicting transient wave propagation and scattering problems.[9,10]

Consider a wave function ψ governed by the time-dependent Schrödinger equation,[11]

$$i\frac{\partial}{\partial t}\psi(x,t) = H\psi(x,t), \tag{1}$$

where

$$H = \underbrace{p^2/2m}_{K} + V$$

is the Hamiltonian operator; K and V the kinetic and potential operators, respectively; and p and m are the momentum and mass, respectively.

The solution of Eq. (1), with H independent of time, is given by

$$\psi(x,t+\tau) = e^{-iH\tau}\psi(x,t), \tag{2}$$

where $\hat{Q}=e^{-iH\tau}$ denotes the quantum wave propagator, which maps the wave function $\psi(x,t)$ at any time t to that at next time $t+\tau$.

To implement the operation of the exponential propagator, $\hat{Q}$ is often approximated to a finite polynomial expansion. A comprehensive discussion on various expansion schemes can be found in the review article by Balakrishnan *et al.*[12] It is worth noting that Kosloff and Tal-Ezer[13–15] did a pioneering work on the pseudospectral method in the field of quantum chemistry and atomic physics, which is an important step in developing time-domain numerical methods. In particular, they described a Chebyshev expansion of the evolution operator as an efficient and accurate method for calculating the time-dependent Schrödinger equation. After that, they proposed a modified Chebyshev pseudospectral method with an $O(N^{-1})$ time step restriction, which was the main drawback in restrictive stability properties.[16] Don and Solomonoff[17] studied a similar method in reducing the round-off error and calculated spatial derivatives using Chebyshev collocation methods. Leforestier *et al.*[18] discussed the advantages and drawbacks of several numerical methods in terms of the numerical accuracy, computational efficiency, and stability. Therefore, it is not surprising that great effort has been devoted to finding the optimum numerical method with an efficient, accurate, and stable numerical procedure to solve the time-dependent PDEs. Recently, Pan

[a]Author to whom correspondence should be addressed. Electronic mail: mmszpeng@polyu.edu.hk

and Wang[19] developed an explicit acoustical wave propagator (AWP) method to describe the time-domain evolution of acoustical waves. The Fourier transform scheme and the modified Bessel function of the first kind were used to evaluate the spatial derivatives, and to implement the acoustical wave propagator $e^{-\hat{H}\tau}$, respectively. The main difference between the quantum wave propagator $e^{-iH\tau}$ and the acoustical wave propagator $e^{-\hat{H}\tau}$ is that the former is complex while the latter is real in a matrix form representing the selected variables.

Examining the spatial and temporal discretization of the method shows that it is necessary to use a higher-order time differencing scheme to save computational time (by an increase in the time step size). The numerical accuracy of the method can be ensured by a good approximation of the spatial derivatives and the temporal exponential. Theoretically speaking, if $e^{-(t-t_0)\hat{H}}$ can be implemented together with the known initial-value vectors $\psi(x,t_0)$, then the above-mentioned AWP method can be used to predict the acoustical wave propagation like the exact analytical solutions [the order of the prediction error in dimensionless form is around $O(10^{-7})$] in Refs. 20–22. However, most practical engineering applications involve complex boundary conditions that need to be properly treated. Therefore, the existing problem is that the previous AWP method including the Fourier transform scheme can hardly deal with the nonperiodic problems such as asymmetrical boundary conditions.

More recently, as a further development of the AWP method, Peng and Huang[23] introduced the Lagrange–Chebyshev interpolation polynomials (by fully considering the boundary conditions) to replace the previous Fourier transform scheme in the implementation of the AWP. Despite the improvement on the spatial derivatives, however, the additional restriction on the time step is still left to be solved. The present paper aims to eliminate this restriction by remapping one set of points with Chebyshev–Gauss–Lobatto points to another set of points with the modified Chebyshev–Gauss–Lobatto points. By choosing an optimal parameter γ in the mapped Chebyshev method, a larger time step with higher computational efficiency and stability can be achieved. It is worth noting that the most distinct feature of the Chebyshev polynomial expansion scheme (still kept in this implementation of the AWP) is that the maximum polynomial expansion order n can be chosen such that the numerical accuracy is dominated by the machine accuracy of the computer, and the error is uniformly distributed over all the range of eigenvalues. The proposed method is essentially a transition between Fourier (γ=0) and Chebyshev spatial discretization methods.

The outline of the paper is as follows. In Sec. II, we derive an explicit AWP, which includes acoustical waves in a one-dimensional duct structure. Section III introduces the Lagrange–Chebyshev interpolation polynomials scheme for spatial derivatives, describes the treatment of nonperiodic/periodic boundary conditions, implements the AWP with the Chebyshev polynomial expansion, and studies the Fourth-order Kunge–Kutta (RK4) method and the mapped Chebyshev method. In Sec. IV, a numerical example is presented to carry out the error analysis together with the relevant maximum time step. Then the numerical performance (numerical accuracy, computational efficiency, and stability) of the improved AWP scheme is demonstrated. Finally, conclusions are drawn in Sec. V.

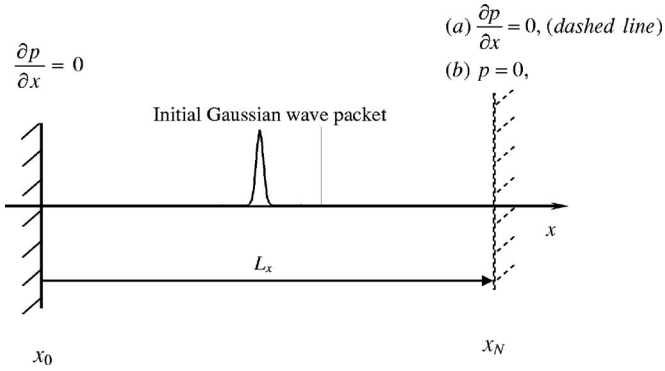


FIG. 1. Illustration of a duct structure with the initial-boundary-value problems: (a) Periodic boundaries (both rigid walls) and (b) nonperiodic boundaries (the left-hand side is rigid wall, the right hand side is pressure release wall).

II. A DUCT STRUCTURE MODEL AND THE AWP

The theoretical model under investigation consists of a one-dimensional duct structure (Fig. 1) with two different boundaries: (a) Periodic (rigid walls at both ends) and (b) nonperiodic (the left-hand side is a rigid wall, with the right-hand side being a pressure-release wall). Given an external sound source $P(x,t)$, the wave will propagate inside the duct.

Acoustical wave motion in the duct is described by the following partial differential equations (PDEs):

$$\frac{\partial p}{\partial t}=-\rho_0 c_0^2\frac{\partial V}{\partial x},$$

$$\frac{\partial V}{\partial t}=-\frac{1}{\rho_0}\frac{\partial p}{\partial x}, \tag{3}$$

where ρ_0 is the air density; c_0 the speed of sound; p the sound pressure; and V the particle velocity along the x direction in the duct.

To derive the acoustical wave propagator in the duct, a state vector ϕ_D consisting of sound pressure p and particle velocity V is formed, transforming Eq. (3) into the following state equation:

$$\frac{\partial}{\partial t}\underbrace{\begin{bmatrix}p\\V\end{bmatrix}}_{\phi_D}=-\hat{H}_D\begin{bmatrix}p\\V\end{bmatrix}, \tag{4}$$

(4)where

$$\hat{H}_D=\begin{bmatrix}0 & \rho_0 c_0^2\frac{\partial}{\partial x}\\ \frac{1}{\rho_0}\frac{\partial}{\partial x} & 0\end{bmatrix}. \tag{5}$$

Integrating Eq. (4) with respect to time yields

$$\phi_D(x,t) = e^{-(t-t_0)\hat{H}_D}\phi_D(x,t_0), \tag{6}$$

where $e^{-(t-t_0)\hat{H}_D}$ is defined as the AWP in a one-dimensional duct structure with the subscript D denoting the duct structure.

The numerical solution of Eq. (4) includes both the first-order spatial and temporal derivatives. When the initial values $\phi_D(x,t_0)$ are known, there are two key steps to obtain $\phi_D(x,t)$: (a) to calculate the spatial derivatives in $\hat{H}_D$ and (b) to adopt an efficient and accurate method to implement the exponential expansion $e^{-(t-t_0)\hat{H}_D}$.

III. CALCULATION OF THE SPATIAL DERIVATIVES AND IMPLEMENTATION OF THE AWP

As a class of methods, pseudospectral methods are used in solving the above-mentioned PDEs. The Fourier transform is often involved in the calculation of the spatial derivatives. For well-behaved problems (isotropic medium, symmetrical structure with periodic boundary conditions), the Fourier transform is very useful to evaluate the spatial derivatives due to its numerical accuracy and computational efficiency,

$$F\left[\frac{\partial'''}{\partial x'''}\psi(x,t)\right] = (jk_x)'''F[\psi(x,t)], \quad \frac{\partial'''}{\partial x'''}\psi(x,t) = F^{-1}\{(jk_x)'''F[\psi(x,t)]\}, \tag{7}$$

where $F[\ ]$ and $F^{-1}\{\ \}$ denote the Fourier and inverse Fourier transforms, respectively, and k_x is the wave number. The spatial approximation for the derivatives utilizes the property of the Fourier transforms that a derivative in the spatial domain becomes a multiplication by ik_{x_l} (k_{x_l} is the wave number corresponding to the l-spatial coordinate) in the spatial frequency domain, and then performs an inverse Fourier transform back to the spatial domain, as described in Eq. (7).

In the finite-difference and finite-element methods, time and space are discretized with a uniform grid. Typically, they do not attain good convergence even for an infinitely smooth function. Note the differences between the Fourier transform and the finite-difference method: The former is based on global approximations; and the latter on local approximations based on Taylor expansion, which is accompanied by the truncation errors. However, the numerical errors in the Fourier transform scheme are mainly due to the approximation of double integrals described in a discrete form. More precisely, the Fourier transform scheme can obtain very small truncation errors (high accuracy) by the positive and negative value cancellation in a global error summation except for two end points. It is observed that when the grid spacing Δx is small, the Fourier transform scheme has much higher accuracy than the multipoint finite-difference methods.

The above-mentioned discrete Fourier transform scheme implies periodic boundaries, which are natural for describing spatially periodic problems. However, for problems where the natural boundary conditions are nonperiodic, the Fourier transform scheme will introduce additional numerical dispersion and the numerical accuracy rapidly deteriorates. When this dispersion becomes too severe, the solutions to Eq. (2) no longer exist. The stability concern actually limits these explicit methods (Euler and RK4 methods) to small Δt except for the Chebyshev polynomial expansion scheme, where the time step Δt is mainly dependent on the parameter $R = dt|\lambda_{\max}|$, and normalized matrix $\hat{H}'_D$. For different structures, the system operator $\hat{H}$ has different forms leading to different maximum eigenvalues $\lambda_{\max}$. For example, for the sound pressure in a one-dimensional duct, $R = c_0\Delta t\pi/\Delta x$; for flexural waves in a thin flexible beam, $R = \sqrt{EI/\rho A}(\pi/\Delta x)^2 dt$.

A. Lagrange–Chebyshev interpolation polynomials scheme for spatial derivatives

In Eq. (5), there are two first-order spatial derivatives for sound pressure $\partial p(x,t)/\partial x$ and the particle velocity $\partial V(x,t)/\partial x$. For simplification, only the former $\partial p(x,t)/\partial x$ is derived as follows. The latter can be treated in a similar way.

Usually, the Chebyshev pseudospectral method is based on polynomial interpolation in the canonical interval $[-1,1]$. However, it can be defined on any finite internal $[x_0,x_N]$ for a general case by means of a linear transform of variable χ which maps $x \in [x_0,x_N]$ onto $[-1,1]$,

$$x = \frac{x_N - x_0}{2}\chi + \frac{x_N + x_0}{2}, \tag{8}$$

which is discretized at Gauss–Lobatto points $\chi_i = \cos[(N-i)\pi/N]$, $i = 0,1,\ldots,N$.

In this scheme, $\partial/\partial\chi$ is represented by a matrix $d_\chi = [d_{i,k}]$ with its elements $d_{i,k}$ given in the following (Ref. 23):

$$d_\chi = \begin{bmatrix} -\frac{2N^2+1}{6} & -\frac{2}{\chi_0-\chi_1} & \cdots & \frac{2(-1)^{N-1}}{\chi_0-\chi_{N-1}} & \frac{(-1)^N}{\chi_0-\chi_N} \\ -\frac{1}{2(\chi_1-\chi_0)} & -\frac{\chi_1}{2(1-\chi_1^2)} & \cdots & \frac{(-1)^N}{\chi_1-\chi_{N-1}} & \frac{(-1)^{N+1}}{2(\chi_1-\chi_N)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{(-1)^{N-1}}{2(\chi_{N-1}-\chi_0)} & \frac{(-1)^N}{(\chi_{N-1}-\chi_1)} & \cdots & -\frac{\chi_{N-1}}{2(1-\chi_{N-1}^2)} & \frac{(-1)^{2N-1}}{2(\chi_{N-1}-\chi_N)} \\ \frac{(-1)^N}{\chi_N-\chi_0} & \frac{2(-1)^{N+1}}{\chi_N-\chi_1} & \cdots & \frac{2(-1)^{2N-1}}{\chi_N-\chi_{N-1}} & \frac{2N^2+1}{6} \end{bmatrix}. \tag{9}$$

The spatial derivative of a discretized function is given as

$$\frac{\partial\tilde{p}(\chi,t)}{\partial\chi} = \sum_{k=0}^{N} d_{i,k}\tilde{p}_k(\chi,t), \tag{10}$$

which can also be expressed as $[\partial\tilde{p}(\chi,t)/\partial\chi] = d_\chi[\tilde{p}(\chi,t)]$, where $[\partial\tilde{p}(\chi,t)/\partial\chi]$ and $[\tilde{p}(\chi,t)]$ are column matrices for the $(N+1)$ discrete points.

Theoretically, mth-order spatial derivatives $d_\chi^{(m)}$ can be obtained by the matrix product of

$$\underbrace{d_\chi^{(1)} \times d_\chi^{(1)} \times \cdots \times d_\chi^{(1)}}_{m}$$

[$d_z^{(1)} = d_\chi$ in Eq. (9)]. When $N+1$ grid points in χ axis are given, according to the relations between $x \in [x_0,x_N]$ and χ

$\in[-1,1]$, $\partial\tilde{p}(x,t)/\partial x$ can be obtained by multiplying $\partial\tilde{p}(\chi,t)/\partial\chi$ with the constant $2/(x_N-x_0)$.

B. Nonperiodic/periodic boundary conditions

A mathematical model with nonperiodic boundary conditions is introduced to describe how the boundary conditions are considered in the spatial derivatives. Here, as shown in Fig. 1, two different boundary conditions are considered: A rigid wall with $\partial\tilde{p}(x,t)/\partial x=0\Leftrightarrow\partial\tilde{p}(\chi,t)/\partial\chi=0$ is imposed on the left-hand side and a pressure-release wall with $\tilde{p}(x,t)=0$ is used on the right-hand side. The procedure is described as follows: (a) All values at the initial condition $\phi_D(x,t_0)$ are known, so $\partial\tilde{p}(x,t_0)/\partial x$ [obtained from $\partial\tilde{p}(\chi,t_0)/\partial\chi$] can be calculated; (b) the values at the new time step t_0+dt can be obtained by using $\phi_D(x,t_0+dt)=e^{-dt\hat{H}}\phi_D(x,t_0)$ for all inner points; then $\partial\tilde{p}(x,t_0+dt)/\partial x$ should be recalculated from $\partial\tilde{p}(\chi,t_0+dt)/\partial\chi$; and (c) the boundary conditions are applied to get $\phi_D(x,t_0+dt)$ on all boundaries. For example, the sound pressure $\tilde{p}(x_0,t_0+dt)$ and its spatial derivative on the left-hand side (rigid wall) are calculated by

$$\tilde{p}(x_0,t_0+dt)=-\Big(d_{0,1}\tilde{p}(x_1,t_0+dt)+\cdots+d_{0,N-1}\tilde{p}(x_{N-1},t_0+dt)+d_{0,N}\underbrace{\tilde{p}(x_N,t_0+dt)}_{=0}\Big)\Big/d_{0,0},\tag{11}$$

and

$$\partial\tilde{p}(x_N,t_0+dt)/\partial x=2\Big(d_{N,0}\tilde{p}(x_0,t_0+dt)+\cdots+d_{N,N-1}\tilde{p}(x_{N-1},t_0+dt)+d_{N,N}\underbrace{\tilde{p}(x_N,t_0+dt)}_{=0}\Big)\Big/(x_N-x_0),\tag{12}$$

where $\tilde{p}(x_0,t_0+dt)$ in Eq. (12) should be obtained from Eq. (11). Spatial derivatives on the boundaries are calculated by the product of the Chebyshev derivative matrix with the acoustical quantity in the whole field.

Similarly, periodic boundary conditions [rigid wall conditions with $\partial\tilde{p}(x,t)/\partial x=0\Leftrightarrow\partial\tilde{p}(\chi,t)/\partial\chi=0$ imposed on both the left-hand and right-hand sides] should be fully considered in the spatial derivatives. The sound pressures $\tilde{p}(x_0,t_0+dt)$ and $\tilde{p}(x_N,t_0+dt)$ on the left-hand and right-hand sides are, respectively, calculated by

$$\tilde{p}(x_0,t_0+dt)=-\big((d_{0,1}d_{N,N}-d_{0,N}d_{N,1})\tilde{p}(x_1,t_0+dt)+\cdots+(d_{0,N-1}d_{N,N}-d_{0,N}d_{N,N-1})\tilde{p}(x_{N-1},t_0+dt)\big)/(d_{0,0}d_{N,N}-d_{0,N}d_{N,0}),$$

$$\tilde{p}(x_N,t_0+dt)=\big((d_{0,1}d_{N,0}-d_{0,0}d_{N,1})\tilde{p}(x_1,t_0+dt)+\cdots+(d_{0,N-1}d_{N,0}-d_{0,0}d_{N,N-1})\tilde{p}(x_{N-1},t_0+dt)\big)/(d_{0,0}d_{N,N}-d_{0,N}d_{N,0}).\tag{13}$$

It should be noted that the above-mentioned procedures are entirely explicit.

C. Chebyshev polynomial expansion schemes with *I* or *J* expansions

It is worth noting that the strategy chosen for the propagating scheme is to expand the evolution operator $AWP=e^{-\hat{H}_d dt}$ in a polynomial series. This strategy becomes a choice of the best polynomial approximation for this series, and the operator $e^{-(t-t_0)\hat{H}_D}$ on the initial wave packet can then be evaluated. The most accurate and stable method to date is the Chebyshev polynomial expansion scheme, which is used to implement the temporal approximation of the exponential propagator $e^{-(t-t_0)\hat{H}_D}$. Since the argument of these polynomials in $\phi_D(x,t)$ is defined in the range of $[-1,1]$, one needs to normalize the system operator $\hat{H}_D$ by $\hat{H}'_D=\hat{H}_D/|\lambda_{\max}|$, where $\lambda_{\max}$ denotes the maximum eigenvalue of $\hat{H}_D$. It is worth noting that the time-domain Chebyshev polynomial expansion scheme has the potential to keep a low error with very large time steps, which will be described in Sec. IV B.

The Chebyshev polynomial expansion schemes with *I* and *J* expansions can be used to expand the acoustical wave propagator. Their respective advantages and existing problems will be discussed in the following.

Denoting $R=(t-t_0)|\lambda_{\max}|$, Eq. (6) can be presented by

$$\phi_D(x,t)=\Big[I_0(R)\underbrace{I}_{T_0(\hat{H}'_D)}+2I_1(R)\underbrace{\hat{H}'_D}_{T_1(\hat{H}'_D)}+2\sum_{n=2}^{\infty}I_n(R)T_n(\hat{H}'_D)\Big]\phi_D(x,t_0),\tag{14}$$

where I is a unit matrix with the same size as that of $\hat{H}'_D$; $I_n(R)$ is the nth-order modified Bessel function of the first kind; and $T_n(\hat{H}'_D)$ is the nth-order Chebyshev polynomials, which can be calculated by the following recursive relations: $T_n(\hat{H}'_D)=2\hat{H}'_D T_{n-1}(\hat{H}'_D)-T_{n-2}(\hat{H}'_D)$, where $n\geqslant 2$. The main advantage of the Chebyshev polynomial expansion schemes with *I* expansion is that expansion coefficients of Chebyshev polynomials decay exponentially when the order of the coefficient function is sufficiently larger than its argument R, which is the important parameter as function of the size of time step and the maximum eigenvalue of the system operator.

However, the existing problem of the *I* expansion with the modified Bessel functions is that there is a limitation due to the dynamic range of expansion functions covered by a single AWP propagation [for a large R, $I_0(30)=7.8\times10^{11}$ comes close to the dynamic range of 10^{16}]. Therefore, the Bessel functions of the first kind, as an alternative expansion, are considered to replace the previous modified Bessel functions of the first kind. Thus, Eq. (6) can be represented by

$$\phi_D(x,t)=\Big[J_0(R)\tilde{T}_0(\hat{H}'_D)+2J_1(R)\tilde{T}_1(\hat{H}'_D)+2\sum_{m=2}^{M}J_m(R)\tilde{T}_m(\hat{H}'_D)\Big]\phi_D(x,t_0),\tag{15}$$

where $J_m(R)$ is the mth-order Bessel function of the first kind

and $\tilde{T}_m(\hat{H}'_D)$ is the mth-order Chebyshev polynomials, which can be calculated by the following recursive relations: $\tilde{T}_m(\hat{H}'_D)=2\hat{H}'_D\tilde{T}_{m-1}(\hat{H}'_D)-\tilde{T}_{m-2}(\hat{H}'_D)$, where $m\geqslant 2$.

The number of terms in the expansions (15) are governed by the behavior of the Bessel function $J_m(R)$. Equation (15) indicates the orthogonality properties that justify the expansion. When $R=1$, the expansion coefficients decay quickly. As R increases, different from the previous $I_n(R)$, the main advantage of the Chebyshev polynomial expansion schemes with J expansion is that expansion coefficients $J_m(R)$ are only bounded in $[-0.5,0.5]$ for $R=5$ and $[-0.2,0.2]$ for $R=50$. Once again, for large R values, the J expansion is better because large prediction errors can be avoided when some extra terms fall outside the dynamic range defined by the machine accuracy of the computer (10^{-16}). However, for small R values, the previous I expansion is strongly recommended due to its exponential decaying property. At the expense of sacrificing computational efficiency, high accuracy can be further achieved by the time-step splitting method ($e^{-R\hat{H}'_D}=\exp(-\Sigma_{m=1}^{M_R}(R/M_R)\hat{H}'_D)$, where the splitting slice $e^{-(R/M_R)\hat{H}'_D}$ can preserve sufficient prediction accuracy) introduced in Ref. 20.

A proper expansion scheme and the relevant optimal parameters hold the key to achieve accurate prediction results efficiently. Theoretically speaking, if the I expansion is selected, one can choose any large R with enough expansion term $n_{\min}$ to ensure that $I_0(R)/I_{n_{\min}}(R)$ reaches the order of 10^{-16}. However, due to the limitation of the dynamic range (increasing R is accompanied by the increased prediction error in the calculation), R should be chosen in a safety range to ensure its numerical accuracy. If possible, a smaller R is better to get the highly accurate prediction results. When R is given, the expansion term n will significantly affect the prediction results. Therefore, there exists an optimal selection among the time step dt and the minimum expansion term $n_{\min}$, which is very complex compared with the following RK4 method.

The main difference between the present improvement and the previous AWP method in Ref. 20 is that $[\partial\tilde{p}(\chi,t)/\partial\chi]=d_\chi[\tilde{p}(\chi,t)]\rightarrow[\partial\tilde{p}(x,t)/\partial x]=2d_\chi[\tilde{p}(\chi,t)]/(x_N-x_0)$ is included in $\hat{H}'_D$ to replace $F[\partial\tilde{p}(x,t)/\partial x]=ik_xF[\tilde{p}(x,t)]$ included in $\hat{H}'_D$. Once again, the former is nonuniform distribution of Chebyshev–Gauss–Lobatto points, but the latter is uniform distribution of points.

For a fixed step size, the Runge–Kutta method is regarded as a classical technique for the solution of differential equations, especially ordinary differential equations with constant coefficients. Here, the RK4 method is introduced to demonstrate the numerical performance of the proposed technique. Equation (6) can then be expressed as[23]

$$\phi(x,t+dt)=\phi(x,t)-K_{\mathrm{RK4}}\hat{H}_D dt, \tag{16}$$

where K_{RK4} is called the modified coefficient vector, which can be calculated by

$$K_{\mathrm{RK4}}=\phi(x,t)\left[1-\tfrac{1}{2}\hat{H}_D dt+\tfrac{1}{6}\hat{H}_D^2 dt^2-\tfrac{1}{24}\hat{H}_D^3 dt^3\right]. \tag{17}$$

The RK4 method retains only the first four terms in the Taylor expansion, while the Euler method keeps the first term $\phi(x,t)$ in the Taylor expansion. The stability criterion using the Chebyshev–Fourier method, especially its convergence property, has been extensively discussed by Peng and Pan.[20] Euler and RK4 are well-established methods, which have been extensively presented in the literature.[24] The stability criteria of these methods mainly depend on the time step size and the absolute maximum eigenvalue $|\lambda_{\max}|$ of the system operator $\hat{H}_D$. For the Euler method $\phi(x,t+dt)=\phi(x,t)(1-\hat{H}_D dt)$, the stability is ensured when $|I+\lambda_i(-\hat{H}_D dt)|<1\,\forall\, i$, where $\lambda_i(\hat{H}_D)$ denotes the ith eigenvalue of $\hat{H}_D$. Similarly, the stability of the RK4 method is governed by $|I+\lambda_i(-\hat{H}_D dt)+\lambda_i^2(-\hat{H}_D dt)/2+\lambda_i^3(-\hat{H}_D dt)/6+\lambda_i^4(-\hat{H}_D dt)/24|<1\,\forall\, i$. Detailed discussions will be given in Sec. IV. However, the fatal drawback of the Euler and RK4 methods is that the associated numerical error is usually proportional to the time step used in the simulation. This often leads to significant accumulated error in both the magnitude and the phase of the acoustical wave in propagation. The above-mentioned drawback can be further demonstrated by comparing its calculated result with those obtained by the Chebyshev polynomial expansions based on a simple function $f(t)=e^{-t}$. Here, it is necessary to explicitly mention that there are two Chebyshev polynomial expansions (interpolation and extrapolation) to obtain the approximation results. Actual analysis errors not only include truncation errors due to interpolation or extrapolation, but also the stability effects. More details about Chebyshev polynomial expansions for this function can be found in the Appendix.

The main attention does not focus on the high numerical accuracy provided that the extremely small time step is chosen. On the contrary, there is increasing interest in the use of a very long time step, even just one step to complete the calculation for some specific problems. As shown in Fig. 2, the Chebyshev polynomial expansion has much better accuracy for a larger time step, which has the most important effect on computational efficiency, in particular the three-dimensional structure calculations. An important observation is that the Chebyshev polynomial interpolation expansion agrees well with the Chebyshev polynomial extrapolation expansion (AWP Chebyshev) for both cases when $n=4$ and $n=10$. For the Chebyshev methods, when the order of the expansion terms n increases, there are more Chebyshev nodes with higher accuracy. As shown in Fig. 2, the curves go up in value and down to another Chebyshev interpolation node again. An overwhelming performance is that the approximation error nearly keeps the same order from the first iteration to the last iteration. For the RK4 method, the curve is nearly symmetrical to the central position $t=0$. As the absolute value of t increases, the approximation accuracy deteriorates exponentially. It is the reason that only very small time steps can be adopted and the number of steps required for modeling a complete propagation is large. On the other hand, irrespective of orders used, for a long-term calculation the final result will be divergent. The Chebyshev

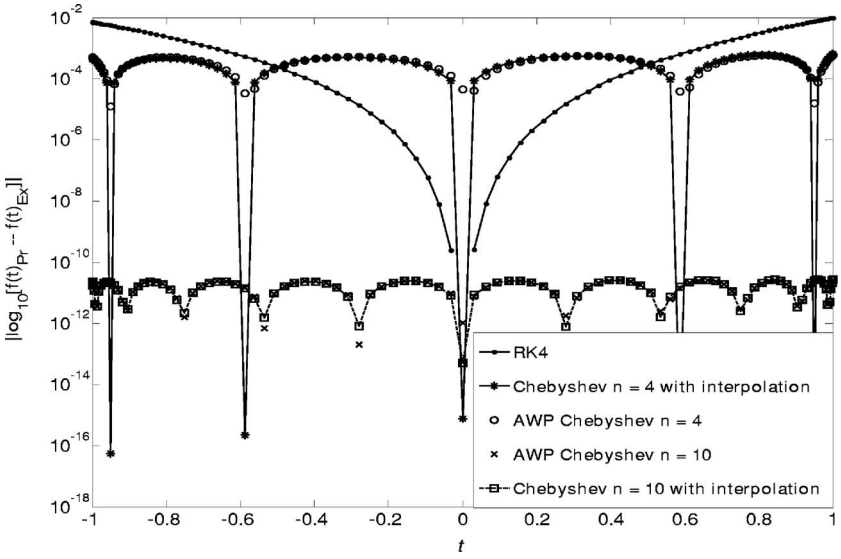


FIG. 2. Absolute errors of the RK4, fourth and tenth Chebyshev polynomial interpolation and extrapolation methods compared with the exact solutions for the function $f(t)=e^{-t}$ based on the variable time-step sizes (the subscripts Pr and Ex represent prediction and exact results).

polynomial expansion overcomes this fatal disadvantage. The calculation is still convergent even when the time is nearly infinite provided that the calculation parameters are properly chosen, such as the time step, the number of grid points, and the order of the Chebyshev polynomial expansion terms.

D. The mapped Chebyshev method

The above-presented Lagrange–Chebyshev interpolation polynomials are different from the traditional Chebyshev pseudospectral method: The following Chebyshev polynomials are used to implement the AWP by $[I_0(R)I+2I_1(R)\hat{H}'_D+2\Sigma_{n=2}^{\infty}I_n(R)T_n(\hat{H}'_D)]$ or $\Sigma_{m=0}^{M}\beta_m J_m(R)\tilde{T}_m(\hat{H}'_D)$. Therefore, the spatial differentiation matrix d_χ is included in $\hat{H}'_D$. In particular, the matrix d_x is not a well-behaved one with eigenvalues scattered in the left-hand side of the complex plane. While most of the eigenvalues grow like $O(N)$, a few of them are $O(N^2)$. This is the reason why, for a traditional Chebyshev pseudospectral method, Δt should be well below $O(N^{-2})$ so that the time marching scheme (the RK4 and Euler methods) will stay in the stable domain. Furthermore, different from the uniform grid size, for Chebyshev–Gauss–Lobatto points, the minimum grid interval $\Delta\chi_{\min}=|\chi_0-\chi_1|=1-\cos(\pi/N)=O(N^{-2})$, which is an important influential factor in choosing the time step. Here, a modified Chebyshev pseudospectral method (also called the mapped Chebyshev method) in Ref. 16 is introduced to keep both the appealing feature in time discretization and original advantage of Lagrange–Chebyshev interpolation polynomial method (for space discretization), and to improve the restrictive stability condition of the former. As a result, the computational efficiency for the previous Lagrange–Chebyshev interpolation polynomials will be improved to some extent. Based on this motivation, a brief introduction of this modified Lagrange–Chebyshev interpolation polynomials is given in the following.

Chebyshev–Gauss–Lobatto points are highly dense near the boundaries with minimal spacing. Since the pseudoespectral method is global, there is no direct relationship between the minimal spacing and the stability condition as in the finite-difference method. However, numerical experience shows that the superfine grid near the boundaries leads to the severe stability condition. Therefore, sharp gradients exist near the boundaries, the highly dense points are needed for resolution and the smaller time step is also needed for physical reason. To overcome the above-noted numerical difficulty, a transform algorithm is applied to map these Chebyshev–Gauss–Lobatto points χ_i to another set of points X_i in the same range $[-1,1]$. Of course, they can be redefined on any original finite internal $[x_0,x_N]$ by means of a linear transform of variable X which maps $[-1,1]$ onto $x\in[x_0,x_N]$, which is discretized at the modified Chebyshev–Gauss–Lobatto points of $X_i=\arcsin(\gamma\chi_i)/\arcsin(\gamma)$, where $\gamma\in]0,1[$ is an optimal parameter; and the notation $[0,1]$ denotes a range, excluding the two end values 0 and 1. Similarly, when $N+1$ grid points in X axis are given, according to the relations between $X\in[-1,1]$ and $x\in[x_0,x_N]$, $\partial\tilde{p}/\partial x$ can be obtained by multiplying the above-noted $\partial\tilde{p}/\partial X$ with the constants $2/(x_N-x_0)$. Differences among $\partial\tilde{p}/\partial x$, $\partial\tilde{p}/\partial\chi$, $\partial\tilde{p}/\partial X$ and three coordinates x (structural coordinate), χ (the coordinate transform with the Chebyshev–Gauss–Lobatto points), and X (the coordinate transform with these modified Chebyshev–Gauss–Lobatto points) should be noted.

The minimal spacing near the boundaries is stretched with larger minimal spacing,

$$X_i=\Re(\chi_i,\gamma). \tag{18}$$

As a result, this mapped method not only reduces the roundoff error but also requires less calculated points compared with previous Lagrange–Chebyshev interpolation polynomials.

Similarly, the spatial derivative of this new function can be obtained by the chain rule,

$$\Re'(\chi_i,\gamma)=\frac{\gamma}{\arcsin(\gamma)\sqrt{1-\gamma^2\chi_i^2}}. \tag{19}$$

Thus, the values at the new grid points $X_i=\Re(\chi_i,\gamma)$, $i=0,1,\ldots,N$ can be calculated by

TABLE I. The effect of N on the parameters $\kappa, \gamma, \underbrace{\Delta X_{\min}(\varepsilon,N)}_{0<\gamma<1}, \underbrace{\Delta X_{\min}(\gamma,N)}_{\gamma\to 0}, \underbrace{\Delta X_{\min}(\gamma,N)}_{\gamma\to 1}$.

N	γ	$O(N^{-2})$	$\Delta X_{\min}(\varepsilon,N)$	$\underbrace{\Delta X_{\min}(\gamma,N)}_{0<\gamma<1}$	$\underbrace{\Delta X_{\min}(\gamma,N)}_{\gamma\to 0}$	$\underbrace{\Delta X_{\min}(\gamma,N)}_{\gamma\to 1}$	$\frac{\Delta X_{\min}(\gamma,N)}{\Delta X_{\min}(0,N)}$	$\frac{\Delta X_{\min}(\gamma,N)}{\Delta X_{\min}(1,N)}$
1	$2.0\times10^{-16}\to0$	1×10^{0}	8.53×10^{-2}	2×10^{0}	2×10^{0}	2×10^{0}	1.0	1.0
2	$2.0\times10^{-8}\to0$	2.5×10^{-1}	4.26×10^{-2}	1×10^{0}	1×10^{0}	1×10^{0}	1.0	1.0
4	$2.0\times10^{4}\to0$	6.25×10^{-2}	2.13×10^{-2}	2.929×10^{-1}	2.929×10^{-1}	5×10^{-1}	1.0	0.5858
8	$2.0\times10^{-2}\to0$	1.56×10^{-2}	1.07×10^{-2}	7.61×10^{-2}	7.61×10^{-2}	2.5×10^{-1}	1.0	0.3044
16	$1.98\times10^{-1}\to0$	3.9×10^{-3}	5.3×10^{-3}	1.95×10^{-2}	1.92×10^{-2}	1.25×10^{-1}	1.0156	0.156
32	$5.75\times10^{-1}\to0$	9.7656×10^{-4}	2.7×10^{-3}	5.5×10^{-3}	4.8×10^{-3}	6.25×10^{-2}	1.1458	0.088
64	$8.545\times10^{-1}\to0$	2.4414×10^{-4}	1.3×10^{-3}	1.9×10^{-3}	1.2×10^{-3}	3.13×10^{-2}	1.5833	0.0607
128	$9.6\times10^{-1}\to0$	6.1035×10^{-5}	6.6620×10^{-4}	8.0093×10^{-4}	3.0118×10^{-4}	1.56×10^{-2}	2.6593	0.0513
256	$9.897\times10^{-1}\to0$	1.5259×10^{-5}	3.3310×10^{-4}	3.6411×10^{-4}	7.5298×10^{-5}	7.8×10^{-3}	4.8356	0.0467
512	$9.974\times10^{-1}\to0$	3.8147×10^{-6}	1.6655×10^{-4}	1.7354×10^{-4}	1.8825×10^{-5}	3.9×10^{-3}	9.2196	0.0445
1024	$9.994\times10^{-1}\to1$	9.5367×10^{-7}	8.3275×10^{-5}	8.8227×10^{-5}	4.7062×10^{-6}	2×10^{-3}	18.7470	$0.0441\to0.0426$

$$\frac{\partial\tilde{p}}{\partial X}=\underbrace{\aleph d_{\chi}}_{d_X}\tilde{p}, \tag{20}$$

where the diagonal matrix $\aleph$ has elements $\aleph_{i,i}=\arcsin(\gamma)\sqrt{1-\gamma^2\chi_i^2}/\gamma$. For simplicity, a new differentiation matrix d_X is defined as the product of $\aleph$ and d_χ.

The parameter γ effectively balances between the accuracy associated with the Chebyshev method and improved stability of the Fourier method. To obtain an effective computation, the parameter γ must be carefully chosen. It should be noted that the parameter γ has significant impact on $\aleph_{i,i}$, $\partial\tilde{p}/\partial X$ and subsequently on the prediction result of the sound pressure $\tilde{p}$. When $\gamma\to0$, $\aleph_{i,i}=\arcsin(\gamma)\sqrt{1-\gamma^2\chi_i^2}/\gamma=\arcsin(\gamma)/\gamma\to1$, the minimal spacing $\Delta X_{\min}=1-\cos(\pi/N)$ is the same as in standard Chebyshev methods $\Delta\chi_{\min}$. However, it is worth noting that, $\gamma\in\,]0,1[$ is an optimal parameter, excluding the two end values 0 and 1. In other words, γ can only indefinitely come nearer to 1, but not $\gamma=1$. It means that $d_X=d_\chi$, the mapped method does not reach its original target. When $\gamma\to1$, $\aleph_{i,i}=\arcsin(\gamma)\sqrt{1-\gamma^2\chi_i^2}/\gamma=\pi\sin(i\pi/N)/2$, the minimal spacing

$$\Delta X_{\min}=|X_{N-1}-X_N|=\underbrace{\left|\arcsin(\gamma\chi_{N-1})/\arcsin(\gamma)-1\right|}_{\gamma\to1}$$
$$=|2\arcsin(\cos(\pi/N))/\pi-1|=2/N,$$

which is the same order as the uniform spacing (Fourier case) Δx $(O(L_x/N))$. In other words, the restriction on the time step related to the stability condition has been removed.

According to the previously defined range, the parameter γ can be expressed as $\gamma=(N^2-\kappa^2)/N^2$, where $\kappa\in\,]0,N[$, while excluding the two end values: 0 and N. Then the minimal spacing $\Delta X_{\min}=2\pi/(N(\sqrt{\pi^2+2\kappa}+\sqrt{2\kappa}))$. When $\kappa=|\ln\varepsilon|^2/2$ and $\gamma=\operatorname{sech}(|\ln\varepsilon|/N)$, the approximation error (the accuracy) is close to ε, which is the machine precision of the computer.

Thus, the minimal spacing $\Delta X_{\min}$ can be calculated by

$$\Delta X_{\min}=\frac{1}{\sqrt{1+|\ln\varepsilon|^2/\pi^2}+\sqrt{|\ln\varepsilon|^2/\pi^2}}\frac{2}{N}\cong\frac{\pi}{N|\ln\varepsilon|}. \tag{21}$$

When ε is fixed, $\Delta X_{\min}$ is only a function of N.

Table I shows the effect of the number of the modified Chebyshev–Gauss–Lobatto points on the parameters

$$\kappa,\gamma,\underbrace{\Delta X_{\min}(\varepsilon,N)}_{0<\gamma<1},\underbrace{\Delta X_{\min}(\gamma,N)}_{\gamma\to0},\underbrace{\Delta X_{\min}(\gamma,N)}_{\gamma\to1},\Delta X_{\min}(\gamma,N)/$$
$$\Delta X_{\min}(0,N)\text{ and }\Delta X_{\min}(\gamma,N)/\Delta X_{\min}(1,N).$$

As shown in the second column, the value of γ changes from 0 to 1 as the number N increases from 1 to 1024. The ratios in the eighth and ninth columns demonstrate the effect of the different values of γ on $\Delta X_{\min}(\gamma,N)$, which is directly related to the time step selected (computational efficiency).

IV. NUMERICAL EXAMPLES AND RESULTS

A. Numerical examples and exact analytical solutions for periodic and nonperiodic boundary conditions

First of all, to demonstrate the above-presented methods, the following modified Gaussian impulse is selected as the initial wave packet with corresponding boundary conditions:

$$p(x,0)=f(x)=0.04^2x^2(x-x_N)^2\exp\left(-\left[\frac{(x-x_c)^2}{4\sigma^2}\right]\right),$$

$$\frac{\partial p(x,0)}{\partial t}=g(x)=0, \tag{22}$$

and

$$\frac{\partial p(x_0,t)}{\partial x}=\frac{\partial p(x_N,t)}{\partial x}=0, \tag{23}$$

where x_c and σ denote the position and Gaussian factor of the initial wave packet, respectively; and to ensure the maximum initial value with positive unit, the sound pressure and its first-order spatial derivative with zero at two ends, the

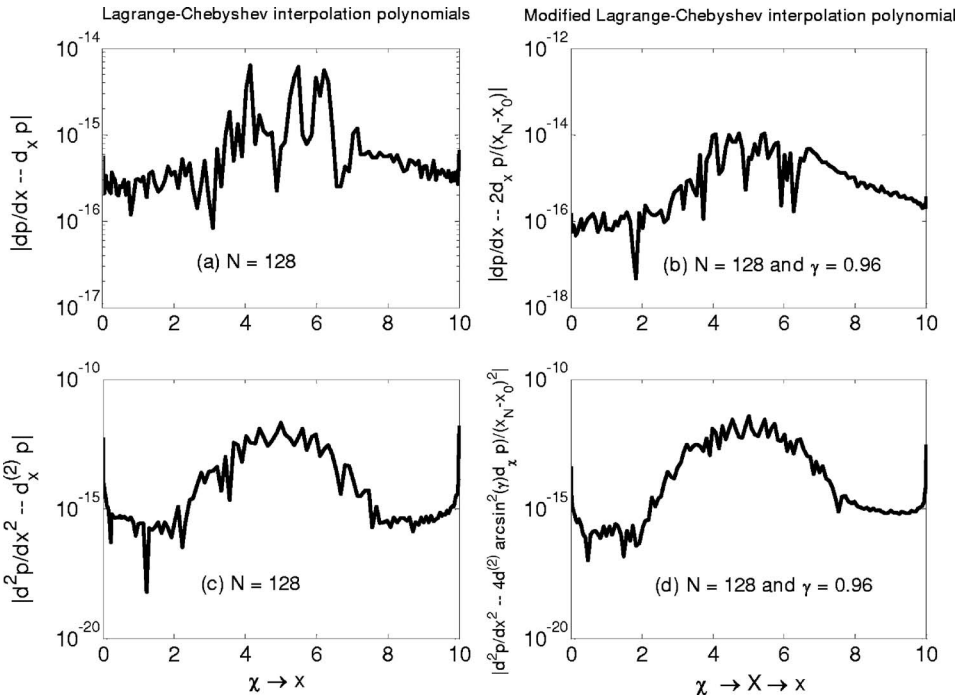


FIG. 3. The first-order and second-order derivatives of sound pressure with Lagrange–Chebyshev interpolation polynomial and modified Lagrange–Chebyshev interpolation polynomial methods.

constant (0.04^2), the terms x^2 and $(x-x_N)^2$ are introduced, respectively.

The exact solution of the sound pressure $p(x,t)=[f(x+ct)+f(x-ct)]/2$ is used for the purpose of comparison. The integral coefficients can be evaluated numerically by numerical integration using MATLAB functions. When $N=16$, the approximation error is the order $O(10^{-14})$, which is the same as that of the specified quadrature accuracy $\delta=1\times10^{-14}$. However, due to the few points used, the curve is not smooth. As N increases ($N\geqslant128$), the approximation error and its variation trend becomes stable. Therefore, the following numerical calculations use $N=128$.

B. Analysis of numerical accuracy and computational efficiency

For the AWP method implementation as in Ref. 20, the Fourier transform scheme was adopted to evaluate the spatial derivative such as $\hat{p}_x=F[\partial p/\partial x]=(jk)F[p(x,t)]=(jk)\hat{p}$, where $\hat{p}$, $\hat{p}_x$, k, and $F[\ ]$ represent the sound pressure and its spatial derivative in the wave number domain, the wave number along the x axis, and the Fourier transform, respectively. Then the inverse Fourier transform is applied to get $\partial p(x,t)/\partial x=F^{-1}\{\hat{p}_x\}$.

For the above-mentioned initial-value problems with nonperiodic boundary condition [a hard wall with $\partial p(x_0,t)/\partial x=0$ imposed on the left-hand side and a pressure-release wall with $p(x_N,t)=0$ imposed on the right-hand side], the Lagrange–Chebyshev interpolation polynomial method with modified Chebyshev–Gauss–Lobatto points is used for calculating the spatial derivatives. Here, only two spatial derivatives $\partial p(x,t)/\partial x$ and $\partial^2 p(x,t)/\partial x^2$ are demonstrated to evaluate this new feature.

The exact solutions of the first-order and second-order derivatives are given by

$$\frac{\partial p(x,t)}{\partial x}=\sum_{m=0}^{\infty}[-c_m\eta_m\sin(\eta_m x)\cos(\zeta_m t)],$$

$$\frac{\partial^2 p(x,t)}{\partial x^2}=\sum_{m=0}^{\infty}[-c_m\eta_m^2\cos(\eta_m x)\cos(\zeta_m t)], \tag{24}$$

where $\eta_m=(m+1/2)\pi/L_x$ and $\zeta_m=c_0(m+1/2)\pi/L_x$.

According to Eq. (22), $f'(x)$,$f''(x)$ can be obtained by

$$f'(x)=\frac{\partial p(x,0)}{\partial x}=0.04^2\left[2x(x-x_N)(2x-x_N)-x^2(x-x_N)^2\frac{(x-x_c)}{2\sigma^2}\right]\exp\left(-\left[\frac{(x-x_c)^2}{4\sigma^2}\right]\right),$$

$$f''(x)=\frac{\partial^2 p(x,0)}{\partial x^2}=0.04^2\left[12x^2-12xx_N+2x_N^2-\frac{(x-x_N)(9x^3-8x^2x_c+4xx_Nx_c-5x^2x_N)}{2\sigma^2}+x^2(x-x_N)\frac{(x-x_c)^2}{4\sigma^4}\right]\exp\left(-\left[\frac{(x-x_c)^2}{4\sigma^2}\right]\right). \tag{25}$$

Figure 3 shows the comparison of the prediction results by the Lagrange–Chebyshev method [(a) $dp/d\chi=d_\chi p$ and (b) $d^2p/d\chi^2=d_\chi^{(2)}p$]; the modified Lagrange–Chebyshev method [(c) $dp/dX=\aleph d_\chi p$ and (d)

$$d^2p/dX^2=d(dp/dX)/dX=\aleph d_\chi(dp/dX)+d_\chi(d\aleph/dX)p=\underbrace{\aleph^2 d_\chi^{(2)}p}_{d_\chi^{(2)}}-\underbrace{\arcsin^2(\gamma)\chi}_{\tilde{d}_\chi^{(1)}}\underbrace{d_\chi p}_{dp/d\chi}$$

and the exact expressions for the initial wave packet given in Eq. (22), where $\aleph^2$ is the square of the diagonal matrix $\aleph$ and

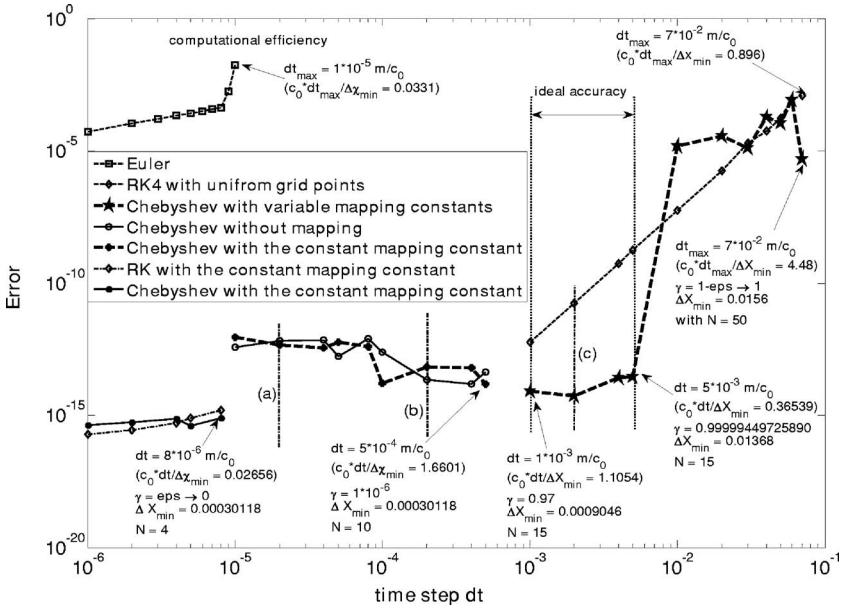


FIG. 4. Errors of the Euler, RK4, and AWP methods with/without the mapped Chebyshev method compared to the exact solutions in different time steps.

$\tilde{d}_x^{(1)}$ is the diagonal matrix with entries $\tilde{d}_{X_{ii}}^{(1)} = \Re''(\chi_i,\gamma)/(\Re'(\chi_i,\gamma))^3 = \arcsin^2(\gamma)\chi_i$. For the first-order and second-order derivatives, the maximum errors of the Lagrange–Chebyshev interpolation polynomial method are 6.3283×10^{-15} and 2.0970×10^{-12}, respectively. For the present modified Lagrange–Chebyshev interpolation polynomial method, as shown in Figs. 3(c) and 3(d), the maximum errors have slightly increased to 1.088×10^{-14} and 3.7788×10^{-12}, respectively. Therefore, the above-presented analysis demonstrates that the modified Lagrange–Chebyshev method keeps the high accuracy for calculating the spatial derivatives. Later discussion will also show that the computational efficiency is also greatly improved at the same time.

The sound pressure inside the duct is calculated hereafter. The main parameters used in this computation are given as follows: The speed of sound $c_0=344$ m/s, the structure sizes are $x_0=0$ m, $x_N=10$ m, $x_c=5$ m, and $\sigma=0.5$. Figure 4 shows the error comparison between the Euler method, the RK4 method, and the Chebyshev method with/without the mapped Chebyshev method. It also shows the effect of γ in the mapped Chebyshev method on the prediction results, in particular the numerical accuracy and computational efficiency. Due to the nonuniform Gauss–Lobatto points, the traditional error evaluation methods (the maximum absolute error, root mean square) are not appropriate. In what follows, the error is defined as the difference between the calculated results and exact results in terms of $\int_{x_0}^{x_N} p(x,t)dx$, where x_N and x_0 represent the upper and lower limits of integration. To this end, the multiple-application trapezoidal rule is used for integration due to the unequal segments ($dx \neq L_x/N$). For discussion purposes, the size of the time step used can be roughly divided into three zones: (a) Small time step $dt \in [1\times10^{-6}, 1\times10^{-5}]m/c_0$; (b) moderate time step $dt \in [1\times10^{-5}, 1\times10^{-3}]m/c_0$; and (c) large time step $dt \in [1\times10^{-3}, 7\times10^{-2}]m/c_0$.

Euler method can only be used in zone one, which requires a very small time step ($dt_{\text{Euler}}=1\times10^{-6}m/c_0$). In particular, as the size of the time step increases, the approximation errors increase linearly. Beyond a certain critical value of $dt=8\times10^{-6}m/c_0$ ($c_0dt/\Delta X_{\min}=0.02648$), the error increases dramatically and the calculation becomes divergent. Overall speaking, the accuracy of the Euler method, even within its validity zone, is still significantly lower compared to other methods. In the same zone, the Chebyshev methods with/without mapping do not show any noticeable difference (not shown in Fig. 4). Figure 4 shows that, in this particular zone, the RK4 method and Chebyshev methods provide comparable calculation accuracy. Generally speaking, as the size of the time step increases, the approximation errors obtained by both methods increase roughly in a linear pattern.

In zone two with moderate time steps, the Chebyshev methods with/without mapping are shown to demonstrate the effect of γ on the calculation errors. The impact of performing mapping starts to be obvious. With the increase of the time step, error curves do not necessarily undergo monotonous increase, suggesting a possible optimization on the combination of the expansion term n and γ to achieve high numerical accuracy.

The effect of the uniform grid points and Chebyshev–Gauss–Lobatto points (especially the modified Chebyshev–Gauss–Lobatto points by the mapped Chebyshev method) on the numerical accuracy and computational efficiency can be clearly seen in zone three. Within this zone, it is observed that the previous Lagrange–Chebyshev method without the mapped Chebyshev method or any higher-order RK methods (>4) fail to provide converged results. Therefore, Fig. 4 only compares the proposed method to the RK4 method with the uniform grid points. Figure 4 shows that the maximum time step allowed by modified Chebyshev–Gauss–Lobatto points by the mapped Chebyshev method can be up to $dt_{\max}=7.0\times10^{-2}m/c_0$ ($c_0dt_{\max}/\Delta x_{\min}=0.896$). Within this zone, however, there exists an optimal range in which modified Chebyshev–Gauss–Lobatto points by the mapped Chebyshev method outperforms the RK4 method by providing a much better accuracy for the same time step used. This range can be determined by properly choosing the expansion term n

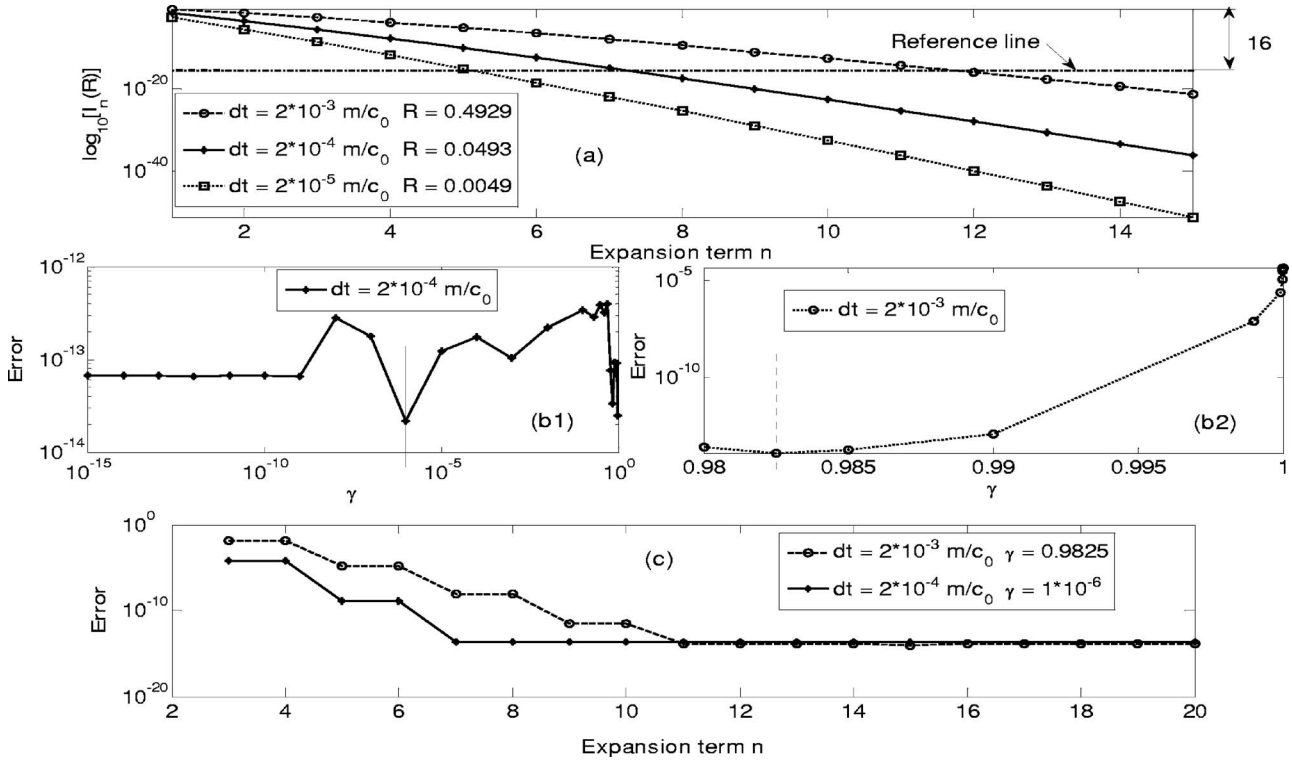


FIG. 5. Effects of the expansion term n and γ on the prediction error.

and γ, such as the following two sets of combinations: (1) $dt=1.0\times10^{-3}m/c_0$ with $\gamma=0.97$ and $n=15$ and (2) $dt=5.0\times10^{-3}m/c_0$ with $\gamma=0.99999449725890$ and $n=15$, as shown in the right-hand lower corner of Fig. 4. Therefore, the proposed technique allows the use of relatively large time steps while maintaining the good calculation accuracy, in which circumstance all other investigated methods fail.

An appealing feature of the Chebyshev polynomial method lies in its ability to prevent accumulation of the truncation errors. Given a time step, the RK method, however, accumulates the truncation error as the selected expansion term increases. It is the reason why both the Euler and RK4 methods are not quite suitable for a long-term calculation. For the same reason, only lower-order RK methods are widely used. The above-presented analysis shows that the modified AWP scheme (double Chebyshev methods both in temporal operator and spatial derivatives) has much better performance in terms of both numerical accuracy and computational efficiency due to the use of larger time step.

The allowable maximum time step $dt_{\max}$ is a crucial parameter governing the numerical accuracy and computational efficiency, as shown in Fig. 4. The effect of the expansion term n and γ on the calculation error based on the modified Lagrange–Chebyshev method with the mapped Chebyshev method is investigated. Three typical time steps are selected from each of the three zones mentioned earlier (Fig. 4): (a) $dt=2.0\times10^{-5}m/c_0$; (b) $dt=2.0\times10^{-4}m/c_0$; and (c) $dt=2.0\times10^{-3}m/c_0$, respectively. The number of the expansion term n in the Chebyshev method is first examined in terms of the coefficients of the first kind of Bessel function $I_n(R)$. Currently, the state-of-the-art computer provides a dynamic range of about 10^{-16}. We attempt to determine the minimum expansion terms needed to ensure that the truncation errors do not contribute to the final result and the sum of the polynomials converges to the order of $I_{n_{\min}}(R)$. The convergence properties of $I_n(R)$ for three given R values (0.0049, 0.0493, 0.4929) calculated from the three time steps mentioned earlier are illustrated in Fig. 5(a). It can be seen that $I_n(R)$ decreases monotonically with the increase of n. For a threshold value $I_0(R)/I_{n_{\min}}(R)=10^{16}$, there exists a minimum value of $n(n_{\min})$. Small R corresponds to a small $n_{\min}$. With a reference line defined by $|\log_{10}^{I_{\text{ref}}(R)/I_0(R)}|=16$, Fig. 5(a) shows that $n_{\min}$ takes the value of 4, 6, and 10 for $R=0.0049$, 0.0493, and 0.4929, respectively. For safety's sake, however, we use $n=10$ and 15 for $R=0.0493$ and 0.4929, respectively. Furthermore, when R takes other values within the range of 0.0049–0.4929, corresponding $n_{\min}$ value can be estimated by interpolation using Fig. 5(a).

The effect of γ on the prediction error is shown in Figs. 5(b1) and (b2), using two time steps ($dt=2.0\times10^{-4}m/c_0$, $dt=2.0\times10^{-3}m/c_0$). With a moderate time step $dt=2.0\times10^{-4}m/c_0$, Figure 5(b1) shows that γ has no effect on the calculation error when its value is small enough ($<1\times10^{-9}$). As γ further increases, γ shows strong influence on the calculation error with several "local minima," suggesting a possible optimization on γ to achieve the highest computation accuracy. Generally speaking, with this γ range (from $\gamma=1\times10^{-15}$ to $\gamma\to1$), the prediction results are found to be highly accurate [the order of error below $O(10^{-12})$]. With a larger time step $dt=2.0\times10^{-3}m/c_0$, Fig. 5(b2) shows that the prediction error remains rather low [$O(10^{-12})$] within a very narrow range from $\gamma=0.98$ to $\gamma=0.985$. Exceeding this range, the prediction accuracy rapidly deteriorates and the

calculation error increases drastically with the increase of γ. The above-presented analysis demonstrates that constant $\gamma = 1$ is not a good choice. Therefore, the value of γ needs to be carefully chosen in order to ensure the accuracy, or even the convergence of the calculation. When the time step dt and the parameter γ are fixed, the effect of the expansion term n on the prediction error is also shown in Fig. 5(c). It can be seen that, as n increases, the prediction error gradually improves until a certain value.

V. CONCLUSIONS

An AWP technique with the mapped Chebyshev method is proposed to describe the time-domain evolution of acoustical waves. Using a numerical example in a duct structure, the numerical accuracy, computational efficiency, and stability of the technique are investigated, leading to the following conclusions:

(1) Drawbacks and limitations of fast Fourier transform and the existing Chebyshev–Fourier scheme in dealing with nonperiodic boundaries are surmounted by the proposed combined scheme: A Chebyshev polynomial expansion scheme in the temporal AWP operator and the modified Lagrange-Chebyshev interpolation polynomials scheme with the mapped Chebyshev method in the spatial derivatives evaluation. The latter not only keeps the high accuracy for calculating the spatial derivatives, but also significantly improves the computational efficiency. Meanwhile, a mathematical model with nonperiodic boundary conditions is introduced in the spatial derivatives, allowing the consideration of any boundary conditions.
(2) For large R values, the J expansion is better because large prediction errors can be avoided when some extra terms fall outside of the dynamic range defined by the machine accuracy of the computer (10^{-16}). For small R values, the previous I expansion is recommended due to its exponential decaying property.
(3) The time step restriction due to the high-density grid near the boundaries with minimal spacing has been overcome by introducing the mapped Chebyshev method. Three zones are observed in which the proposed method shows different characteristics with respect to other existing methods. In zone one with very small time steps, the Chebyshev methods with/without mapping do not show any noticeable difference. Apart from the Euler method, other conventional numerical methods studied in this paper provide very comparable numerical accuracy. In zone two with moderate time steps, the effect of γ (is an optimal parameter $\gamma \in]0,1[$, excluding the two end values 0 and 1) on the Chebyshev methods with mapping is obvious. With the increase of the time step, error curves do not necessarily undergo monotonous increase, suggesting a possible optimization on the combination of the expansion term n and γ to achieve high numerical accuracy. In zone three with large time steps, γ dominates the numerical accuracy and computational efficiency. By properly choosing the expansion term n and γ, an optimal result with high numerical accuracy and computational efficiency with much larger time step used can be obtained. In this time step zone, neither the previous Lagrange–Chebyshev method without the mapped Chebyshev method nor the high–order RK methods can be used.

The numerical analyses carried out in this paper demonstrate that the modified Lagrange–Chebyshev method with the mapped Chebyshev method can satisfactorily handle the nonperiodic boundary conditions and initial-value problems. The proposed method provides a good combination of numerical accuracy, computational efficiency, and stability, in particular for long-term calculations. The work presented in this paper provides significant improvement to the existing AWP method, opening doors to a large number of engineering problems. Typical examples may include problems involving fluid–structural interactions and physical systems exhibiting strong nonlinear behavior. The framework established in this paper allows straightforward extension to the first type of problems, while the second one is much more challenging and requires further in-depth investigations.

ACKNOWLEDGMENTS

The authors are grateful to Dr. L. Huang for his help in numerical analysis, especially his original contribution to the Lagrange–Chebyshev method developed for the nonperiodic boundary condition. The authors also benefited from valuable discussions with Professor J. Pan and Dr. K. S. Sum. S.Z.P. is thankful for financial support from the Postdoctoral Fellowship Scheme of the Hong Kong Polytechnic University (G-YX62).

APPENDIX: THE RK4 METHOD FOR A SIMPLE FUNCTION $F(T)=E^{-T}$

For the function $f(t)=e^{-t}$, there are two kinds of Chebyshev polynomial expansions: (a) Interpolation approximation and (b) extrapolation expansions. For the former, the Chebyshev polynomial expansion $f_{\mathrm{Ch}}(t)$ of degree n for $f(t)$ over the interval $[-1,1]$ can be written as a sum of $T_j(t)$:

$$f(t) \approx f_{\mathrm{Ch\_Inter}}(t) = \sum_{j=0}^{n} c_j T_j(t). \tag{A1}$$

The coefficients c_j are computed with the following formulas:

$$c_0 = \frac{1}{n+1}\sum_{k=0}^{n} f(t_k)T_0(t_k) = \frac{1}{n+1}\sum_{k=0}^{n} f(t_k); \quad c_1 = \frac{2}{n+1}\sum_{k=0}^{n} f(t_k)T_1(t_k) = \frac{2}{n+1}\sum_{k=0}^{n} f(t_k)t_k \tag{A2}$$

and

$$c_j = \frac{2}{n+1}\sum_{k=0}^{n} f(t_k)T_j(t_k) = \frac{2}{n+1}\sum_{k=0}^{n} f(t_k)(2t_kT_{j-1}(t_k) - T_{j-2}(t_k)), \tag{A3}$$

where $j=2,\ldots,n$ and t_k denote the Chebyshev interpolation nodes calculated by $t_k=-\cos((2k+1)\pi/(2n+2))$ for $k=0,1,2,\ldots,n$. Strictly speaking, the coefficients c_j are obtained by borrowing the future results of the function $f(t_k)$, as shown in Eq. (A3).

Thus, when $n=4$ (taken the same terms as that in the RK4 method), Eq. (A1) can be given by

$$\begin{aligned} f_{\text{Ch4\_Inter}}(t) = {} & \underbrace{1.2660658}_{c_0} - \underbrace{1.1303182}_{c_1}t + \underbrace{0.2714951}_{c_2}(2t^2 - 1) - \underbrace{0.0443337}_{c_3}(4t^3 - 3t) + \underbrace{0.0054293}_{c_4}(8t^4 - 8t^2 + 1) \\ = {} & 1.0 - 0.9973173t + 0.4995562t^2 - 0.1773346t^3 + 0.0434341t^4. \end{aligned} \tag{A4}$$

For the latter,

$$f(t) \approx f_{\text{Ch}}(t) = \left[I_0(R)T_0\left(\underbrace{t-t_0}_{\hat{H}'}\right) + 2I_1(R)T_1(t-t_0) + \sum_{m=2}^{M} 2I_m(R)T_m(t-t_0) \right] f(t_0),$$

where $I_m(R)$ is the mth-order Bessel function of the first kind. This simple function $f(t)=e^{-t}$ can be simplified from $f(t)=e^{-R\hat{H}'}$ provided that $R=1$, $\hat{H}'=t-t_0$ and $t_0=0$. $T_0(t)=1$, $T_1(t)=t$, and the rest can be calculated by the following recursive relations: $T_{m+1}(t)=2tT_m(t)-T_{m-1}(t)$. Different from the previous interpolation expansions, it is worth noting that the coefficients $I_m(R)$ and $T_m(t)$ are calculated by known results at present time step.

[1] J. Dollimore, "Some algorithms for use with the fast Fourier transform," J. Inst. Math. Appl. **12**, 115–117 (1973).
[2] J. P. Coyette, "Transient acoustics: Evaluation of finite element and boundary element methods," Proceedings of ISMA 19, Leuven, Belgium, September 12–14, 1994, pp. 223–234.
[3] R. Renaut and J. Frohlich, "A pseudospectral Chebyshev method for the 2D wave equation with domain stretching and absorbing boundary conditions," J. Comput. Phys. **124**, 324–336 (1996).
[4] F. Q. Hu, M. Y. Hussaini, and J. L. Manthey, "Low-dissipation and low-dispersion Runge-Kutta schemes for computational acoustics," J. Comput. Phys. **124**, 177–191 (1996).
[5] Z. Jackiewicz and B. D. Welfert, "Stability of Gauss-Radau pseudospectral approximations of the one-dimensional wave equation," SIAM J. Sci. Comput. (USA) **18**, 287–313 (1996).
[6] K. Burrage and E. Platen, "High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations," Appl. Numer. Math. **22**, 81–101 (1996).
[7] K. Y. Fung, "Time-domain computation of acoustics in confinements," Proceedings of the Fifth International Congress on Sound and Vibration, Adelaide, Australia, December 15–18, 1997, pp. 1839–1847.
[8] J. B. Schneider and O. M. Ramahi, "The complementary operators method applied to acoustic finite-difference time-domain simulations," J. Acoust. Soc. Am. **104**, 686–693 (1998).
[9] D. R. Hans, K. Michielsen, J. S. Kole, and M. T. Figge, "Solving the Maxwell equations by the Chebyshev method: A one-step finite-difference time-domain algorithm," IEEE Antennas Propag. Mag. **51**, 3155–3160 (2003).
[10] A. Gelb, Z. Jackiewicz, and B. Welfert, "Absorbing boundary conditions of the second order for the pseudospectral Chebyshev methods for wave propagation," SIAM J. Sci. Comput. (USA) **17**, 501–512 (2002).
[11] M. D. Feit, J. A. Fleck, and A. Steriger, "Solution of the Schrödinger equation by a spectral method," J. Comput. Phys. **47**, 412–433 (1982).
[12] N. Balakrishnan, C. Kalyanraman, and N. Sathyamurthy, "Time-dependent quantum mechanical approach to reactive scattering and related processes," Phys. Rep. **280**, 79–144 (1997).
[13] H. Tal-Ezer and R. Kosloff, "An accurate and efficient scheme for propagating the time-dependent Schrödinger equation," J. Chem. Phys. **81**, 3967–3971 (1984).
[14] R. Kosloff and D. Kosloff, "Absorbing boundaries for wave propagation problems," J. Comput. Phys. **104**, 363–376 (1985).
[15] R. Kosloff, "Time-dependent quantum-mechanical methods for molecular dynamics," J. Phys. Chem. **92**, 2087–2100 (1988).
[16] D. Kosloff and H. Tal-Ezer, "Modified Chebyshev pseudospectral method $O(N^{-1})$ with time step restriction," J. Comput. Phys. **104**, 457–469 (1993).
[17] W. S. Don and A. Solomonoff, "Accuracy enhancement for higher derivatives using Chebyshev collocation and a mapping method," SIAM J. Sci. Comput. (USA) **18**, 1040–1057 (1995).
[18] C. Leforestier, R. H. Bisseling, C. Cerjan, M. D. Feit, R. Friesner, A. Guldberg, A. Hammerich, G. Jolicard, W. Karrlein, H.-D. Meyer, N. Lipkin, O. Roncero, and R. Kosloff, "A comparison of different propagation schemes for the time domain dependent Schrödinger equation," J. Comput. Phys. **94**, 59–80 (1991).
[19] J. Pan and J. B. Wang, "Acoustical wave propagator," J. Acoust. Soc. Am. **108**, 481–487 (2000).
[20] S. Z. Peng and J. Pan, "Acoustical wave propagator for time-domain flexural waves in thin plates," J. Acoust. Soc. Am. **115**, 467–474 (2004).
[21] S. Z. Peng and J. Pan, "A study of time-domain stress concentration in a plate with sharp change of section using the acoustical wave propagator method," J. Acoust. Soc. Am. **117**, 492–502 (2005).
[22] S. Z. Peng, "Dynamic stress concentration in a ribbed plate using acoustical wave propagator method," J. Sound Vib. **279**, 75–88 (2005).
[23] S. Z. Peng and L. Huang, "The improved acoustical wave propagator method for predicting time-domain acoustical wave propagation in a duct structure," Proceedings of the Ninth Western Pacific Acoustics Conference, Seoul, Korea, June 26–28, 2006.
[24] S. C. Chapra and R. P. Canale, *Numerical Methods for Engineers* (McGraw-Hill, Boston, 2006).

Three wave mixing test of hyperelasticity in highly nonlinear solids: Sedimentary rocks

R. M. D'Angelo, K. W. Winkler, and D. L. Johnson[a)]
Schlumberger-Doll Research, One Hampshire St., Cambridge, Massachusetts 02139

(Received 23 April 2007; accepted 11 November 2007)

Measurements of three-wave mixing amplitudes on solids whose third order elastic constants have also been measured by means of the elasto-acoustic effect are reported. Because attenuation and diffraction are important aspects of the measurement technique results are analyzed using a frequency domain version of the KZK equation, modified to accommodate an arbitrary frequency dependence to the attenuation. It is found that the value of β so deduced for poly(methylmethacrylate) (PMMA) agrees quite well with that predicted from the stress-dependent sound speed measurements, establishing that PMMA may be considered a hyperelastic solid, in this context. The β values of sedimentary rocks, though they are typically two orders of magnitude larger than, e.g., PMMA's, are still a factor of 3–10 less than those predicted from the elasto-acoustic effect. Moreover, these samples exhibit significant heterogeneity on a centimeter scale, which heterogeneity is not apparent from a measurement of the position dependent sound speed.
 [DOI: 10.1121/1.2821968]

PACS number(s): 43.25.Dc, 43.25.Ba, 43.25.Cb [MFH] Pages: 622–639

I. INTRODUCTION

The simplest and most commonly investigated aspects of nonlinear acoustics are (a) the elasto-acoustics (EA) effect, in which one measures the rate of change of sound speed with applied stress and (b) three wave mixing (3WM), in which two incident waves at two different frequencies, f_1 and f_2, mix to generate a third frequency component at $f_2 - f_1$, $f_1 + f_2, 2f_1$, and $2f_2$. These effects are described in Hamilton and Blackstock (1998). Within the context of hyperelasticity both these processes are governed by the same nonlinear material parameters, called third order elastic (TOE) constants if one is dealing with a solid. We shall use the term hyperelasticity, sometimes called classical nonlinearity, to mean that there is an underlying deformation energy which is a well-defined function of the strain tensor (Truesdell and Noll, 1965; Eringen and Şuhubi, 1974). (In any real material there may also be additional mechanisms that give rise to dispersion and attenuation of acoustic waves. If these effects are small then one may say the material is hyperelastic to a first approximation.) The hyperelasticity hypothesis is a very basic assumption that is generally made, often implicitly, in discussing nonlinear effects.

Specifically, the strain energy of a hyperelastic solid is path independent. One may first apply a compression and then a shear, or vice versa, or apply them simultaneously. If the work done on the system depends only upon the final strain, and not the order, the system is said to be path independent. A counter example, in which the system is nonlinearly elastic along a given deformation path but the strain energy depends upon the deformation path taken, is given by a collection of loose unconsolidated solid grains (Norris and Johnson, 1997; Johnson and Norris, 1997). Likewise, the hyperelastic part of the strain energy does not depend upon the rate at which the final strain is achieved. It goes without saying that the hyperelasticity hypothesis does not allow for hysteresis in the stress-strain relation.

For such a hyperelastic material it is useful to expand the deformation energy in powers of the strain tensor. The coefficients of the first order terms (in strain) are simply the static stresses. The coefficients of the terms second order in strain, known as the second order elastic constants, give rise to a linear relationship between differential stress and differential strain. These elastic constants also predict the speeds of small amplitude sound waves and predict them to be independent of frequency and of wave amplitude for small enough amplitudes. One may say that many solids, including rocks, obey the hyperelasticity hypothesis at this second order, at least approximately, in the sense that:

(1) For "small enough" strain amplitudes, the measured speeds of sound are observed to be independent of strain amplitude.

(2) The measured speeds of sound are not very frequency dependent, even over a wide frequency range. Such dispersion effects that are seen are accounted by means of specific mechanisms, additional to hyperelasticity, such as, e.g., "the squirt effect" in sedimentary rocks (Dvorkin *et al.*, 1995).

(3) For "small enough" stress-strain cycles there is no accumulated hysteresis. An acoustic experiment, using frequencies above 1 kHz, and strain amplitudes on the order of 10^{-8}, may run for a considerable length of time with no measurable change in the acoustic properties of the sample. Such a measurement will have incorporated thousands if not millions of small amplitude stress-strain cycles.

(4) The numerical values of the elastic constants deduced in one context may be transferred to another. The measured compressional and shear speeds, say, allow one to compute the speed of an extensional mode in a bar, for ex-

[a)]Author to whom correspondence should be addressed. Electronic mail: johnson10@slb.com

ample, or in other complicated geometries. This property speaks to the issue of path independence of the strain energy through second order in the strain.

These properties, shared by most solids, have been so well established for so long that it is not questioned that most solids do obey the hyperelasticity hypothesis at the second order. In most textbooks on elasticity the hypothesis is not even questioned. At the third order and higher, such is not the case, however. At the fourth order the hypothesis is demonstrably not true for sedimentary rocks. (See below.) In this article we examine whether it is satisfied at the third order.

An isotropic hyperelastic solid has three TOE coefficients, A, B, and C. A straightforward discussion of the underlying theory is given by Norris (1998) who makes connection with other systems of notation. In this regard, sedimentary rocks are particularly interesting because their sound speeds are known to be more sensitive to applied stress than are homogeneous solids, by 2–3 orders of magnitude. This fact makes measurements of *in situ* acoustic properties useful as a way of determining information about the state of stress in rock formations (Sinha *et al.*, 2000; Tang and Patterson, 2001). In order to determine the actual numerical value of the stress from the sound speed measurements one needs to have an independent measurement of the TOEs, or some relevant combination thereof. A candidate technique would be to measure the efficiency of 3WM effects, but this would presuppose that both the EA effect and the 3WM effect are governed by the same nonlinear parameters A, B, and C, i.e., that the hyperelastic hypothesis is satisfied at third order. This article is devoted to an investigation of whether the measured values of the TOE constants of rocks can be used quantitatively to compute the amplitudes of three wave mixing signals.

There are other nonlinear effects which are quite significant in sedimentary rocks. We mention a few: (1) Hysteresis in a static stress-strain measurement. (2) The shift of the resonant frequency in a resonant bar experiment as the amplitude of the standing wave is increased. (3) The efficiency of the generation of the third harmonics such as $3f_1$, etc. (4) Amplitude dependent attenuation. The evidence in the literature is that classical nonlinearity fails to describe the results of these experiments on sedimentary rocks: (1) Classical nonlinearity does not admit of any static hysteresis. (2) It has been observed that the shift in the resonant bar frequency is proportional to the amplitude of the wave, whereas classical nonlinearity predicts that it should be proportional to the square of the amplitude. (3) Classical nonlinearity predicts that the amplitude of the third harmonic should be proportional to the cube of the fundamental amplitude, whereas experiments indicate it is proportional to the square. This is a direct indication of the breakdown of hyperelasticity at the fourth order in the strain tensor. (4) There is no universally accepted "classical nonlinear" theory of the amplitude dependence of the attenuation. These results have motivated researchers to develop more sophisticated nonlinear acoustic theories in which the underlying hysteretic elements in the system are accounted at the beginning (Nazarov *et al.*, 1988; Zimenkov and Nazarov, 1994). The so-called Presisach-Mayergoyz (PM) protocol developed by Guyer and coworkers (Guyer *et al.*, 1997, and references therein), in particular, allows one to deduce the relevant distribution of hysteretic elements and predict the outcome of these various experiments. Qualitatively, this PM-based theory does account for the observed nonclassical nonlinear behavior seen in rock samples. The situation, theoretical and experimental, is reviewed by Guyer and Johnson (1999); Ostrovsky and Johnson (2001).

It should be mentioned, however, that very recent results cast doubt on the claimed failure of the classical nonlinear theory as regards the shift of the resonant bar frequency with amplitude in sedimentary rocks. TenCate *et al.* (2004) and Pasqualini *et al.* (2007) have redone these experiments and have carefully reanalyzed previous experiments. They conclude that as long as the sample is kept in the reversible, nonlinear regime the shift in the resonance frequency is indeed proportional to the square of the amplitude, as would be predicted by the classical theory. Here, the relevant classical nonlinear theory would involve the fourth order elastic constants, not just the third order ones. For strains that exceed a specific, sample dependent value, ϵ_M, the sample shows irreversible behavior.

Notwithstanding the foregoing discussion it seems to be the case that the amplitudes of the dominant 3WM signals in sedimentary rocks (e.g., $2f_1$ if there is only a single fundamental, f_2-f_1 in a 3WM experiment) scale quadratically with the amplitude of the fundamental(s) (Nazarov *et al.*, 1988; Johnson and Shankland, 1989; Johnson *et al.*, 1996). Moreover, in a one-dimensional propagation experiment the amplitude of the second harmonic is seen to grow linearly with propagation distance (Meegan *et al.*, 1993). These facts are in accord with the classical theory, based on the hyperelasticity assumption; the relevant nonlinear parameters would be the TOE constants. It makes sense to ask the question: Do the TOE constants measured in an elasto-acoustic experiment accurately predict the amplitude of the generation of these 3WM waves? Surprisingly, until very recently this question had not been addressed for any solid (Jacob *et al.*, 2003; D' Angelo *et al.*, 2004). In order to answer that question one must use calibrated transducers in an experimental geometry simple enough that the relevant calculations can be done. Here, we report a continuation of our earlier results (D' Angelo *et al.*, 2004), extended this time to dry as well as water-saturated samples. We find that a sample of poly(methylmethacrylate) (PMMA) is rather well described by the classical theory of nonlinear acoustics, based on the hyperelasticity assumption. That is, for this sample of PMMA, the amplitudes of f_2-f_1, f_1+f_2, $2f_1$, and $2f_2$ can be accurately predicted from the measured third order elastic constants. The rock samples generally exhibit a regime in which the fundamentals, f_1 and f_2, behave linearly and the nonlinear signals, f_2-f_1, f_1+f_2, $2f_1$, and $2f_2$, behave quadratically (as a function of the fundamental's amplitude) but the values of the so-deduced nonlinear amplitudes are smaller by a factor of 3–10 from those expected from the EA measurements. Moreover, these samples are significantly heterogeneous, on a one centimeter scale, which certainly complicates the interpretation.

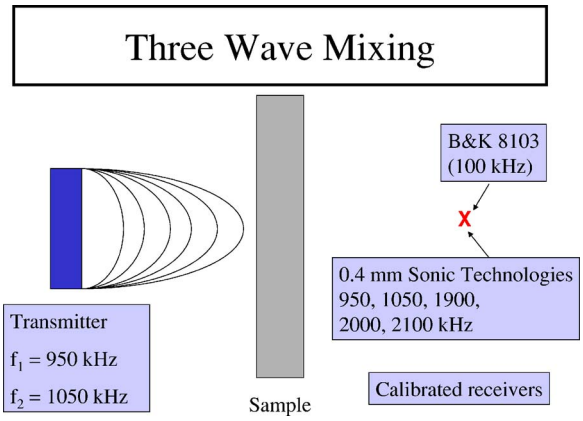


FIG. 1. (Color online) Schematic of experiment for measuring three-wave mixing amplitudes.

The organization of the paper is as follows: In Sec. II we describe our experimental technique, both with respect to the three wave mixing experiments as well as the elasto-acoustic measurements. In Sec. III we describe the modified KZK equation and our means of solving it in our experimental geometry. The results are presented in Sec. IV. We summarize our conclusions in Sec. V.

II. EXPERIMENTAL TECHNIQUE

A. Three wave mixing measurements

We performed our three wave mixing measurements in water using a high-power source transducer and calibrated receivers in a tank approximately 2 m×1 m×1 m deep. The source signal comprised the beating, for 50 μs, of two sinusoidal components, one at f_1=0.95 MHz and the other at f_2=1.05 MHz. The drive signal (f_1&f_2) voltage was varied in precisely known steps, such that the transducer face pressure ranged from approximately 1 to 1000 kPa. We always maintain a constant ratio of the amplitudes of the two drive signals: $A(1.05\ \text{MHz})=0.6119*A(0.95\ \text{MHz})$. Subsequently, as we discuss varying the amplitude of the fundamentals, it will be understood that this amplitude ratio is held fixed. The source transducer was an air-backed, 1.0 MHz, 1.0-in.diam. planar-piston transducer.

We positioned receivers at a distance of (typically) 203 mm from the source face, on axis with the transmitter beam. One receiver was a B&K 8103, used for measuring the difference frequency (100 kHz). The other was a 0.4 mm Sonic Technologies membrane-hydrophone probe, used for all of the measurements around and above 1 MHz. Both receivers were calibrated by the manufacturers in Pa over wide frequency ranges. See Fig. 1 for a schematic view of the experiment. We did utilize a variety of different standoff values and transmitter-receiver separations, but we needed to ensure that the receivers are not located near the diffraction nodes and that we are not plagued by multiple reflections at interfaces which interfere with the direct signals. Our measurements are all taken in the far field of the transmitting transducer.

With our calibrated receivers in a water-only setup (β =3.5) we used the KZK calculations (Sec. III, below) to calibrate the transmitter pressures at each drive voltage. We did this by quantifying the axial and radial variations of all received signals. This test also has the effect of ruling out any significant contamination of 3WM frequencies being generated either in the electronics or in the transmitter; either of these possibilities would not contribute to nonlinear signals that grow with transmitter-receiver separation.

Then when we insert a sample in the acoustic path, as indicated in Fig. 1, we know the initial amplitudes of the fundamentals, f_1 and f_2, at the transmitter face, for each value of the voltage amplitude of the drive signal. The measured received signals, calibrated in Pascales (Pa), are analyzed in the context of the modified KZK equation in order to determine an estimate of β, the nonlinear material parameter of interest.

We used standard immersion techniques to derive the linear velocity-dispersion and attenuation-dispersion curves for each sample, which we measured over as wide a frequency range as possible, depending on the sample. We mea-

TABLE I. Values of relevant parameters for three air-saturated sedimentary rocks and for PMMA, a commercially available plastic. We measured the density, speed, attenuation, and the third-order elastic constants, as described in the text. α_Δ, $\alpha_{1.0}$, and $\alpha_{2.0}$ refer to the nominal attenuation values at 100 kHz, 1.0 MHz, and 2.0 MHz, respectively. $\beta(stress)$ is computed from Eq. (1) using the elasto-acoustic measurements and $\beta(3wm)$ is determined from the analysis of our three-wave mixing data. The indicated range of numerical values for the latter reflects the heterogeneous nature of the samples.

	Water	PMMA	Portland Sandstone	Indiana Limestone	Berea Sandstone
ρ(gm/cc)	1.0	1.19	2.18	2.17	2.00
V(km/s)	1.48	2.72	3.35	3.25	1.72
$\alpha_{1.0}$(1/m)	2.5×10^{-2}	12.4	19.4	53.1	190
α_Δ(1/m)	2.5×10^{-4}	1.55	1.38	3.34	0.96
$\alpha_{2.0}$(1/m)	1.0×10^{-1}	23.0	27.6	66.8	1200
$-A$(GPa)	⋯	22±6	1672±119	3943±356	5011±395
$-B$(GPa)	⋯	18±3	161±52	5643±205	1692±11
$-C$(GPa)	⋯	19±3	134±78	5275±573	158±31
$\beta(stress)$	⋯	9±2	124±19	849±50	1193±28
$\beta(3wm)$	3.5	7.1±0.5	[150, 200]	[−50, +100]	[300, 600]

TABLE II. Values of relevant parameters for three water-saturated sedimentary rocks and for PMMA, a commercially available plastic. Same conventions as in Table I.

	Water	PMMA	Portland Sandstone	Indiana Limestone	Berea Sandstone
ρ(gm/cc)	1.0	1.19	2.33	2.40	2.26
V(km/s)	1.48	2.72	3.47	3.89	2.45
$\alpha_{1.0}$(1/m)	2.5×10^{-2}	12.4	64.0	77.0	288
α_{Δ}(1/m)	2.5×10^{-4}	1.55	6.25	0.89	14.3
$\alpha_{2.0}$(1/m)	1.0×10^{-1}	23.0	129	295	711
$-A$(GPa)	⋯	22 ± 6	2362 ± 147	6559 ± 313	-2234 ± 395
$-B$(GPa)	⋯	18 ± 3	1310 ± 106	11715 ± 286	4939 ± 269
$-C$(GPa)	⋯	19 ± 3	1846 ± 210	9377 ± 644	8787 ± 517
$\beta(stress)$	⋯	9 ± 2	332 ± 28	1201 ± 43	1155 ± 93
$\beta(3wm)$	3.5	7.1 ± 0.5	[60, 150]	[400, 525]	[170, 670]

sured the sample densities. In Table I we list values of the relevant parameters that we measured on three air-dried sedimentary rocks, as well as a sample of PMMA, a commercially available plastic. Table II lists those values for the water-saturated samples. The symbols α_{Δ}, α_1, and α_2 refer to the measured attenuation coefficients at 100 kHz, 1.0 MHz, and 2.0 MHz, respectively. However, in our analysis of the nonlinear acoustics, for which we need accurate values of the attenuation, we have adjusted, slightly, the input values for the attenuation at 950 kHz and 1050 kHz for the rock samples so as to match the measured amplitudes of these two fundamentals. This was not necessary for the PMMA sample.

The samples were then located in the beam of the high-power transmitter, typically at a stand-off of 50.8 mm from the transmitter face. We varied the drive signal $(f_1 \& f_2)$ voltage in the same steps as for the water-only test, recording the received tone bursts for both high and low frequency receivers. The received signals were then processed into calibrated pressure values at each pertinent frequency, as above.

B. Stress-dependent sound speeds

The third order elastic constants for a given material can be deduced by measuring the rate at which the sound speeds vary under the influence of applied stress: $\lim_{\sigma\to0}\frac{dV}{d\sigma}$, where V may be either a compressional wave or a shear wave speed and σ is the amplitude of the relevant applied stress. Each of these rates of change of sound speeds is simply related to a specific linear combination of third order, and second order, elastic constants. For an isotropic medium there are three different third order constants, A, B, C in the Landau notation, and so there must be at least three such rates of change measured in order to deduce the values of A, B, C. Winkler *et al.*(1996, 2004) have described a practical way of doing this by measuring the effect of hydrostatic pressure, as well as that of uniaxial stress, on P and S wave speeds. In the uniaxial case the propagation direction is perpendicular to the applied force; in addition to the P wave there are two distinct polarizations of the S wave. Thus, altogether there are five independent measurements of rate of change of speed from which we deduce the three values of A, B, C.

The system of five equations over-determines the three unknowns but does so consistently for the dry rock samples. For the water-saturated samples, however, the χ^2 of the fit is not nearly as good, implying that a more complete theory of stress dependence of sound speeds than that implied by standard third order elasticity may be required. In Table I we list the relevant parameter values for the three air-saturated rocks and in Table II we list those values for the same three rocks when they are water saturated. The three rock samples reported here are among those reported earlier by Winkler and McGowan (2004).

The parameter β describes the amplitude of nonlinearly generated waves as seen in three wave mixing experiments. For a hyperelastic isotropic solid it may be computed from the third order elastic constants (Norris, 1998), viz

$$\beta = -\left[\frac{3}{2} + \frac{A+3B+C}{K+(4/3)\mu}\right], \tag{1}$$

where K and μ are the bulk and shear moduli, respectively, of the material. The values of β computed from the measured values of A, B, C using Eq. (1) are listed in Tables I and II as $\beta(stress)$.

III. THEORY: KZK EQUATION

In order to interpret our nonlinear measurements, we need a theoretical tool which incorporates nonlinearity, diffraction, attenuation, and reflection losses at the various interfaces. We used the modified KZK equation for this purpose (Norris, 1998):

$$\frac{\partial^2 p}{\partial z\,\partial \tau} + \frac{\partial F}{\partial \tau} - \frac{V}{2}\nabla_{\perp}^2 p = \frac{\beta}{2\rho V^3}\frac{\partial^2 p^2}{\partial \tau^2}. \tag{2}$$

Here, p is the acoustic pressure, V is the linear speed of sound, ρ is the density, and β is the nonlinear parameter. If the solid is hyperelastic, then β is given by Eq. (1). In our work, however, we consider β to be an adjustable parameter whose value we deduce from an analysis of the three wave mixing measurements; we compare this value, $\beta(3wm)$, against that deduced from elasto-acoustic measurements implied by Eq. (1), $\beta(stress)$. Inasmuch as our experimental geometry is axis symmetric, the received pressure is a function of the cylindrical coordinates r and z, as well as of retarded time $\tau = t - z/V$, i.e., $p = p(r, z, \tau)$. The linear frequency dependent attenuation coefficient, $\alpha(\omega)$, is incorpo-

rated in the quantity F, which is conveniently written in the frequency domain as

$$\tilde{F} = \alpha(\omega)\tilde{p}. \tag{3}$$

If $\alpha \propto \omega^2 \leftrightarrow F \propto \frac{\partial^2 p}{\partial \tau^2}$ one recovers the usual KZK equation appropriate to a viscous fluid. Equation (2) is solved in the frequency domain by expanding the solution as a Fourier series, as described in Aanonsen *et al.* (1984) and Naze Tjøtta *et al.* (1990):

$$p = \sum_n [a_n(z,r)\sin(\omega_n \tau) + b_n(z,r)\cos(\omega_n \tau)]. \tag{4}$$

In principle the nonlinearity causes the generation of new frequency components related to the two initial frequencies by

$$\omega_n = N_1\omega_1 + N_2\omega_2, \tag{5}$$

where N_1, N_2 are any integers but we shall focus our attention on the dominant nonlinear frequencies given by the difference, $f_2 - f_1 = 100$ kHz, the sum, $f_1 + f_2 = 2.0$ MHz, and the second harmonics, $2f_1 = 1.9$ MHz and $2f_2 = 2.1$ MHz. The amplitudes of these frequency components are all predicted, within this classical nonlinear theory, to scale quadratically as a function of the initial pressure amplitude of the two fundamentals. All other frequency components are predicted to be much weaker than these.

The code was modified by us to include discontinuities of material properties and to allow for an arbitrary frequency dependence for the attenuation, $\alpha(\omega)$. As the KZK equation is a paraxial approximation, intended to describe the forward propagating part of the wave, the reflection losses were incorporated simply by multiplying each frequency component of the wave by the appropriate impedance-dependent factor as it crosses a boundary. Although the code accounts for reflection losses at boundaries it does not, however, account for multiple reflections, which, in fact, are quite small for our attenuative samples. Inasmuch as the computations are done in the frequency domain, it was a simple matter to incorporate an arbitrary frequency dependence to the (linear) attenuation. We simply read in the appropriate values for the desired frequencies.

IV. RESULTS

In this section we present our results for a test case, consisting of PMMA, a commercially available plastic which is obtainable in large, homogeneous blocks. We then discuss our results on three different air-saturated sedimentary rocks, and finally on the same three rocks, water saturated.

A. PMMA

We chose to investigate PMMA partly because of the convenience of its availability in large, homogeneous pieces and partly because we can make the effects of attenuation in the 3WM experiments as large as those in the rocks (see below). The speed of sound is similar to those of the rocks, though the density is quite different. Because, it turns out, β for PMMA is not significantly larger than that of water, we found that the forward solution to the KZK equation with a 1-in.-thick sample of PMMA was relatively insensitive to the assumed β value; the acoustic path is dominated by water and so the nonlinear signals are dominated by β(water).

Because the attenuation is much smaller than in the rock samples, however, we were able to get our most sensitive determination of β(PMMA) using a 146.15-mm-thick sample directly abutting the transmitter. The receiver was located in water just outside the PMMA block. There are no multiple reflections to speak of. This geometry, of course, loads the transmitter so that it is no longer calibrated as it was in water. Assuming that the measured attenuation coefficients of the two fundamentals are accurate, we can recalibrate the system using the computed proportionality between the transmitter face pressure and the received pressure amplitudes of the two fundamentals, f_1 and f_2. We do this for data for which the amplitudes of the fundamentals are well within the linear regime. Then, for any of the received signals we know the corresponding transmitter face pressure. This procedure is similar to that which we used to calibrate the transmitter in a water-only system but with the additional complication that attenuation within PMMA is appreciable (see below).

Using this so-determined initial amplitude we compute the non-linear signals for a range of β values, seeking to find that which best fits the measured data. An example of this is shown in Fig. 2. Because essentially the entire acoustic path is in PMMA the computed nonlinear signals are all directly proportional to β of PMMA.

For each of the four nonlinear components β may be determined as that value for which the computed amplitude equals the measured one. We see that each of these four signals is fit by nearly the same value of β. We repeat this procedure for the other received signals, corresponding to other initial transmitter amplitudes, and thereby determine an overall best fit value: $\beta(3wm) = 7.1 \pm 0.5$ for PMMA. This value is listed in Tables I and II where we see that it reasonably matches the value determined from the elasto-acoustic effect, $\beta(stress) = 9 \pm 2$.

Our values of β are to be compared to the value $\beta = 10$ deduced by Landsberger and Hamilton (2001) from the measured amplitude of second harmonic generation in a similar geometry and the value $\beta = 6$ deduced from earlier elasto-acoustic measurements (Winkler and Liu, 1996). It has been alleged (Batzle *et al.*, 2006) that PMMA samples vary widely from batch to batch; thus there may not be any real issue to the different numbers determined by the different measurements reported in the literature. Our measurements (stress and three wave mixing) were made on samples machined from the same large block so to the extent that the two determinations of β are equal, within the error bars, it represents the first confirmation of this aspect of the hyperelastic hypothesis in a solid.

Using the value $\beta = 7.1$ we have computed the amplitudes of all the received frequencies in our experiment. The results are plotted in Fig. 3. Over essentially the entire range of amplitudes we are covering: (1) The received signals for f_1 and f_2 are directly proportional to the initial transmitter amplitude. (2) The nonlinearly generated signals have a quadratic dependence on the amplitude of the initial waves, f_1

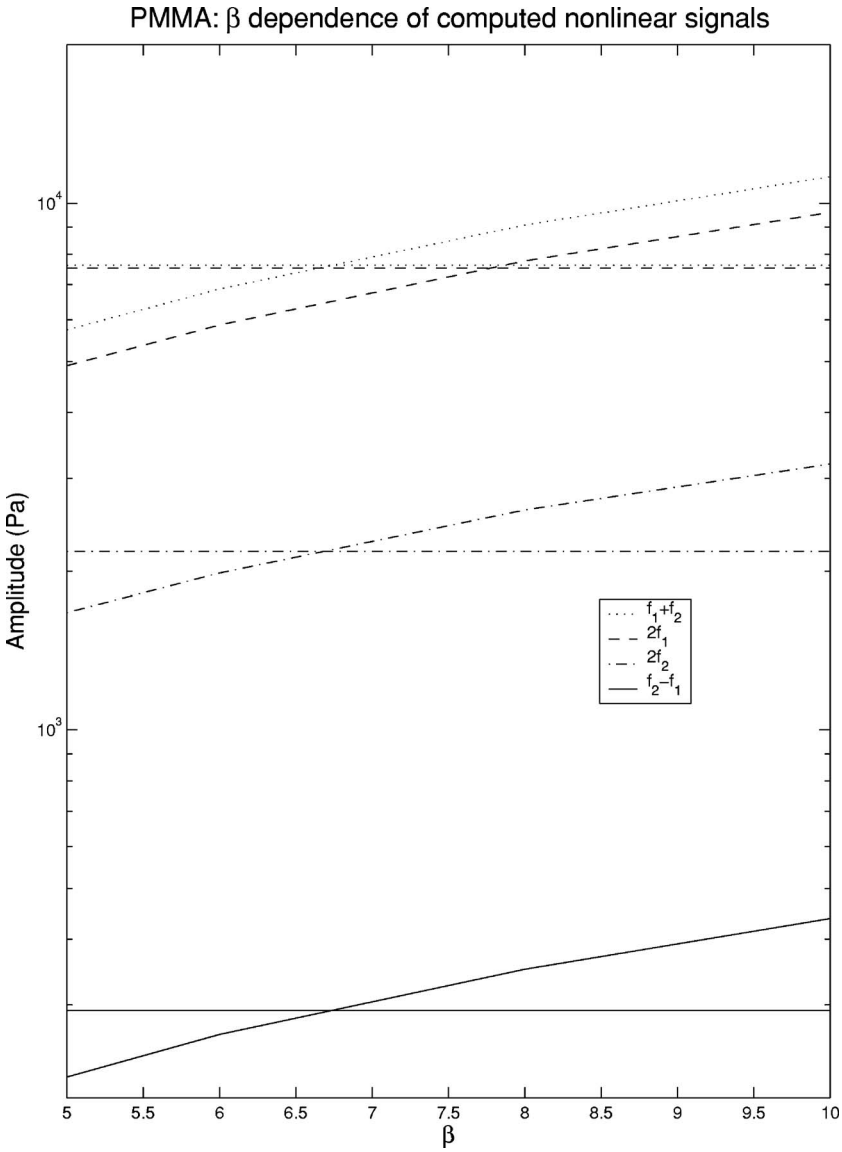


FIG. 2. Comparison of measured against computed nonlinear signals for PMMA as a function of β. The assumed transmitter pressure is $P_{f_1}(z=0)=1.21\times10^6$ Pa, which corresponds to a received amplitude $P_{f_1}(z=146^+\text{mm})=1.51\times10^5$ Pa. The horizontal lines are the measured data; the measured amplitudes of f_1+f_2 and of $2f_1$ are virtually the same. All four computed nonlinear signals fit the data within a narrow range of the same β value.

and f_2. (3) The KZK code does an excellent job of predicting these amplitudes. This last point is significant evidence that both our experimental procedure as well as our theoretical processing of the data are working well.

In order to emphasize the role of attenuation and diffraction, we plot the computed amplitudes of all relevant waves as a function of distance along the z axis, in Fig. 4, for a specific assumed value of the initial pressure. We see destructive interference due to diffraction effects in the fundamental (f_1 and f_2) components as well as in the higher harmonics within the first 5 cm of the sample. We see a large decrease in the amplitudes due to attenuation; we have plotted the plane wave attenuation coefficients $\exp[-\alpha(f)z]$ for $f=1$ MHz, $f=2$ MHz, and $f=100$ kHz, as guides to the eye. We see the discontinuity in the transmitted amplitudes at the PMMA-water interface.

Because the value determined from these three wave mixing measurements is in substantial agreement with that determined from the stress dependence of the sound speeds, we draw the following conclusions: (1) Through third order in the strain the deformation energy in PMMA is well described by standard hyperelastic theory. (2) Our experimental

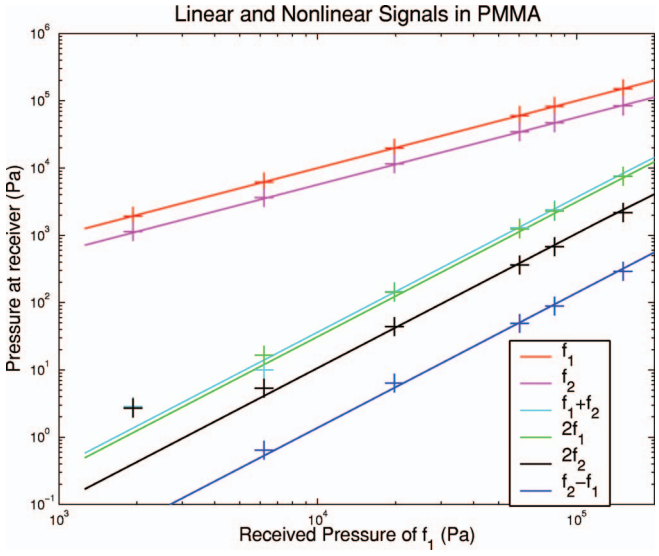


FIG. 3. Amplitudes of the two fundamental frequencies, f_1 and f_2, as well as those of the nonlinearly generated signals plotted against the received amplitude of f_1. The acoustic path is 146 mm of PMMA and less than 1 mm water. The symbols represent the measured data and the solid curves are the solutions to the KZK equation with $\beta=7.1$ for PMMA.

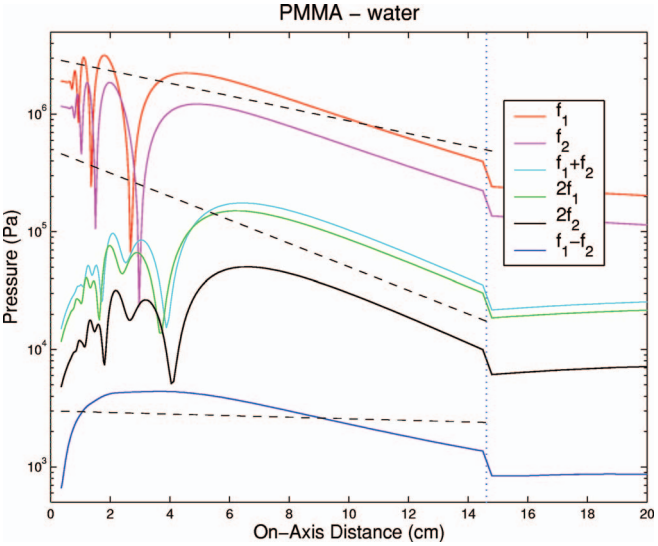


FIG. 4. Computed amplitudes of the relevant frequency components of Fig. 3 on the z axis from the transmitter. The dashed lines represent linear decay at 1 MHz, 2 MHz, and 100 kHz, from top to bottom. The vertical dotted line is the PMMA-water boundary. The receiver is located just outside this boundary.

techniques for determining β both from the stress dependence of the sounds speeds and from the three wave mixing measurements are accurate. (3) The modified KZK equation, Eq. (2), accurately describes the combined effects of nonlinearity, diffraction, attenuation, and the amplitude loss at the PMMA-water interface. It would be unlikely if the agreement between theory and experiment would hold for all four nonlinear signals if these three conclusions were not true.

B. Air-saturated sedimentary rocks

The primary samples for this investigation consist of a set of three different sedimentary rocks: a Berea Sandstone, a Portland Sandstone, and an Indiana Limestone. Each has a porosity in the range 17–24%. The samples were dried in a partial vacuum and then equilibrated under room-air conditions. We made no attempt to monitor the humidity of the air in the pores; we note that Van Den Abeele *et al.* (2002) had previously reported that certain linear and nonlinear acoustics properties of their rock samples were essentially independent of humidity for relative humidities less than 80%, which certainly applies to our air-saturated samples. In order to use our immersion technique in a water bath, each air-saturated sample was sealed with a very thin ($\approx$150 μm) layer of polyurethane [Cytec, EN-20] which we had extensively tested to ensure that it added no measurable acoustic properties of its own. Specifically, we coated the exterior of a 1-in.-thick slab of PMMA and the measured amplitudes and arrival times of the linear and nonlinear components were unaffected by this coating. We also coated the interior of a 1.50 in. i.d. steel tube and monitored the characteristics of a tube wave propagating down the inside of the pipe; we found no measurable changes in modal characteristics for this tube wave. We had investigated several other possible coating materials before we were able to identify one which did not

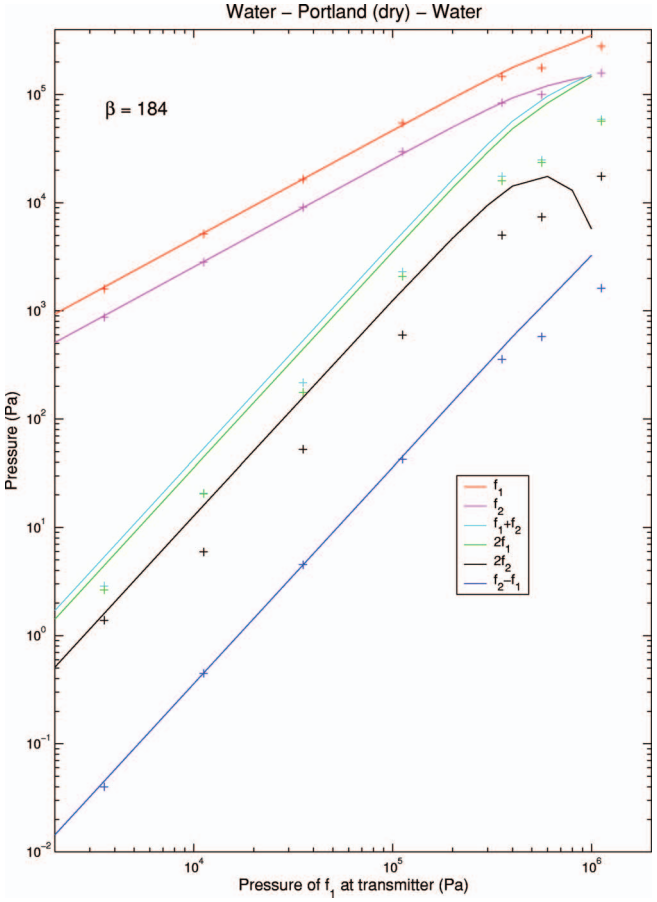


FIG. 5. Measured amplitudes of f_1, f_2 and of the nonlinearly generated signals as a function of the initial amplitude of f_1 for Portland Sandstone (dry). The sample is 25.5 mm thick and it is set 47.6 mm in front of the transmitter. The receivers are 201.8 mm from the transmitter. The symbols represent the measured data and the curves represent the results of the calculations, using parameter values as described in the text.

contribute its own acoustic properties to the experiment. For the purposes of measuring the elasto-acoustic effect, we did not use the coated samples.

First, we consider Portland Sandstone. In Fig. 5 we show a fairly typical plot of the received amplitudes of the two fundamentals, f_1 and f_2, as well as those of the three-wave mixing signals, $f_1+f_2, 2f_1, 2f_2$ and f_2-f_1, as a function of the initial amplitude of f_1 on the calibrated transmitter face. We see that, for the range of pressure amplitudes accessible to us, the nonlinearity of the systems is weak enough that the amplitudes of the received fundamentals are essentially linearly related to the initial amplitudes. Accordingly, by adjusting the assumed values of the attenuations $\alpha(f_1)$ and $\alpha(f_2)$ in the computations, we are able to match our computed amplitudes of f_1 and f_2 against those measured. Similarly, both the computed and measured amplitudes of the difference frequency are nearly quadratic in their dependence on the initial amplitude, as expected for weak nonlinearity. With a value of β=184 the computed and measured amplitudes are in essential agreement, at least for the lower amplitudes.

In principle we could analyze the three nonlinear components near 2 MHz, as we did for PMMA. The problem is that the attenuation coefficients around 2 MHz are very large

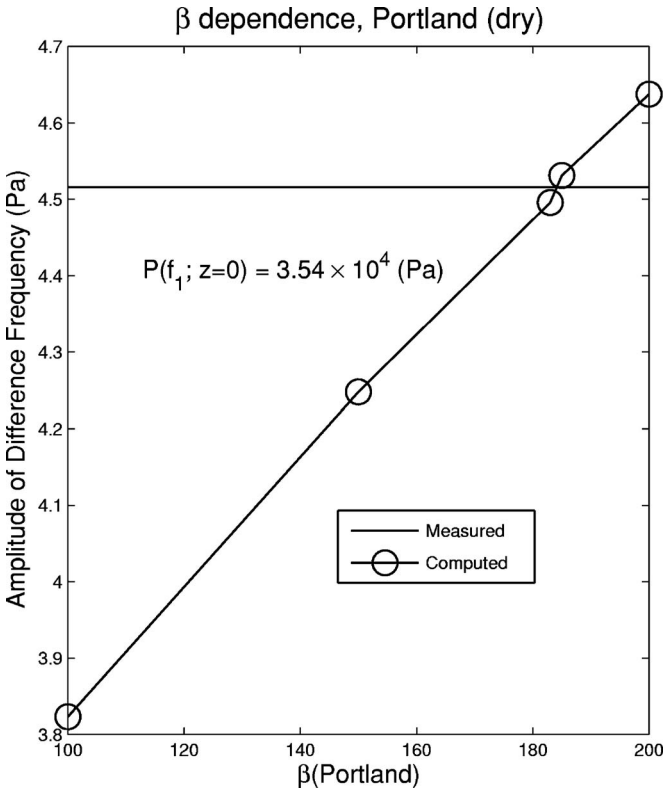


FIG. 6. Sensitivity of the computed amplitude of the difference frequency to variations in β (Portland, dry). The measured amplitude is plotted as a horizontal line. Where these two curves intersect yields the experimentally determined value of β.

and the computed amplitudes of these components are very sensitive to them. We have no direct way of determining those attenuation values in a way analogous to how we handle the attenuation coefficients at 0.95 and 1.05 MHz. This is why, we believe, the computed amplitudes of these signals are much larger than the measured ones. This point is discussed in Sec. IV D. We could adjust the assumed values of the attenuation near 2 MHz in order to bring them into agreement but this hardly seems fruitful. In order to deduce experimental values of $\beta(3wm)$, then, we confine our analysis to the difference frequency components only, in the remainder of this article.

We note in passing that the measured amplitudes of these three components near 2 MHz scale quadratically with the initial amplitude, in agreement with the observations of others on sedimentary rocks (Nazarov *et al.*, 1988; Johnson and Shankland, 1989; Johnson *et al.*, 1996). We observed this quadratic behavior in all three air-saturated samples and to a limited extent in the water-saturated ones.

In Fig. 6 we show how the calculated amplitude of the difference frequency varies with β (Portland, dry) for a given value of the initial pressure of f_1. The horizontal line is the measured amplitude, corresponding to the same initial pressure as for the calculation. β (Portland, dry) is determined as the point of intersection of the two curves. We note that the computed curve has reasonable sensitivity to β (Portland, dry) and that the variation is essentially linear in β (Portland, dry). (The scatter in the computed data reflects convergence issues in the code.) Indeed, on general grounds, one expects that, in this limit of weak nonlinearity, the dependence may be written:

$$P_\Delta = [a_1\beta(\mathrm{H_2O}) + a_2\beta(Portland)][P_{f_1}(z=0)]^2, \tag{6}$$

where a_1 and a_2 are independent of the initial amplitude and of either value of β but do depend on all the other parameters in the problem, including the positions of the sample and the receiver, etc. For the case at hand our calculations indicate:

$$a_1 = 6.86 \times 10^{-10}(\mathrm{Pa})^{-1}$$
$$a_2 = 6.50 \times 10^{-12}(\mathrm{Pa})^{-1}. \tag{7}$$

The fact that a_1 is two orders of magnitude larger than a_2 reflects the facts that most of the acoustic path is in the water, the attenuation of the rock sample is significantly larger than that of water, and the region where the fundamentals, f_1 and f_2, are strongest is in water. A similar situation held for the 1-in.-thick PMMA sample, which necessitated that we use the much thicker sample (previous subsection). All of the rock samples, however, have β values so much larger than that of water that it is possible to determine these values with reasonable sensitivity, as we see in Fig. 6, for example.

We note from Fig. 5 that as the initial amplitudes of the fundamentals are increased toward 1 MPa the measured amplitudes of all three signals are significantly lower than those we compute. We believe this is evidence for an amplitude dependent attenuation, which becomes very significant for the water-saturated samples. (See Sec. IV C).

Of necessity the KZK code contains the underlying assumption that the samples are homogeneous. In order to test this assumption we have taken six sets of measurements by shifting the sample position laterally relative to the incident beam, at 1 cm collinear intervals, keeping the distance from the transducers to the sample fixed. The results are shown in Fig. 7 for dry Portland Sandstone, at a fixed initial pressure for f_1 of 1.12×10^4 Pa. We see that, judging from the apparent sound speed alone, one would conclude that the sample is essentially homogeneous. But the measured amplitudes of all three signals have significant position dependence. The observation that nonlinear parameters are significantly more sensitive to heterogeneities in a variety of systems has been made by others (Rudenko and Soluyan, 1977; Zaitsev *et al.*, 2006). We are somewhat surprised that the linear attenuation in the rocks we studied are so sensitive to heterogeneities but the linear moduli are not. In order to gain some insight into the distribution of values for the attenuation and for β we have processed this data as before, as if the sample was homogeneous. The results for Portland Sandstone are plotted in Fig. 8. The variability of these rock parameters on a centimeter scale is significant.

We repeat the process for our other two samples. In Fig. 9 we show a plot of a typical amplitude dependence of the received signals for a dry sample of Indiana Limestone. We also show the results of the KZK computation using values of the attenuation and of β, deduced as described above. In Figs. 10 and 11 we show the results of a transverse scan of this sample. Figures 12–14 are the equivalent results for the dry sample of Berea Sandstone. We summarize our results

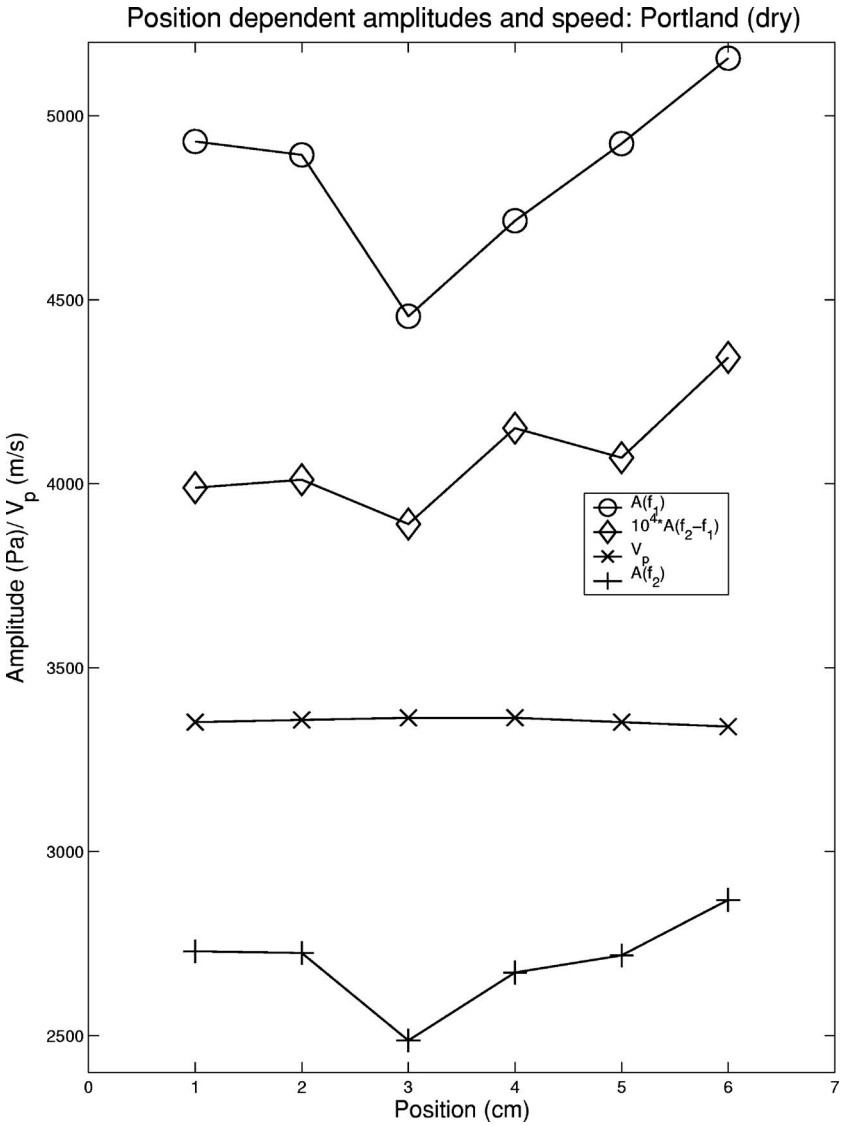


FIG. 7. Measured values of the amplitudes of the two fundamental frequencies, of the amplitude of the difference frequency, and of the speed of sound as a function of lateral sample position. The initial amplitude of f_1 is 1.12×10^4 Pa.

for β for these rocks in Table I. The values for the attenuation coefficients, in the Table, are taken as nominal ones.

We make the following observations about these air-saturated samples:

(1) In all three samples, and at each position within each sample, we find that there is a low-amplitude region in which the amplitudes of the received fundamental frequencies, f_1 and f_2, are linearly related to the initial amplitude of f_1 at the transmitter. Similarly, the received amplitude of the difference frequency, f_2-f_1, is quadratically related to the initial amplitude of f_1. This basic property is a prerequisite for any system that may be hyperelastic; it allows us to make a sensible estimate of the parameter of nonlinearity, β. (Additionally, the amplitudes of $f_1+f_2, 2f_1$ and $2f_2$ also scale quadratically with the initial amplitude of f_1.)

(2) As the initial amplitude of f_1 increases beyond 10^5 Pa, the received amplitudes of f_1 and of f_2 in all three samples fall significantly below the values predicted by the calculations. This did not happen to any significant extent for the PMMA sample. We ascribe this behavior to a non-linear attenuation mechanism, one which increases with increasing amplitude. We discuss this point further in Sec. IV E.

(3) At the centimeter scale there is significant heterogeneity in each of the samples. This implies a range of values for the attenuations and for β which we have estimated by our point-by-point method. This range of values is summarized in the relevant entries for $\beta(3wm)$ in Table I. Notice that, as expected, these samples have very significantly larger β values than non-rocks.

(4) With the possible exception of the dry Portland Sandstone sample, for which $\beta(3wm)\approx\beta(stress)$, we find a significant violation of the hyperelasticity hypothesis in the sense that $\beta(3wm)\ll\beta(stress)$ for the other two samples.

C. Water-saturated sedimentary rocks

In this subsection we report the results of our nonlinear investigations on the same three samples, but under the condition that they are water saturated by means of a standard vacuum impregnation technique. In fact, these results were obtained earlier and were summarized in a previous publication (D' Angelo *et al.*, 2004). In particular, all data were taken without the presence of an external polyurethane coating. In Figs. 15–17 we show fairly typical examples of the

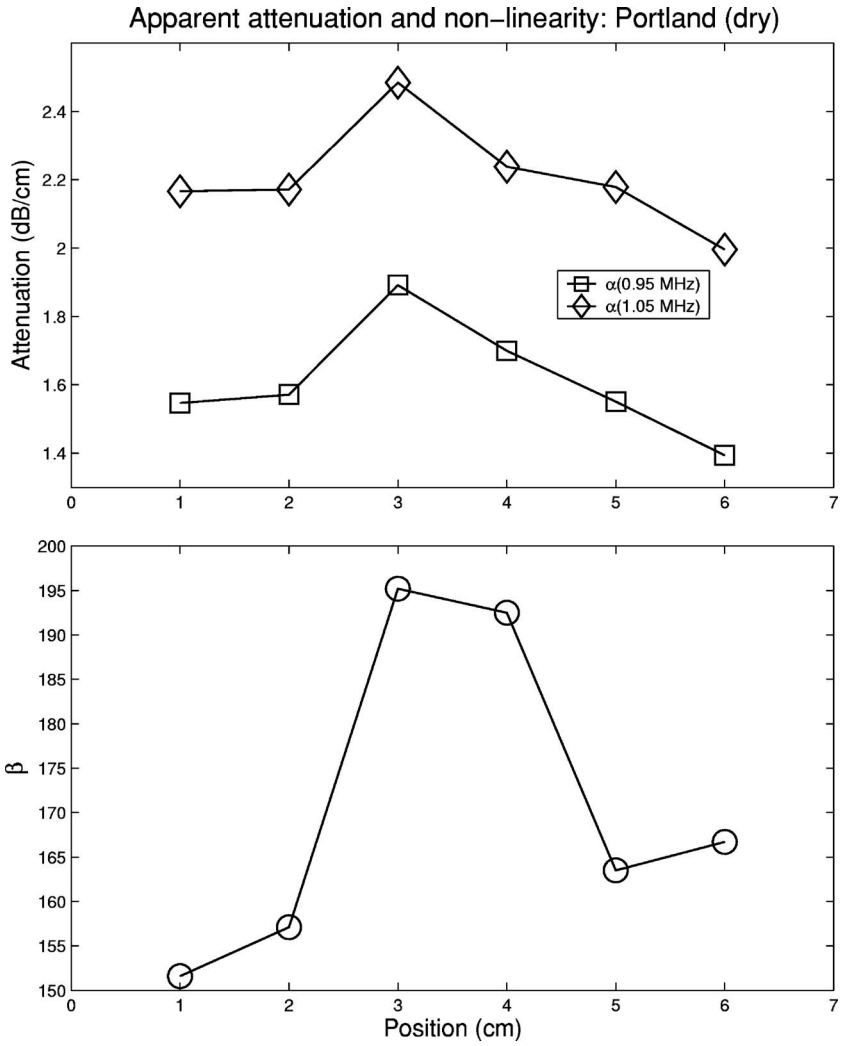


FIG. 8. Apparent values of the attenuation of f_1 and of f_2 (top) and of β (bottom) as a function of lateral position, implied by the data of Fig. 7 for Portland (dry).

amplitude dependence of the received signals in saturated Portland Sandstone, Indiana Limestone, and Berea Sandstone, respectively. Apart from the heterogeneity issue discussed in the previous subsection, the Indiana Limestone and the Berea Sandstone samples have the additional feature that there is no unambiguous regime in which the difference frequency amplitude is quadratically related to the incident amplitude. In the case of Berea, there is not even a clearly defined regime in which the amplitudes of the fundamentals, f_1 and f_2, are linearly related to the initial amplitude. The attenuation in these saturated samples is so large in the vicinity of 1 MHz that we are unable to accurately quantify the amplitudes of the received signals in a low-amplitude regime such that the fundamentals would be linear, and the difference frequency quadratic, in the initial amplitude of f_1. Not only is the attenuation large but, we believe, it is significantly amplitude dependent.

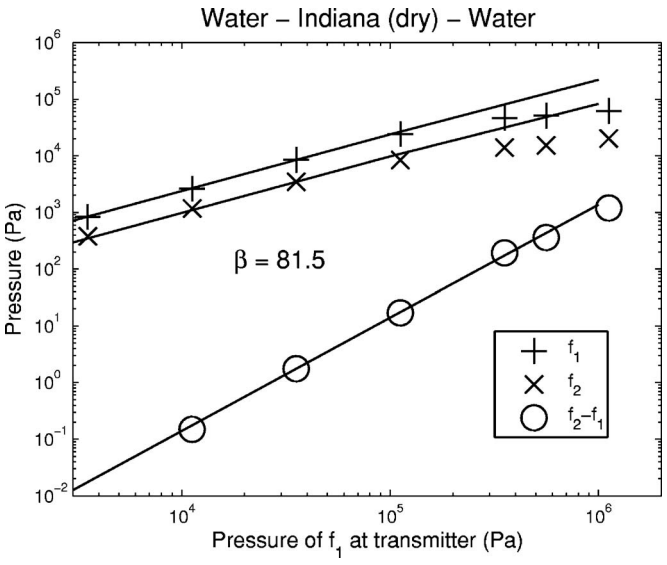


FIG. 9. Measured and computed linear and non-linear amplitudes for Indiana Limestone (dry) vs. initial amplitude of f_1. Same conventions as in Fig. 5.

Notwithstanding, we have estimated a range of applicable β values for these samples by fitting the calculated values to the measured ones at specific incident amplitude values. We have done this using different transmitter-sample spacings, as well as different lateral positions. (At each step we also fit the assumed attenuation values for f_1 and f_2, as described in the previous subsection.) We believe this is a reasonable procedure for the saturated Portland, inasmuch as a clearly defined regime of linear and quadratic behavior exists, for the fundamentals and the difference frequencies, respectively. To an extent it is approximately valid for the

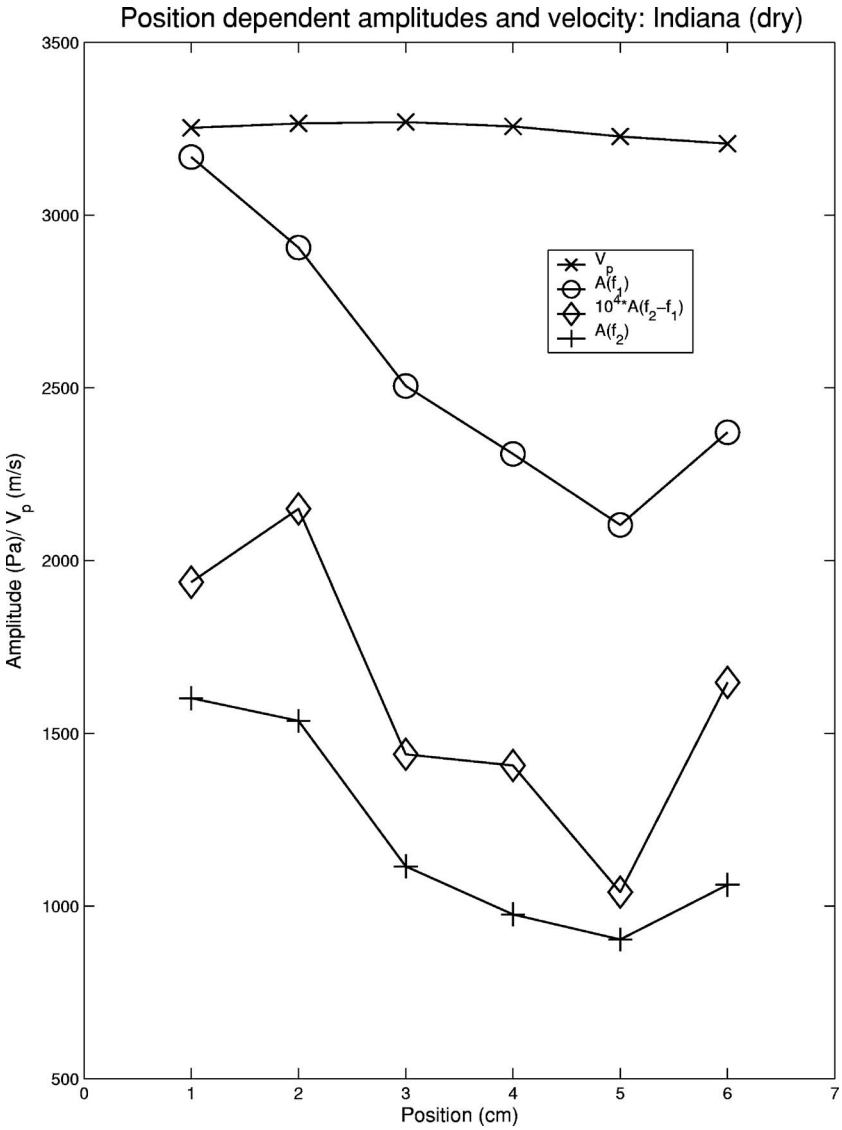


FIG. 10. Position dependent amplitudes of received signals for dry Indiana Limestone. Same conventions as in Fig. 7.

Indiana Limestone, too. For the Berea sample, however, the received difference frequency amplitude seems to scale as a power law: $A(f_2-f_1) \propto A^n(f_1)$, where $n \approx 1.82$. The meaning of this result is unclear; the theory on which the KZK equation is based has $n \equiv 2$ for small amplitudes and just this behavior is seen in the dry Berea data (Fig. 12). The range of so-deduced β values for these water-saturated samples is entered into Table II. Again, the listed values of the attenuation are nominal ones. As with the dry samples (previous subsection) we see that there is a considerable range of apparent $\beta(3wm)$ values. Much of this variation is due to the heterogeneity in the samples, discussed previously. We did not systematically investigate the position dependence of the received signals for these saturated samples. Notwithstanding, it is clear that for all three saturated rocks $\beta(3wm) \ll \beta(stress)$.

In addition to the heterogeneity effect, we believe the role of attenuation, and its amplitude dependence, is a significant effect, especially for these saturated samples. That was our motivation for extending our earlier results to the case of dry samples, where attenuation is generally smaller.

D. Sensitivity to attenuation

As an illustrative example in Fig. 18 we show the results of a calculation of the axial dependence of the computed linear and non-linear signals for saturated Portland Sandstone. In the computation we used $\beta=59$ and we had adjusted the attenuation values, $\alpha(f_1)$ and $\alpha(f_2)$, in order that the computed amplitudes of f_1 and of f_2 matched those that were measured. In this computation we used attenuation values $\alpha(f_2-f_1)$, $\alpha(2f_2)$, $\alpha(2f_1)$ and $\alpha(f_1+f_2)$ that we deduced from the broad-band measurement. This figure is much like that of Fig. 4 except that the rock sample is confined between the two vertical dotted lines, the remainder being water. For comparison we indicate the exponential decay of a plane wave implied by the nominal attenuation values at 1.0 MHz, 2.0 MHz, and 100 kHz, as dashed lines, top to bottom. We draw the following conclusions which apply to all the samples to varying degrees.

(1) The attenuation at the difference frequency within the rock is small enough that errors in a determination of its value do not significantly affect its computed amplitude. The sample thickness is significantly less than the decay distance

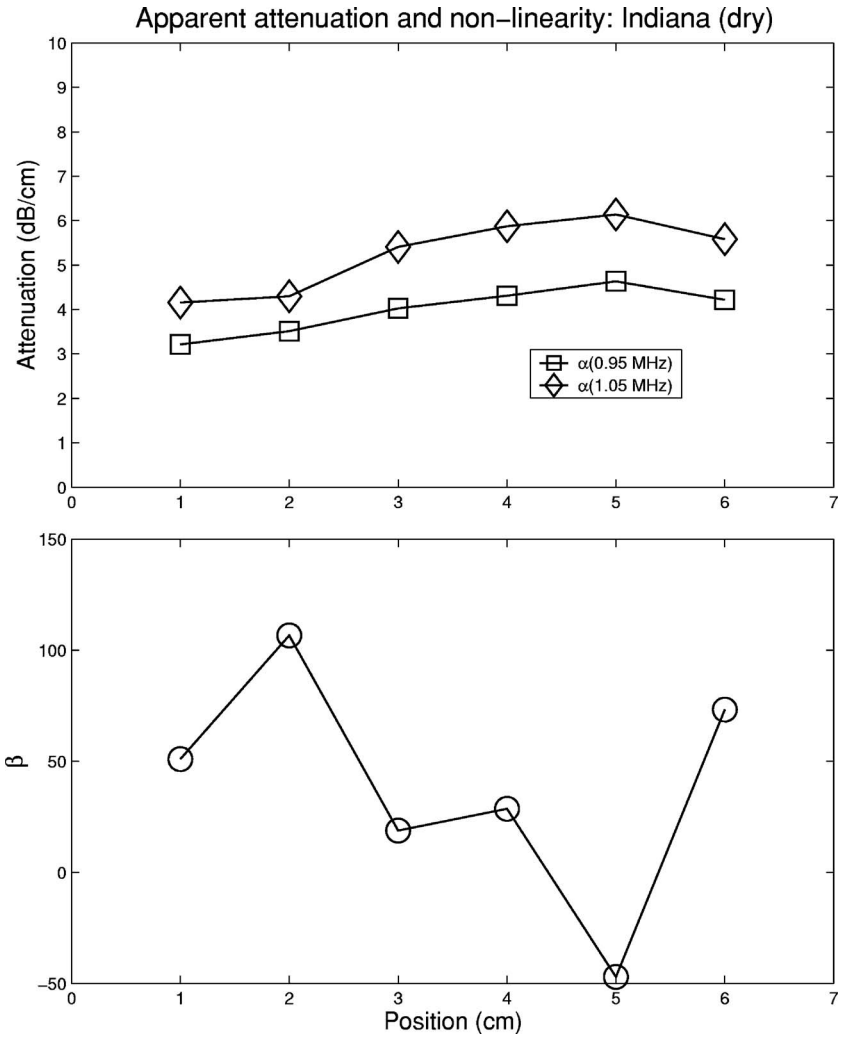


FIG. 11. Apparent values of the attenuation of f_1 and of f_2 (top) and of β (bottom) as a function of lateral position, implied by the data of Fig. 10 for Indiana Limestone (dry).

at 100 kHz. For this reason we think that uncertainties in our determination of the attenuation at 100 kHz do not translate to significant uncertainties in $\beta(3wm)$. We have directly verified this by introducing small variations in α(100 kHz) in our calculations.

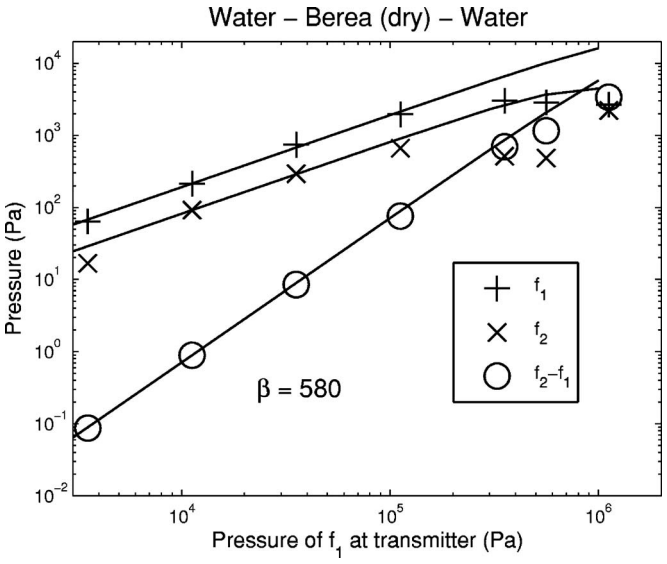


FIG. 12. Measured and computed linear and non-linear amplitudes for Berea Sandstone (dry). Same conventions as in Fig. 5.

(2) The attenuation of each of the two fundamentals is significant across the 25.4 mm thickness of the sample and small errors in those values have a very significant effect on the transmitted amplitudes. The effect of attenuation on the fundamental amplitudes can be seen in Fig. 18. For this reason we have tried to set the values of the attenuation coefficients of f_1 and f_2 by requiring that the computed amplitudes match the measured ones, in the strictly linear regime, as discussed earlier.

Because the fundamentals act as a source for the nonlinear signals, the latter computed amplitudes are also very sensitive to slight changes in the attenuation values of the fundamentals, as we have verified by direction calculations. A simple intuitive way to see this is in the over-simplified case in which there is negligible diffraction so the nonlinear acoustics problem becomes one dimensional. The generalized KZK equation, Eq. (2), may be integrated once with respect to τ and the result is the generalized Burgers equation. A perturbation theory gives the result for the amplitude of the nonlinearly generated difference frequency, in an arbitrarily dispersive and attenuative medium (Tserkovnyak and Johnson, 2004, Eq. 17). In the present instance where we have neglected dispersion in the phase velocity, but retained the effects of attenuation, we have:

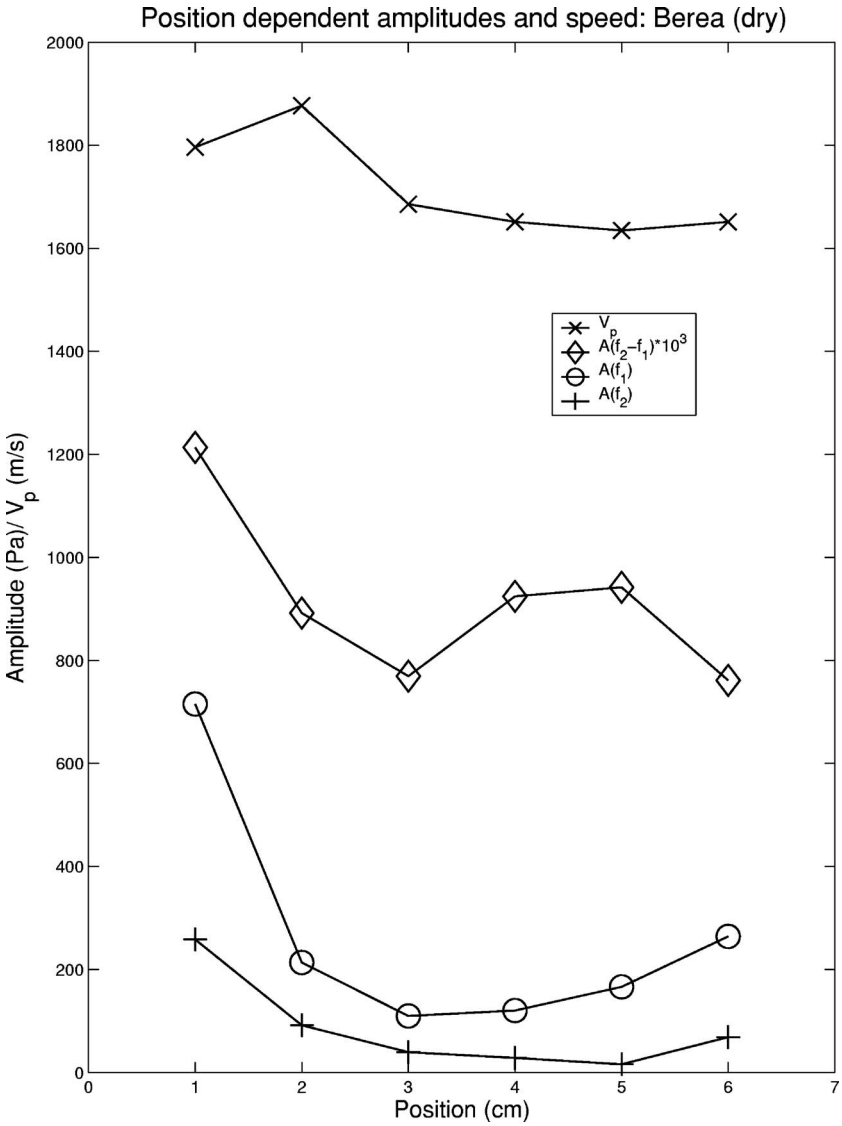


FIG. 13. Position dependent amplitudes of received signals for dry Berea Sandstone. Same conventions as in Fig. 7.

$$p_\Delta(z) = p_\Delta(0) + \frac{\Delta\omega\beta}{2\rho_f V^3} \times \left[\frac{e^{-[\alpha(\omega_1)+\alpha(\omega_2)]z} - e^{-\alpha(\Delta\omega)z}}{\alpha(\omega_1)+\alpha(\omega_2)-\alpha(\Delta\omega)}\right]\cos(\Delta\phi)p_1(0)p_2(0). \quad (8)$$

$\Delta\phi$ is a relative phase factor and $p_\Delta(0)$ represents the amplitude of the difference frequency component in the rock sample that was generated in the water path. For "small enough" path lengths, the amplitudes of the nonlinearly generated waves increase linearly with propagation distance; we have

$$\lim_{z\to 0}\left[\frac{e^{-[\alpha(\omega_1)+\alpha(\omega_2)]z} - e^{-\alpha(\Delta\omega)z}}{\alpha(\omega_1)+\alpha(\omega_2)-\alpha(\Delta\omega)}\right] = -z. \quad (9)$$

But this is not a relevant regime for our experiments. From Eq. (8) we see that the source for the generation of the difference frequency, being the product of the amplitudes of the two fundamentals, decays over a distance $\xi = 1/[\alpha(\omega_1)+\alpha(\omega_2)]$. This distance is on the order of one centimeter, or less, for all the samples, as can be seen from Tables I and II. We have $\xi = 0.78$ cm for saturated Portland Sandstone. The source for the generation of the difference frequency component is effectively confined to a distance ξ from the front surface. Inasmuch as the thicknesses of our samples, L, are generally 2.54 cm we have, from Eq. (8),

$$p_\Delta(L) \approx p_\Delta(0) + \frac{\Delta\omega\beta}{2\rho_f V^3} \times \left[\frac{e^{-\alpha(\Delta\omega)L}}{\alpha(\omega_1)+\alpha(\omega_2)}\right]\cos(\Delta\phi')p_1(0)p_2(0), \quad (10)$$

where we have used the fact that $[\alpha(\omega_1)+\alpha(\omega_2)] \gg \alpha(\Delta\omega)$. Thus, the amplitude of the nonlinearly generated difference frequency is very sensitive to the assumed values of the attenuation coefficients of the fundamentals; this conclusion holds true, both for the one-dimensional problem and for the full KZK based calculation.

(3) The computed amplitudes of the components whose frequencies are near 2 MHz, ($2f_1, 2f_2$ and f_1+f_2), are also extremely sensitive to the input values of the attenuation coefficients at f_1 and f_2, for basically the same reason as above. In addition they are very sensitive to the assumed

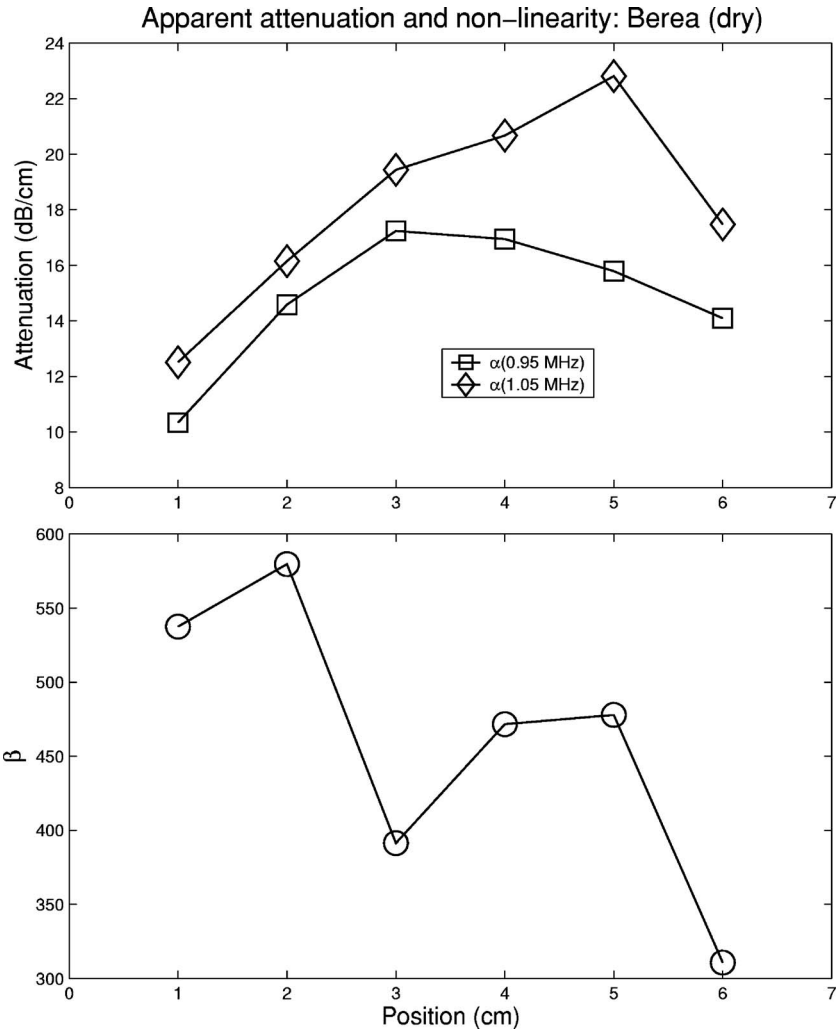


FIG. 14. Apparent values of the attenuation of f_1 and of f_2 (top) and of β (bottom) as a function of lateral position, implied by the data of Fig. 13 for Berea Sandstone (dry).

values near 2 MHz, which is more or less obvious from Fig. 18: There is a very significant decrease in amplitude of these nonlinear components upon traversing the sample. Inasmuch as we do not have an *in situ* means of determining these 2 MHz attenuation coefficients, as we did for the fundamentals, we did not pursue any attempt to quantify these measurements near 2 MHz.

(4) Based on the foregoing observations it might make sense to perform the 3WM measurements on thinner samples, but this approach has potential problems of its own. As we mentioned earlier, our incident signals consist of a 50 μs gate of f_1 and f_2. While this corresponds to $\approx$50 cycles of each of the fundamentals it is only five cycles of the difference frequency, $f_2-f_1=100$ kHz. An appreciably shorter pulse would introduce time-domain complications that we are trying to avoid by using effectively a continuous wave approach. Even so a multiply reflected component starts to come through even before our 50 μs gate is over. The two-way travel time through our samples is $\approx$20 μs, depending on the speed of sound in the sample. Of course, the KZK solver is not able to accommodate these multiple reflections but here the rather large attenuation in our samples is helpful. The additional two-way attenuation of the fundamental frequency components, $\exp(-\alpha_1 2L)$, is generally extremely small, except for Portland (dry) for which it equals 0.4. Moreover, for all relevant components there are significant reflection losses at the interfaces. Notwithstanding, we were reluctant to consider using thinner samples for our measurements.

E. Onset of amplitude dependent attenuation

In each of Figs. 5, 9, 12, and 15–17 there is a significant reduction in the measured amplitudes of the two fundamentals, relative to those computed from the KZK equation, as the amplitude increases above a certain level. Of course, for large enough amplitudes the KZK solutions predict a deviation from a linear relation between received and transmitted signals because some of the energy in the fundamentals is converted into the other harmonics. But the measured rollover in received amplitudes is occurring at amplitudes small enough that the fundamentals are still predicted to be in the linear regime, essentially. We have tentatively ascribed this effect to nonlinear, or amplitude dependent, attenuation. If, for example, the attenuation term in the KZK equation, Eq. (3), can be expanded symbolically as

$$F = \alpha p + \delta p^2 + \gamma p^3 + \cdots, \tag{11}$$

then it is clear that the cubic term, γp^3, will contribute to a change in attenuation at the fundamental frequencies as the

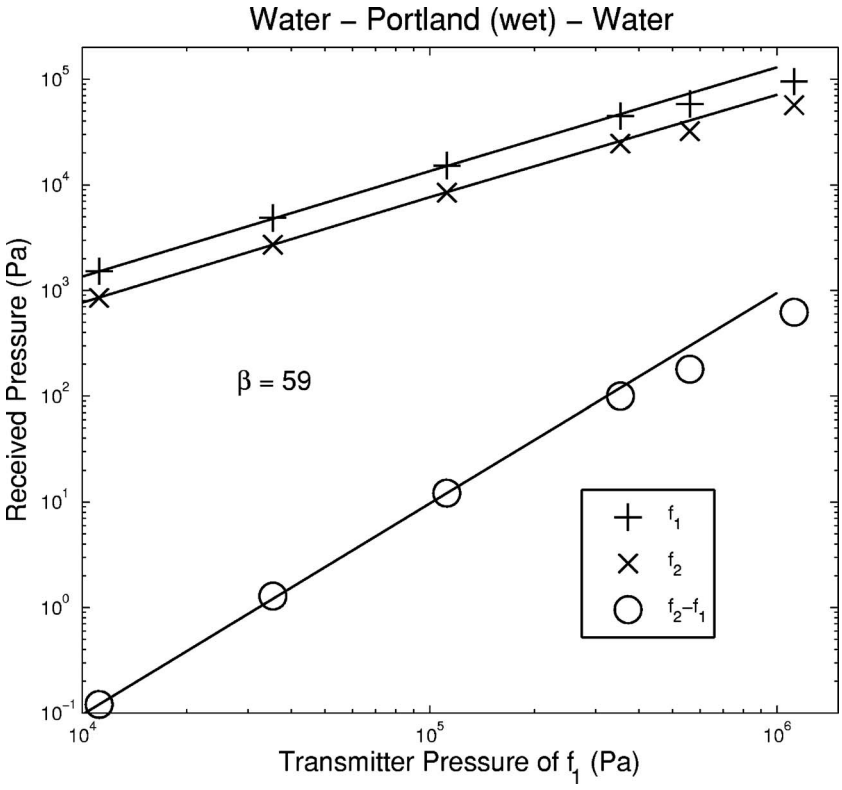


FIG. 15. Measured and computed linear and non-linear amplitudes for Portland Sandstone (saturated). Same conventions as in Fig. 5.

amplitude is increased. The simplest example of this is for the case of a single frequency, viz:

$$[P_0 \cos(\omega t)]^3 = (1/4)P_0^3[\cos(3\omega t) + 3\cos(\omega t)]. \quad (12)$$

Quite generally, a cubic term in an equation of motion will have the effect of directly modifying the behavior of the fundamental frequencies, whereas a quadratic term will do so only indirectly. We have attempted to estimate the onset of nonlinear attenuation as follows: From each of Figs. 5, 9, 12, and 15–17 we estimate that value of the transmitter face pressure, $P_0 = P(z=0)$, at which the measurements "first" deviate from the KZK calculations. Obviously, this is quite subjective. Then, using that value for P_0 in the KZK calculation we find the corresponding maximum amplitude within the rock sample, P_M, assuming the approximate validity of the input attenuation, etc. Finally, the strain at which this first

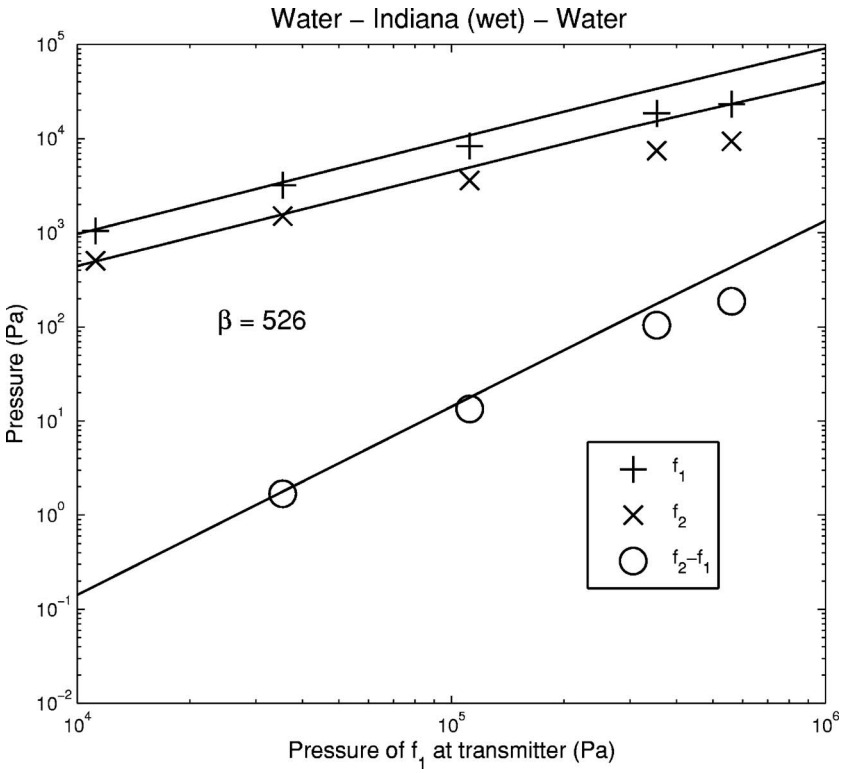


FIG. 16. Measured and computed linear and non-linear amplitudes for Indiana Limestone (saturated). Same conventions as in Fig. 5.

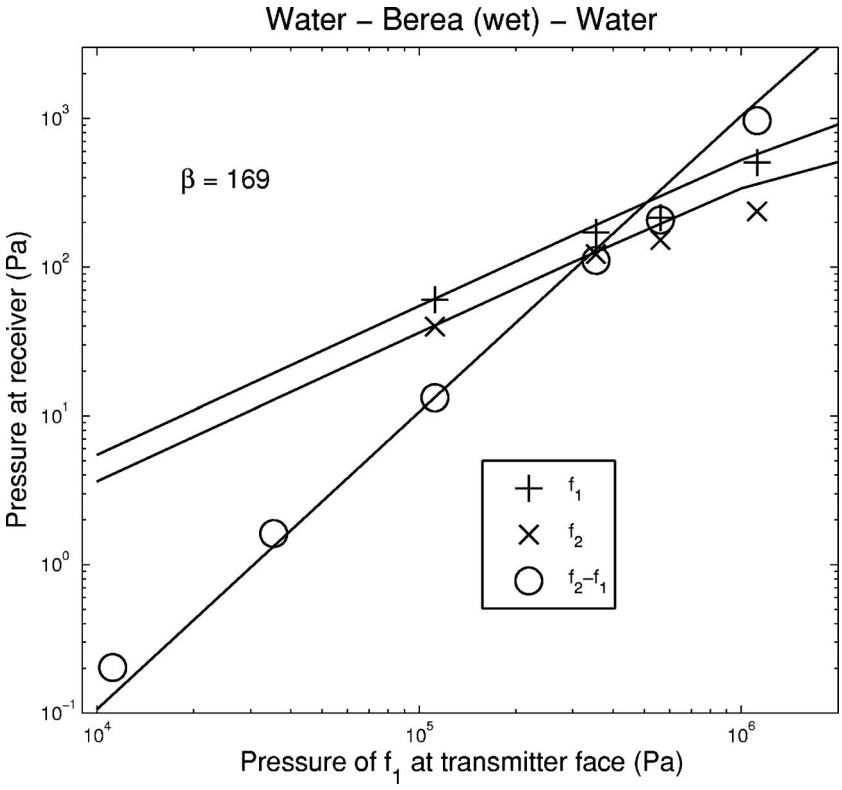


FIG. 17. Measured and computed linear and non-linear amplitudes for Berea Sandstone (saturated). Same conventions as in Fig. 5.

occurs is $\epsilon = P_M / C_{11}$ where C_{11} is the P-wave modulus. Our results are summarized in Table III.

Amplitude dependent attenuation has previously been reported in rocks. Using a resonant bar technique Winkler *et al.* (1979) have shown the onset to occur for strains typically in the range $10^{-8} - 10^{-7}$. Similarly TenCate *et al.* (2004); and Pasqualini *et al.* (2007) have also reported similar results, around 1 kHz. Our own estimates of the onset, if even approximately valid, are several orders of magnitude larger than those previously reported. It is, therefore, not surprising that we may be seeing the effects of nonlinear attenuation. For frequencies around 1 MHz Mashinskii (2006) has reported the onset of nonlinear attenuation in a sample of Nivagalsk sandstone for strains in the range $10^{-7} - 10^{-6}$ but he sees a general decrease in attenuation as strain is increased whereas we, and others (TenCate *et al.*, 2004; and Pasqualini *et al.*, 2007) see a significant increase. We do not understand why there is this difference between Mashinskii (2006) and the others. A possible mechanism for nonlinear attenuation has been proposed by Zaitsev and Matveev (2006). The input to that theory contains a number of parameters which we are unable to measure and so we are not in position to make a meaningful comparison between theory and experiment.

What is surprising is that the onset occurs for strains seemingly so large. The two main differences between our technique and that of Winkler *et al.* (1979) and of TenCate *et al.* (2004); Pasqualini *et al.* (2007)] are: (1) We are operating at frequencies three orders of magnitude higher than they. (2) They have used a resonant bar technique whereas we have used a through transmission technique. In our measurements the high amplitude regions are more localized than in the resonant bar but this hardly seems to account for

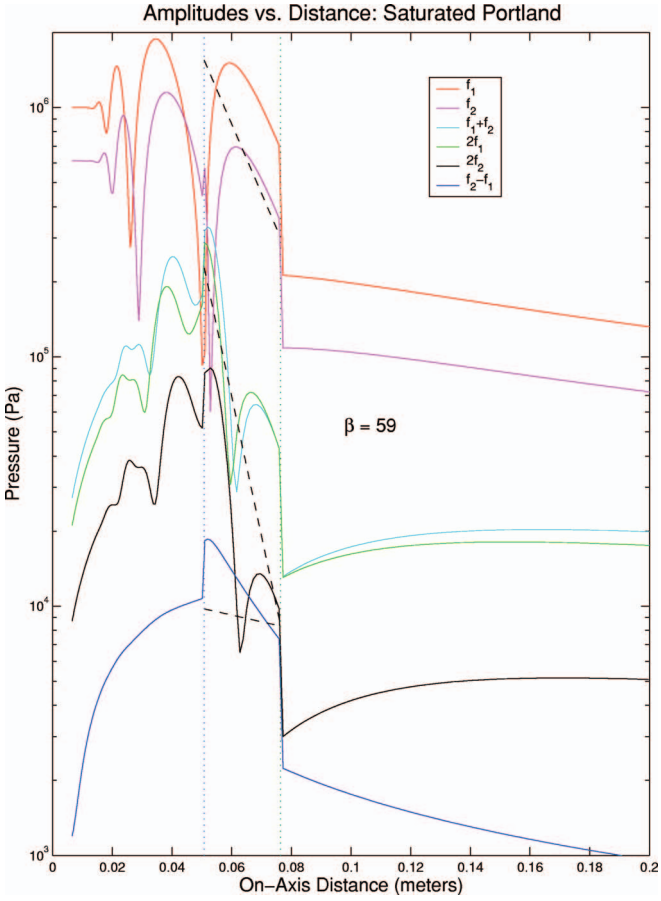


FIG. 18. Computed amplitudes of the relevant frequency components on the z axis from the transmitter. The water-saturated sample of Portland Sandstone is confined to the region between the two dotted lines. The dashed lines represent linear decay at 1 MHz, 2 MHz, and 100 kHz, from top to bottom. Data were taken at 203 mm.

TABLE III. Estimates for the onset of amplitude dependent attenuation. $C_{11}=\rho V_P^2$ is the P-wave modulus. P_0 is that value of the pressure on the transmitter face at which the received amplitudes of the fundamentals first deviate from the predictions based on amplitude independent attenuation. P_M is the maximum value within the rock sample corresponding to that value of P_0. ϵ is the corresponding strain.

	C_{11} (GPa)	P_0(kPa)	P_M (kPa)	$\epsilon=P_M/C_{11}$
Berea (dry)	5.92	100	25	4.2×10^{-6}
Berea (sat.)	13.57	<100	<30	$<2.2\times10^{-6}$
Indiana L.S. (dry)	22.92	100	150	6.5×10^{-6}
Indiana L.S. (sat)	36.32	35	53	1.5×10^{-6}
Portland S.S. (dry)	24.46	300	690	2.8×10^{-5}
Portland S.S. (sat)	28.06	300	450	1.6×10^{-5}

the orders of magnitude difference in the onset strains. We are inclined to think that there is a genuine frequency dependence to the onset.

V. CONCLUSIONS

We have established that PMMA may be considered to be a hyperelastic solid in the sense that the value of β deduced from the elasto-acoustic measurements agrees with that deduced from the three-wave mixing measurements. Inasmuch as diffraction, attenuation, reflection losses at the water-PMMA boundary are all significant effects, we also conclude that our experimental technique is a valid one, as are our solutions of the modified KZK equation. It would be unlikely that we would get good agreement for all four nonlinear signals with the same value of β were this not so.

The three dry samples of rock all have a low-amplitude regime by which we mean the received amplitudes of the two fundamental components, f_1 and f_2, are proportional to the incident amplitude and the amplitudes of the difference frequencies, f_2-f_1 (as well as those of the primary higher harmonics) are proportional to the square of the incident amplitude. The values of the nonlinear parameter, $\beta(3wm)$, though they are much larger than that of PMMA, are generally significantly smaller than the value implied by the elasto-acoustic effect, $\beta(stress)$. A possible exception occurs for dry Portland Sandstone, for which $\beta(3wm)\approx\beta(stress)$. Thus, we may say that the dry Portland Sandstone sample approximately obeys the hyperelasticity hypothesis, to 3rd order in the strain energy, in this context.

So to say, at the high frequencies of our 3wm experiments the nonlinear micro-mechanical elements within the rock samples do not have enough time to exhibit the nonlinearity that would be implied by the static, elasto-acoustic, measurements. Models which can incorporate such frequency dependent nonlinear effects have been proposed by Gusev *et al.* (1998), Zaitsev *et al.* (2001) and Gusev and Tournat (2005) but we are not in a position to test the quantitative validity of these theories, as they contain several additional parameters not readily measurable by us.

There is significant heterogeneity, on a one-centimeter scale, of the deduced apparent values of $\beta(3wm)$ as well as the linear attenuation coefficients, $\alpha(f_1)$ and $\alpha(f_2)$. There is no appreciable heterogeneity evident in the sound speeds.

Because of the much larger attenuation in the water-saturated samples, it was difficult to acquire data in the low-amplitude regime, as had been done for the dry samples (above). This was, however, possible for the saturated Portland Sandstone and, to an extent, for the saturated Indiana Limestone. For the saturated Berea Sandstone the attenuation is so large that we are not able to see a regime in which the received amplitudes of f_1 and f_2 are strictly proportional to the initial amplitudes. Moreover, the received amplitudes of the difference frequency (in saturated Berea) do not vary quadratically with the initial amplitudes of the fundamentals. We find $A(f_2-f_1)\propto A^n(f_1)$, where $n\approx1.82$.

Using our data we have made estimates of the range of admissible values of $\beta(3wm)$ for these saturated samples. We find, in all three cases, that $\beta(3wm)\ll\beta(stress)$. The range of values is due, in part, to the heterogeneous nature of the samples as well as to the fact that there is not always a clearly delineated range in which the amplitude of the difference frequency component scales quadratically with that of the fundamental amplitude.

Because the measured amplitudes of the received fundamentals, f_1 and f_2, are significantly lower than their computed values for high amplitudes, we tentatively conclude that the attenuation coefficients are themselves amplitude dependent. We estimate the onset of nonlinear attenuation as being in the range of strain values $\epsilon\approx10^{-6}-10^{-5}$, which is two orders of magnitude larger than onset values previously reported by others, using lower frequency techniques.

ACKNOWLEDGMENTS

We are grateful for the advice and assistance of L. McGowan, T. J. Plona, and N. Chang. We have benefited greatly from discussions with V. E. Gusev, R. A. Guyer, P. A. Johnson, and A. N. Norris.

Aanonsen, S. I., Barkve, T., Naze-Tjøtta, J., and Tjøtta, S. (**1984**). "Distortion and harmonic generation in the nearfield of a finite amplitude sound beam," J. Acoust. Soc. Am. **75**, 749–768.

D' Angelo, R. M., Winkler, K. W., Plona, T. J., Landsberger, B. J., and Johnson, D. L. (**2004**). "Test of hyperelasticity in highly nonlinear solids: Sedimentary rocks," Phys. Rev. Lett. **93**, 214301.

Batzle, M. L., Han, D.-H., and Hofmann, R. (**2006**). "Fluid mobility and frequency dependent seismic velocity—direct measurements," Geophysics **71**, N1–N9.

Dvorkin, J., Mavko, G., and Nur, A. (**1995**). "Squirt flow in fully saturated rocks," Geophysics **60**, 97–107.

Eringen, A. C., and Şuhubi, E. S. (**1974**). *Elastodynamics, V. I* (Academic, New York, 1974).

Gusev, V. E., Lauriks, W., and Thoen, J. (**1998**). "Dispersion of nonlinearity, nonlinear dispersion, and absorption of sound in micro-inhomogeneous materials," J. Acoust. Soc. Am. **103**, 3216–3226.

Gusev, V., and Tournat, V. (**2005**). "Amplitude and frequency dependent nonlinearities in the presence of thermally-induced transitions in the Preisach model of acoustic hysteresis," Phys. Rev. B **72**, 054104.
Guyer, R., McCall, K., Boitnott, G., Hilbert, L., and Plona, T. (**1997**). "Quantitative implementation of Preisach-Mayergoyz space to find static and dynamic elastic moduli in rock," J. Geophys. Res. **102**, 5281–5294.
Guyer, R. A., and Johnson, P. A. (**1999**). "Nonlinear mesoscopic elasticity: Evidence for a new class of materials," Phys. Today **52**, 30–36.
Hamilton, M. F., and Blackstock, D. T. (**1998**). *Nonlinear Acoustics* (Academic, New York).
Jacob, X., Barriére, C., and Royer, D. (**2003**). "Acoustic nonlinearity parameter measurements in solids using the collinear mixing of elastic waves," Appl. Phys. Lett. **82**, 886–888.
Johnson, D. L., and Norris, N. A. (**1997**). "Rough elastic spheres in contact: Memory effects and the transverse force," J. Mech. Phys. Solids **45**, 1025.
Johnson, P., Zinszner, B., and Rasolofosaon, P. (**1996**). "Resonance and elastic nonlinear phenomena in rock," J. Geophys. Res. **101**, 11,553–11,564.
Johnson, P., and Shankland, T. J. (**1989**). "Nonlinear generation of elastic waves in granite and sandstone: Continuous wave and travel time observation," J. Geophys. Res. **94**, 17,729–17,733.
Landsberger, B. J., and Hamilton, M. F. (**2001**). "Second-harmonic generation in sound beams reflected from, and transmitted through, immersed elastic solids," J. Acoust. Soc. Am. **109**, 488–500.
Mashinskii, E. I. (**2006**). "Nonlinear amplitude-frequency characteristics of attenuation in rock under pressure," J. Geophys. Eng. **3**, 291–306.
Meegan, G. D., Johnson, P. A., Guyer, R. A., and McCall, K. R. (**1993**). "Observation of nonlinear elastic wave behavior in sandstone," J. Acoust. Soc. Am. **94**, 3387–3391.
Nazarov, V., Ostrovsky, L., Soustova, I., and Sutin, A. (**1988**). "Nonlinear acoustics of micro-inhomogeneous media," Phys. Earth Planet. Inter. **50**, 65–73.
Naze Tjøtta, J., Tjøtta, S., and Vefring, E. H. (**1990**). "Propagation and interaction of two collinear finite amplitude sound beams," J. Acoust. Soc. Am. **88**, 2859–2870.
Norris, A. N., and Johnson, D. L. (**1997**). "Nonlinear elasticity of granular media," J. Appl. Mech. **64**, 39.
Norris, A. N. (**1998**). "Nonlinear elasticity of solids" in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), p. 263.
Ostrovsky, L., and Johnson, P. A. (**2001**). "Dynamic nonlinear elasticity in geomaterials," Riv. Nuovo Cimento **24**, 1–46.
Pasqualini, D., Heitmann, K., TenCate, J. A., Habib, S., Higdon, D., and Johnson, P. A. (**2007**). "Nonequilibrium and nonlinear dynamics in Berea and Fontainbleau sandstones: Low strain regime," J. Geophys. Res. **112**, B01204.
Rudenko, O. V., and Soluyan, S. I. (**1977**). "*Theoretical Foundations of Nonlinear Science*," (Consultants Bureau, Plenum, NY).
Sinha, B. K., Kane, M. R., and Frignet, B. (**2000**). "Dipole dispersion crossover and sonic logs in a limestone reservoir," Geophysics **65**, 390–407.
Tang, X. M., and Patterson, D. (**2001**). "Shear wave anisotropy measurement using cross-dipole acoustics logging: An overview," Petrophysics **42**, 107–117.
TenCate, J. A., Pasqualini, D., Habib, S., Heitmann, K., Higdon, D., and Johnson, P. A. (**2004**). "Nonlinear and nonequilibrium dynamics in geomaterials," Phys. Rev. Lett. **93**, 065501.
Truesdell, C., and Noll, W. (**1965**). "The non-linear field theories of mechanics," in *Encyclopedia of Physics, V. III*, edited by S. Flügge (Springer-Verlag, New York, 1965).
Tserkovnyak, Y., and Johnson, D. L. (**2004**). "Nonlinear tube waves in permeable formations: Difference frequency generation," J. Acoust. Soc. Am. **116**, 209–216.
Van Den Abeele, K. E.-A., Carmeliet, J., Johnson, P. A., and Zinszner, B. (**2002**). "Influence of water saturation on the nonlinear elastic mesoscopic response in Earth materials and the implications to the mechanism of nonlinearity," J. Geophys. Res. **107**, B6 ECV4 1–11.
Winkler, K. W., and Liu, X. (**1996**). "Measurements of third-order elastic constants in rocks," J. Acoust. Soc. Am. **100**, 1392.
Winkler, K. W., and McGowan, L. (**2004**). "Nonlinear acoustoelastic constants of dry and saturated rocks," J. Geophys. Res. **109**, B10204, 1–9.
Winkler, K. W., Nur, A., and Gladwin, M. (**1979**). "Friction and seismic attenuation in rocks," Nature (London) **277**, 528–531.
Zaitsev, V. Yu., Nazarov, V. E., and Belyaeva, I. Yu. (**2001**). "The equation of state of a microinhomogeneous medium and the frequency dependence of its elastic nonlinearity," Acoust. Phys. **47**, 178–183.
Zaitsev, V. Yu., and Matveev, L. A. (**2006**). "Strain-amplitude dependent dissipation in linearly dissipative and nonlinear elastic microinhomogeneous media," Russ. Geol. Geophys. **47**, 695–710.
Zaitsev, V. Yu., Nazarov, V. E., and Talanov, V. I. (**2006**). "'Nonclassical' manifestations of microstructure-induced nonlinearities: New prospects for acoustic diagnostics," Phys. Usp. **49**, 89–102.
Zimenkov, S. V., and Nazarov, V. (**1994**). "Nonlinear acoustics effects in rock samples," Izv. Earth Phys. **29**, 12–18.

Measurements of inner and outer streaming vortices in a standing waveguide using laser doppler velocimetry

Solenn Moreau, Hélène Bailliet, and Jean-Christophe Valière
Laboratoire d'Etudes Aérodynamiques, 40 Avenue du Recteur Pineau, 86022 Poitiers Cedex, 86022 France

(Received 4 May 2007; revised 30 July 2007; accepted 1 December 2007)

Measurements of the axial streaming velocity are performed by means of laser doppler velocimetry in an experimental apparatus consisting of a waveguide having loudspeakers at each end for high intensity sound levels. Streaming is characterized by an appropriate Reynolds number $\mathrm{Re}_{\mathrm{NL}}$, the case $\mathrm{Re}_{\mathrm{NL}} \ll 1$ corresponding to the so-called slow streaming and the case $\mathrm{Re}_{\mathrm{NL}} \geqslant 1$ being referred to as fast streaming. The variation of axial streaming velocity with respect to the transverse coordinate is compared to the available slow streaming theory. Streaming fluid flow is measured both in the core region and in the near wall region. Streaming velocity in the center of the guide agrees reasonably well with the slow streaming theory for small $\mathrm{Re}_{\mathrm{NL}}$ but deviates significantly from such predictions for $\mathrm{Re}_{\mathrm{NL}} > 20$ and its evolution for further increasing $\mathrm{Re}_{\mathrm{NL}}$ is discussed. Then streaming behavior in the near wall region is particularly studied. For $\mathrm{Re}_{\mathrm{NL}} < 70$, two vortices are present across the guide section as predicted by slow streaming theory. Then it appears that, when the Reynolds number is increased, two other vortices become visible in the near wall region. Different stages for the generation and evolution of these inner streaming vortices are presented.
 [DOI: 10.1121/1.2828059]

PACS number(s): 43.25.Nm [MFH] Pages: 640–647

I. INTRODUCTION

"Acoustic streaming" is a generic term used to refer to the second order steady velocity that is induced by and superimposed on the dominating first order acoustic velocity. Among the different kind of streaming, Rayleigh streaming owes its origin to the viscous interaction between the sound field and a solid boundary. Since it was first modeled by Rayleigh, this phenomenon has motivated numerous theoretical studies (e.g., Ref. 1). During the last decade, thermoacoustic application motivated several theoretical works to refine thermal effect description[2] and/or to remove the restriction to narrow channels.[3] Besides these theoretical studies and as noted by Thompson and Atchley[4] and by Sharpe *et al.*,[5] experimental works on streaming have been rather scarce and qualitative until recent development of laser techniques.[4–10] In particular, it is well known that inside the boundary layer, there exist streaming vortices, "inner vortices," whose direction of rotation are opposite those of the outer, Rayleigh vortices; but there is still no experimental evidence and characterization of these inner vortices in standing wave guides to our knowledge.

In the present work, outer and inner streaming velocity field generated by an acoustic standing wave in a cylindrical resonator are investigated experimentally by means of laser doppler velocimetry (LDV). The relevant theory is briefly presented in Sec. II and previous measurements of the phenomenon are discussed. Section III provides an overview of the method used for calculating streaming velocity and Sec. IV presents results of measurements together with their comparison to the available theoretical expectation.

II. THEORETICAL AND EXPERIMENTAL BACKGROUND

In a two-dimensional resonator with rigid walls in which a $\lambda/2$ standing wave is set up, as shown by Fig. 1(a), streaming vortices are present either sides of the central axis and spaced at intervals of $\lambda/4$ with $\lambda = c/f_{\mathrm{ac}}$ the acoustic wavelength, c the sound speed in the fluid, and f_{ac} the acoustic frequency.[11] For outer streaming, the steady flow along the tube wall is directed toward the velocity nodes of the standing wave and returns along the central axis of the tube to complete a closed loop. In the near wall region, the inner streaming vortices have directions of rotation opposite to those of the outer. Note that in Fig. 1(a), the ratio R/δ_ν is about 4 with $\delta_\nu = \sqrt{2\nu/\omega}$ the boundary layer thickness, ν the kinematic viscosity of the fluid, and ω the angular frequency of acoustic oscillations.

A. Theory

Streaming cells were first observed by Faraday[12] in 1831, while considering air currents near an oscillating plate, noting that currents of air ascended at displacement antinodes and descended at the displacement nodes of the plate. The problem was solved initially by Rayleigh[13] in 1884 for wide channel (in which the boundary layer thickness is negligible in comparison to channel width and the wavelength is big compared to the tube radius). Applying successive approximations to the equations governing fluid motion, he was able to show that streaming arises due to the generation of Reynolds stresses. His solution describes the steady vortices outside the boundary layer, commonly referred to now as Rayleigh streaming:

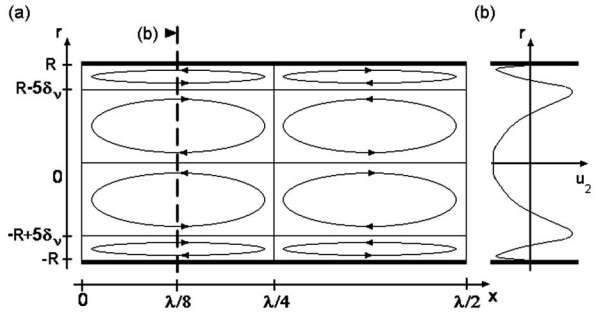


FIG. 1. Streaming velocity field. (a) Acoustic streaming vortices in a resonator is schematized. (b) Variation of axial streaming velocity with respect to the transverse coordinate r is schematized for $x=\lambda/8$.

$$\overline{u_{2,\mathrm{Rayleigh}}} = \frac{U^2}{4c}\sin(2kx)\left(\frac{1}{2}e^{-2R(1+\eta)/\delta_\nu} + e^{-R(1+\eta)/\delta_\nu}\cos\left(\frac{R(1+\eta)}{\delta_\nu}\right) + 2e^{-R(1+\eta)/\delta_\nu}\sin\left(\frac{R(1+\eta)}{\delta_\nu}\right) + \frac{3}{4} - \frac{9\eta^2}{4}\right), \tag{1}$$

with U the acoustic velocity amplitude, c the speed of sound, R the tube radius, x the axial coordinate, k the complex wave number, $\eta=r/R$ the polar coordinates with r the distance from the tube axis; in Eq. (1) and elsewhere, the overbar is used to denote time-averaged quantities, the subscript 0 refers to equilibrium quantities, the subscript 1 to first order quantities and the subscript 2 to second order quantities.

Thermal effects on the outer streaming were first considered by Rott.[14] His result, restricted to wide channels, includes the effects of heat conduction and dependence of the viscosity on temperature in a gas, as well as the effect of a mean temperature gradient imposed along the channel walls. In the case of a zero temperature gradient, Rott obtained an axial streaming velocity given by

$$\overline{u_{2,\mathrm{Rott}}} = (1+\alpha)\overline{u_{2,\mathrm{Rayleigh}}}, \tag{2}$$

where the constant α represents a correction to the magnitude of the streaming velocity. A typical value for α in air at standard condition is $\alpha=0.030$, so that thermal effects alter the streaming velocity in wide channels that do not support any temperature gradient by only a few percent. A result similar to Rott was later obtained independently by Qi[15] who neglected the dependence of viscosity and thermal conductivity on temperature. Qi described the inner streaming vortex confined to the boundary layer, in addition to the outer Rayleigh streaming.

Recently, there has been a renewal of interest in theoretical studies on Rayleigh streaming due to its importance in practical high-amplitude acoustic devices such as thermoacoustic devices. Olson and Swift[2] based their theoretical derivation on Rott's work to derive an expression for the streaming velocity in the case of widely separated parallel plates and used this to propose an appropriate tapering of pulse tube that yields suppression acoustic streaming. Also several authors were concerned with streaming inside the heart of thermoacoustic devices, that is thermoacoustic stack that basically consists in a stack of narrow channels. There was therefore need to remove the restriction to wide channels. Waxler[16] included heat conduction in a study of streaming in a gas between parallel plates. Bailliet *et al.*[3] included temperature dependence of the viscosity in addition to heat conduction in their analysis of streaming in both two-dimensional (2D) channels and cylindrical tubes. In the case of a cylindrical tube that does not support temperature gradient, using polar coordinates, the second-order velocity was found by these authors to be function of acoustic quantities, temperature gradient, geometry of the system and thermophysical properties of the fluid according to

$$\begin{aligned}\overline{u_{2,\mathrm{Bailliet}}} = {} & \frac{4}{\rho_0}(\eta^2-1)\int_0^1 \overline{\rho_1 u_{1x}}\,\eta\,d\eta + \frac{4R^2}{\mu_0}(\eta^2-1)\int_0^1 \eta\int_{-1}^{\eta}\frac{1}{\eta'}\int_0^{\eta'}\partial_x(\rho_0\overline{u_{1x}^2})\,\eta''\,d\eta\,d\eta'\,d\eta'' \\ & + \frac{4R}{\nu_0}(\eta^2-1)\int_0^1 \eta\int_{-1}^{\eta}\overline{u_{1x}u_{1\eta}}\,d\eta\,d\eta' - \frac{4\beta}{T_0}(\eta^2-1)\int_0^1\int_{-1}^{\eta}\overline{T_1\partial_\eta u_{1x}}\,d\eta\,d\eta' \\ & + \frac{R^2}{\mu_0}\int_{-1}^{\eta}\frac{1}{\eta'}\int_0^{\eta'}\partial_x(\rho_0\overline{u_{1x}^2})\,\eta''\,d\eta'\,d\eta'' + \frac{R}{\nu_0}\int_{-1}^{\eta}\overline{u_{1x}u_{1\eta}}\,d\eta' - \frac{\beta}{T_0}\int_{-1}^{\eta}\overline{T_1\partial_\eta u_{1x}}\,d\eta',\end{aligned} \tag{3}$$

with ρ the density, $\mu=\rho\nu$ the dynamic viscosity. T the temperature, ∂_x the axial derivative, and u_{1x} and $u_{1\eta}$ the axial and radial first-order components of the particle velocity. In both the Waxler[16] and Bailliet *et al.*[3] studies, the mean temperature was authorized to vary along the channel walls. Hamilton *et al.*[1] have studied analytical streaming produced by a standing wave in a channel of arbitrary width. They extended their analysis to a gas in which heat conduction and dependence of the viscosity on temperature are taken into account.[17] For the case of a wide tube ($R \gg \delta_\nu$), their results together and those of Waxler[16] and Bailliet *et al.*[3] were in agreement with the results of Rott.

All the above-cited studies assume the streaming to be very slow, that is slow enough to leave the first-order variables unperturbed. As the streaming becomes larger, recent theoretical work by Menguy and Gilbert[18] indicates that nonlinear effect of fluid inertia will render the vortex pattern distorted. They observed that the variation of the axial component of streaming velocity is not parabolic at relatively large acoustic amplitudes. In the case of a waveguide, this effect is determined by an appropriate Reynolds number[19]

$$\mathrm{Re}_{\mathrm{NL}} = \left(\frac{U}{c}\right)^2\left(\frac{R}{\delta_\nu}\right)^2. \tag{4}$$

$\mathrm{Re}_{\mathrm{NL}}$ compares inertia and viscosity and determines the degree to which the streaming velocity field is distorted relative to the field in the slow streaming case. The case $\mathrm{Re}_{\mathrm{NL}} \ll 1$ corresponds to the slow streaming and the case $\mathrm{Re}_{\mathrm{NL}} \geqslant 1$ is

referred to as "nonlinear streaming" or fast streaming. Again, apart from numerical calculation by Menguy and Gilbert valid for $Re_{NL} \approx 1$, all the available theoretical studies are valid for $Re_{NL} \ll 1$ only.

B. Previous measurements

As stated before, only few experimental studies of acoustic streaming are available in the literature and quantitative investigations were only achievable in the last decades thanks to the development of optical methods. To our knowledge, only two research groups succeeded in obtaining quantitative measurements of streaming in a standing wave guide.

The streaming associated with one-dimensional monofrequency acoustic standing wave in a resonator has previously been investigated by Sharpe *et al.*[5] (cylindrical guide, $f=2460$ Hz, $U=2.5$ m/s, $Re_{NL}=4$), Arroyo and Greated[6] (rectangular guide, $f=1910$ Hz, $U=1.5$ m/s, $Re_{NL}=1$), Hann and Greated[8] (square cross section guide, $f=1600$ Hz, $U=2.1$ m/s, $Re_{NL}=3$) and Campbell *et al.*[7] (cylindrical guide, $f=1975$ Hz, $U \approx 1.5$ m/s, $Re_{NL}=4$). These authors used particle image velocimetry to extract the axial component of outer streaming velocity field. Their results obtained agreed reasonably well with Rott's predictions.

Very recent studies by Thompson *et al.*[10] (cylindrical guide, $f=310$ Hz, 2.7 m/s $< U <$ 8.6 m/s, $2 < Re_{NL} < 20$) reported experimental studies of outer Rayleigh streaming in a guide using LDV. Their results for $Re_{NL} \approx 1$ showed that streaming velocities are in better agreement with the theory of Rott than the prediction of Menguy and Gilbert,[18] suggesting that the influence of fluid inertia on the streaming-velocity field is not as deterministic as thermal conditions. They considered three different thermal conditions: isothermal, uncontrolled and insulated. For $Re_{NL} > 1$, they observed that when the magnitude of the temperature gradient increases, the magnitude of the streaming decreases and the shape of the streaming cell becomes increasingly distorted. They found that the thermoacoustically induced axial temperature gradient strongly influences the axial component of the acoustic streaming and that in practice nonparabolic variation of axial streaming velocity with respect to the transverse coordinate is due to the influence of thermal effect more than fluid inertia. For high Reynolds number, they observed that steady-state streaming velocities are not in agreement with any available theory.

In spite of these important recent advancements and although streaming has been reconsidered theoretically during the last years without restricting to outer vortices, there is still no experimental evidence and characterization of inner vortices in standing wave guides. The aim of the present study is to measure particle velocity in the near wall region by means of LDV in order to detect inner vortices and to study its evolution when the Reynolds number is increased.

III. PROCEDURE

A. Experimental apparatus

The setup used to observe the phenomenon of acoustic streaming is shown in Fig. 2 and consists in a cylindrical (2D) tube connected at each end to a loudspeaker so that a

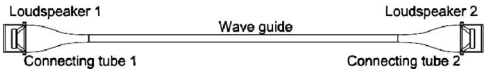


FIG. 2. Diagram of the experimental apparatus.

$\lambda/2$ high level standing wave is sustained in the guide. Loudspeakers are connected to each end of the waveguide via connecting tube designed to avoid separation effect related to the singularities in change of section. The main part of the waveguide is a glass cylinder of inner radius $R=19.5$ mm. The total length of the wave guide $L=\lambda/2$ can vary between 1.1 and 2.0 m with different length of guide tube so that the frequency f_{ac} varies between 84 and 150 Hz. For such frequencies, the boundary layer thickness δ_ν, of order of 0.2 mm, is a very small fraction of the acoustic wavelength λ. Because streaming outer vortices have a $\lambda/4$ periodicity, we expect two Rayleigh streaming cells along the guide length, completely contained within the cylindrical waveguide. The waveguide is filled with atmospheric air.

Wood smoke is introduced into the guide to render the flow visible. A wave generator provides the loudspeaker input signal, whose frequency and amplitude are controlled, as well as a trigger reference signal, used to synchronize the LDV system (Dantec Dynamics Model 2580). The LDV probe is mounted on a three-axis positioning system. The argon krypton laser has an optical wavelength of 514.5 μm, a power of 25 W. The measurement volume is approximately ellipsoidal in shape, 0.399 mm long and 0.047 mm in diameter, and the fringe spacing is 2.168 μm. The longer dimension is oriented perpendicular to the axis of the waveguide. The parameters of the LDV system are adjusted for sound measurement.[20] The axial particle velocity is measured along the centerline of the guide and also across the section: with a 0.05 mm step near the wall and with gradually growing steps until 5 mm in the center of the guide.

B. Determination of axial streaming velocity

For each position in the guide, a burst spectrum analyzer provides a couple time-velocity (t_i, u_i) for each particle crossing the LDV measurement volume. Figure 3 represents

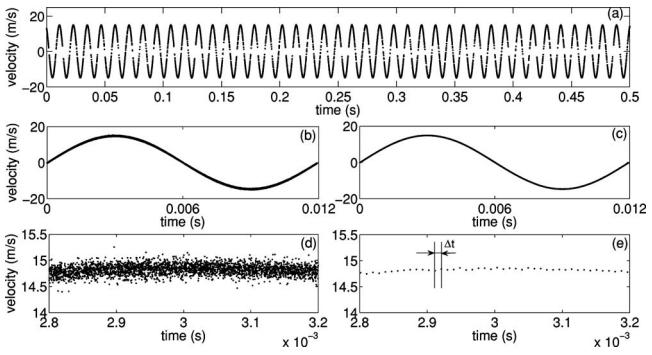


FIG. 3. LDV signal method. (a) Velocity of rough data from LDV measurement are presented as a function of time. (b) and (d) Axial velocity is brought back on one acoustic period. (c) and (e) Axial velocity is averaged over regular time step on the acoustic period. (d) and (e) are zoomed views of (b) and (c).

the method used to compute streaming velocity: Axial velocity issued from LDV measurement u_i [Fig. 3(a)] is brought back on one acoustic period T_{ac} [Fig. 3(b)]

$$u_j(r,t_j) = u_i(r,t_i + T_{ac}(i-1)), \tag{5}$$

corresponding to time

$$t_j = t_i + T_{ac}(i-1), \tag{6}$$

and sorted per growing time. Then, the average over regular time step, Δt, on the acoustic period [Fig. 3(c)] is calculated

$$u_k(r,t_k) = \frac{1}{N_k}\sum_{j=1}^{N_k} u_j(r,t_j + \Delta t(j-1)) \tag{7}$$

corresponding to time

$$t_k = \frac{2k-1}{2}\Delta t, \tag{8}$$

with N_k the particle number between the two times $(k-1)\Delta t$ and $k\Delta t$. The time step $\Delta t = T_{ac}/N$ used for averaging is chosen to be the minimum time step to ensure the presence of at least one couple of measurement in each slot $(N_k \geqslant 1)$ where N is the number of steps. Streaming axial velocity $\overline{u_2}$ is then calculated as the average of velocity points over the period

$$\overline{u_2}(r) = \frac{1}{N}\sum_{k=1}^{N} u_k(r,t_k). \tag{9}$$

C. Preliminary measurements and considerations

1. Effect of harmonic distortion

For the largest acoustic amplitude used in the present study ($U=53$ m/s, $\mathrm{Re_{NL}}=211$) at the position of measurement, the amplitude of the second harmonic of particle velocity is approximately 2% of the amplitude of the first harmonic. Then, following Ref. 10, we can state that: "the assumption of a monofrequency acoustic field is still expected to be valid because the magnitude of the streaming generated by this second-harmonic standing wave is expected to be less than 1% of the magnitude of the streaming generated by the first harmonic, and therefore too small to be detected."

2. Measured onset of streaming

Figure 4 shows a measurement of the onset of streaming associated with the acoustic field being switched on. Each axial streaming velocity is calculated with 70 000 samples. The axial component of streaming velocity reaches its steady state value after 2 min. But for 2 min $< t <$ 26 min, Fig. 4 shows a slow evolution of the axial component of streaming velocity due to the acoustically induced temperature gradient. This is in agreement with Thompson *et al.*[10] who found that for the case of an uncontrolled boundary condition, the axial streaming velocity reaches its steady-state value within approximately 14 min and the axial temperature gradient within 23 min. So finally, in our experiments, measurements are performed more than 26 min after the acoustic field is switched on and are stable after this time.

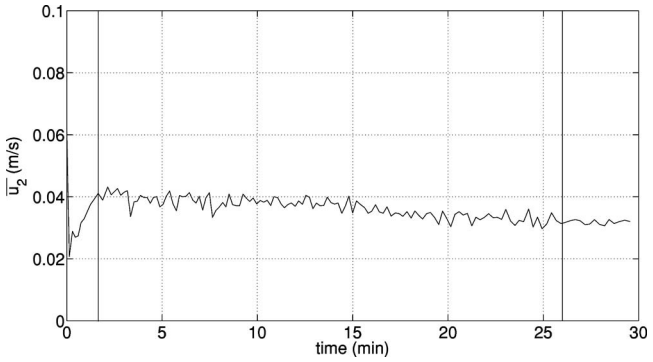


FIG. 4. Onset of axial streaming velocity value measured at the center of acoustic guide as a function of time after the acoustic field is switched on $(x=\lambda/25)$.

3. Measured steady-state streaming

Once the steady-state has been reached, it is important to evaluate how many samples are needed to ensure convergence of streaming velocity.[21] Figure 5 shows that if the calculation of acoustic streaming is performed over 40 000 samples or more, then the calculated value of $\overline{u_2}$ is within 95% of its steady state value, the latter being defined as the streaming velocity calculated with 90 000 samples. Therefore, in order to reach convergence of the measurement, we choose to acquire either 70 000 samples or to stop acquisition after 10 s for good compromise between a sufficient number of point and time runs to get sufficient time to perform a set of measurements with enough seeding particles.

Note that Thompson and Atchley[9] used a different post-processing for their measurements with LDA. Their sample records of velocity contain less than 30 000 samples (during a time interval of 10–115 s) so that the Eulerian streaming velocity measured by LDV channel has not reached its convergence. In order to avoid the bias induced, they determined the Lagrangian streaming velocities of individual tracer particles. Their method assumes that the Eulerian and Lagrangian acoustic-velocity fields are equivalent.

IV. RESULTS

Because measures on outer streaming only are available in the experimental literature, we first focus in Sec. IV A on

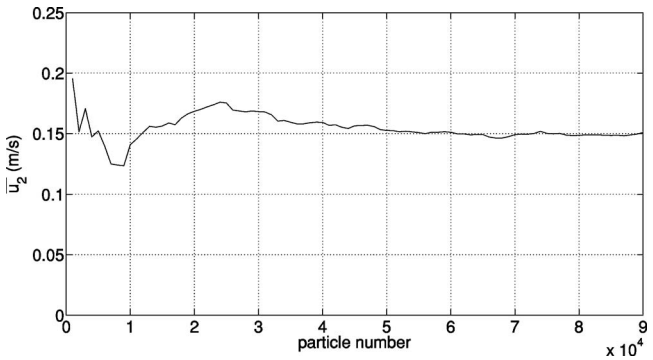


FIG. 5. Axial streaming velocity steady-state value measured at the center of acoustic guide as a function of particle number $(x=\lambda/25)$.

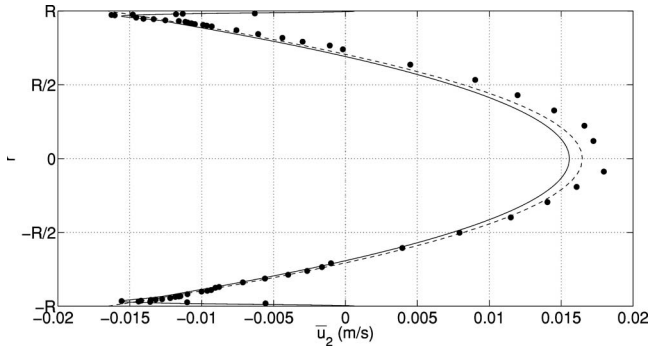


FIG. 6. Axial streaming velocity in guide section for $Re_{NL}=1$, $f_{ac}=88$ Hz, $x=\lambda/8$; —: Rayleigh expression [Eq. (1)]; ---: Bailliet *et al.* (Ref. 3) expression [Eq. (3)]; ●: measurements.

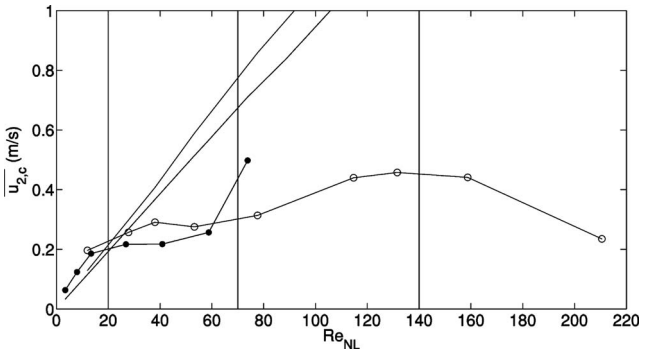


FIG. 8. Evolution of axial component of the centerline streaming velocity ($r=0$). black: $x=\lambda/25$, ●: measurements for $f_{ac}=84$ Hz; gray: $x=\lambda/15$, ○: measurements for $f_{ac}=113$ Hz; —: Rayleigh theory.

outer streaming and compare our results with the literature. Then, in Sec. IV B, results in the near wall region are presented.

A. Outer streaming behavior

Figures 6 and 7 represent the axial streaming velocity in the guide section. Theoretical expression for the streaming derived by Rayleigh [Eq. (1)] and Bailliet *et al.*[3] [Eq. (3)] are also represented for comparison. Recall that Eq. (3) is valid both in the far and in the near wall region whereas Eq. (1) is only valid for outer streaming and that both assume $Re_{NL} \ll 1$. The shapes of these theoretical curves are similar for the outer streaming vortices. At low levels (Fig. 6) the measured axial streaming velocity in the center of the guide is a bit higher than the theoretical curves of Rayleigh in accordance with the theoretical expectation.[3] When the acoustic level is increased, the measured axial streaming velocity in the center of the guide tends to be equal and then smaller than theoretical expectations in accordance with Thompson *et al.*[10] findings. For fast streaming (Fig. 7) the measured axial streaming velocity in the center of the guide tends to zero and is not in agreement with any available theory.

Under steady-state conditions and through cross section, the streaming measurements are consistent with the principle of conservation of mass: the average flow going in the negative direction cancels out the average flow going in the positive direction.

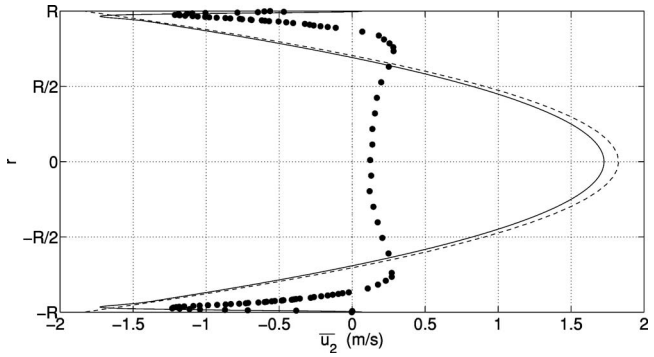


FIG. 7. Axial streaming velocity in guide section for $Re_{NL}=98$, $f_{ac}=88$ Hz, $x=\lambda/8$; —: Rayleigh expression [Eq. (1)]; ---: Bailliet *et al.* (Ref. 3) expression [Eq. (3)]; ●: measurements.

Figure 8 represents the evolution of axial component of the centerline ($r=0$) streaming velocity evolution $\overline{u_{2,c}}$ as a function of the Reynolds number of streaming motion Re_{NL}. The measurements are compared to Eq. (1) (parabolic solution). This diagram shows that the radial dependence of $\overline{u_2}$ departs from parabolic as Re_{NL} increases. For $Re_{NL}<20$, it is a bit higher than parabolic curve but fits it quite well in accordance with previous studies[3] results. The departure from parabolic begins for $Re_{NL}\approx 20$. This is in agreement with the measurements of Thompson *et al.*[10] (Fig. 8, p. 1846, $x\approx 3L/20$) who found for $Re_{NL}\geqslant 20$ significant departure for $x=\lambda/25$ and for uncontrolled boundary thermal case. According to these authors, a significant correlation is observed between the thermoacoustically induced temperature gradient that increases with the Reynolds number and the behavior of the streaming: as the magnitude of the temperature gradient increases, the magnitude of the streaming decreases and the shape of the streaming cell becomes increasingly distorted.

In our experiments, measurements were performed for higher Re_{NL} than in any previous studies and Fig. 8 shows three changes in the slope of the curve that gives the centerline streaming velocity as a function of the Re_{NL}. As stated above, for $Re_{NL}\approx 20$, the experimental curve departs from the theoretical slow streaming ($Re_{NL}\leqslant 1$) expectation (first change). Then, for $20<Re_{NL}<70$, $\overline{u_{2,c}}$ is hardly modified when the Re_{NL} is increased (second change). Then, for $70<Re_{NL}<140$, the slope of the curve increases and in the last stage for $Re_{NL}>140$, when the Re_{NL} is increased the centerline axial streaming velocity decreases. In Sec. IV B of this paper, this behavior will be connected to the evolution of the near wall streaming for increasing Re_{NL}.

In our case, the glass section of the resonator is simply surrounded by air corresponding to an uncontrolled boundary thermal condition (although the axial temperature distribution along the waveguide has not been measured). Figure 9 shows the axial component of the measured streaming velocity field at a Reynolds number of $Re_{NL}=5$. At this Reynolds number the discrepancies between the measured data and the theory are small. Figure 9 is in agreement with the results of Thompson *et al.* at the same Reynolds number[22] and the evolution of streaming velocity wave form for increasing Re_{NL} confirms that our experimental is uncontrolled boundary thermal condition.

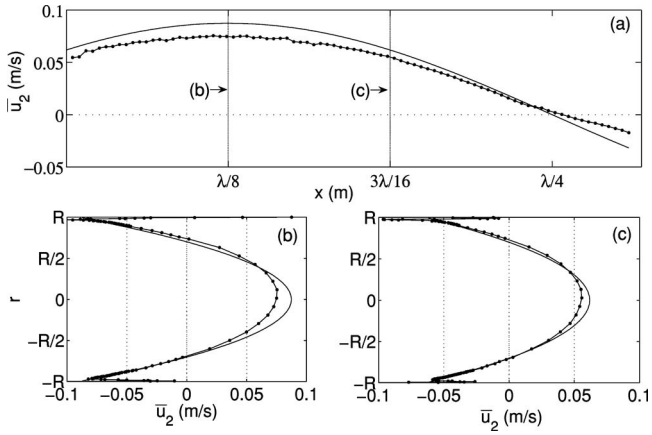


FIG. 9. Axial streaming velocity in guide section for $Re_{NL}=5$; —: Rayleigh theory [Eq. (1)]; ●: measurements. (a) Variation of axial streaming velocity with respect to the axial coordinate x ($r=0$) is compared to the theory of Rayleigh. (b) and (c) Variation of axial streaming velocity with respect to the transverse coordinate r ($x=\lambda/8$ and $3\lambda/16$) is compared to the theory.

B. Inner streaming vortices in the near wall region

Let us turn now to the inner vortices, that are near wall streaming vortices whose directions of rotation are opposite to those of the outer cells [see Fig. 1(a)] and that have never been measured before. When the velocity amplitude is increased, different stages for the generation and evolution of these inner streaming vortices can be determined. In order to focus on inner vortices, we zoom over the near wall region and consider the streaming velocity $\overline{u_2}$ at a distance from 0 to about $30\delta_\nu$ from the wall. Recall that a streaming vortex is detected in the guide when the streaming velocity crosses the abscissa axis $r=0$ corresponding to the center of the vortex [see Fig. 1(b)].

1. First stage

Figure 10 represents the axial streaming velocity dimensioned by the Rayleigh centerline axial streaming velocity $u_{2,\mathrm{Rayleigh},c}$ (normalized streaming velocity) as a function of the distance to the wall $R-r$ dimensioned by the thickness of the boundary layer δ_ν. For each stage, arrows visualize the evolution of streaming velocity maximum displacement for an increasing Re_{NL}. For the first stage, Fig. 10 shows that only one vortex is measured over the half section of the guide. When the Reynolds number is increased, the normalized streaming velocity decreases in the guide center region (see Fig. 8) and increases in the near wall region. The streaming velocity is maximum at the same position for increasing Re_{NL} across the guide section: the center and at $3\delta_\nu$. This last position corresponds to the position of the maximum negative value of expression of Eq. (3) (which is valid only for $Re_{NL}\ll 1$).

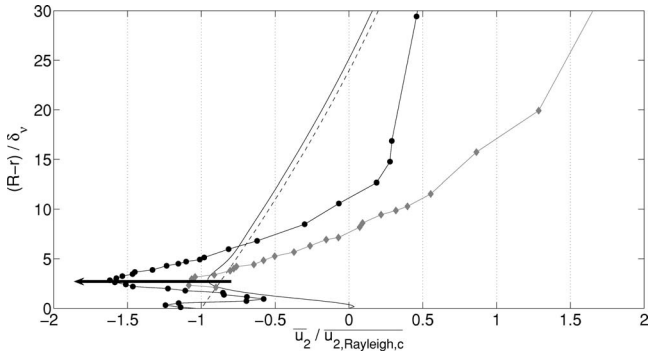


FIG. 10. First stage for the generation and evolution of streaming vortices; ---: Eq. (1); —: Eq. (3); ◇: measurements for $Re_{NL}=3$, ●: measurements for $Re_{NL}=13$ ($x=\lambda/25$). The arrow shows the displacement of the streaming velocity maximum for an increasing Re_{NL}.

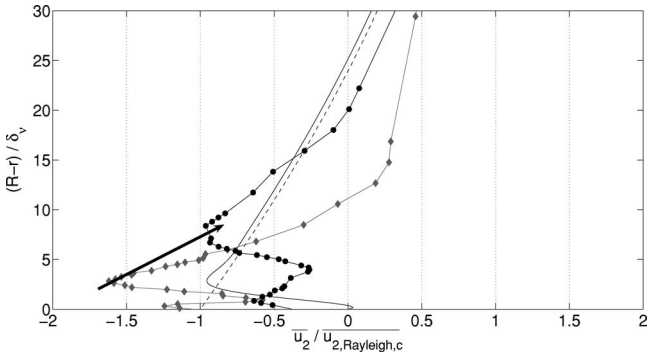


FIG. 11. Second stage for the generation and evolution of streaming vortices; ---: Eq. (1); —: Eq. (3); ◇: measurements for $Re_{NL}=13$, ●: measurements for $Re_{NL}=59$ ($x=\lambda/25$).

2. Second stage

Figure 11 shows that in the second stage, one vortex is measured as in the case of the first stage. When the Reynolds number is increased, the maximum normalized velocity of the streaming decreases near the guide center (in agreement with section A) and near the guide wall (conversely to the first stage). Also, the maximum of streaming vortex velocity near the wall offsets to the center: $3\delta_\nu$ for $Re_{NL}=13$ to $7\delta_\nu$ for $Re_{NL}=59$. Therefore the outer vortex becomes narrower and its maximum normalized velocity keeps on decreasing.

For each position x in the waveguide, only three stages out of four are observed. For $x=\lambda/25$, only the three first stages are visible in our experiments. Then, observations for the stage 3 and the stage 4 are made for $x=\lambda/15$.

3. Third stage

For the third stage, Fig. 12 shows that two vortices are measured in the near wall region. When the Reynolds number is increased, the thickness and the maximum normalized

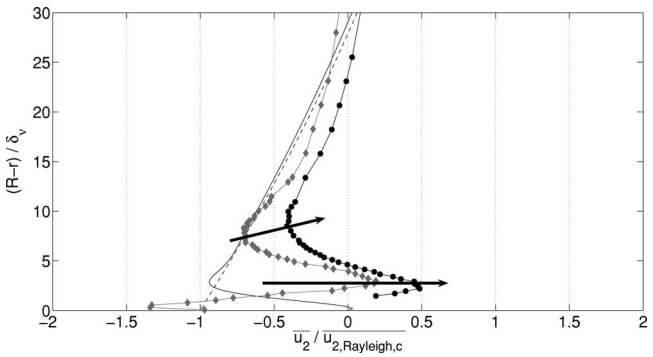


FIG. 12. Third stage for the generation and evolution of streaming vortices; ---: Eq. (1); —: Eq. (3); ◇: measurements for $Re_{NL}=78$, ●: measurements for $Re_{NL}=132$ ($x=\lambda/15$).

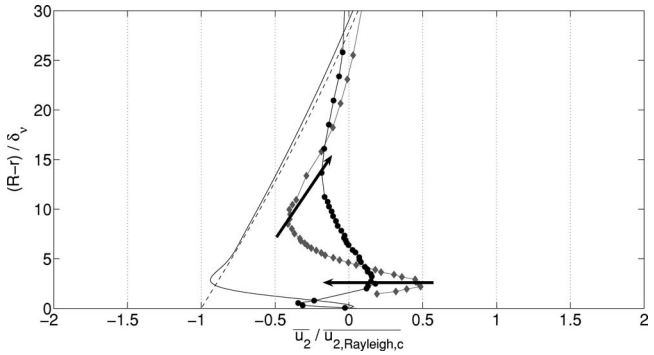


FIG. 13. Fourth stage for the generation and evolution of streaming vortices; ---: Eq. (1); —: Eq. (3); ◇: measurements for $Re_{NL}=132$, ●: measurements for $Re_{NL}=211$ ($x=\lambda/15$).

velocity near the wall of the outer streaming vortex continues to decrease. The thickness of the second vortex increases but the one of the third vortex (vortex near the wall) remains equal. Both the maximum normalized velocity of the second vortex near the wall and the maximum normalized velocity of the third vortex near the center increase. The maximum normalized velocity of the second vortex increases near the guide wall and decreases near the guide center as described for the outer vortex in the first stage. The position of the maximum positive value near the wall ($3\delta_\nu$ in Fig. 12) does not change when the Re_{NL} is increased: it corresponds to the position of the maximum negative value of outer streaming in stage 1 and also in the theoretical expression given by Eq. (3) (see first stage). We can suspect this position to be critical for vortices generation.

4. Fourth stage

For the fourth and last stage of our measurements, Figure 13 shows three vortices detected in half a section of the guide. When the Reynolds number is increased, the thickness and the maximum normalized velocity of the outer streaming vortex continues to decrease. The maximum normalized velocity of the second vortex decreases but its thickness increases. The maximum normalized velocity of the second vortex increases near the guide wall and near the guide center as was the case for the outer vortex in stage 2. The second vortex maximum velocity follows the same evolution as the outer vortex: the maximum normalized velocity is increased (first stage for the outer vortex and third stage for the second vortex) and then decreased (second stage for the outer vortex and fourth stage for the second vortex). The thickness of the third vortex remains equal.

5. Transition from stage to stage

Table I shows an outline of the experimental results. Combinations of frequency f_{ac}, axial coordinate to acoustic wavelength ratio x/λ, acoustic velocity amplitude U and Reynolds number Re_{NL} are defined. Results show that the transition of stage 1 to stage 2 occurs for $Re_{NL}\approx 20$, the transition from stage 2 to stage 3 for $Re_{NL}\approx 70$ and the transition from stage 3 to stage 4 for $Re_{NL}\approx 140$. These critical Re_{NL} should be connected to the ones given by Fig. 8 for the outer streaming velocity evolution so that, remaining the previous analysis of inner vortices evolution, we can think there exists a transfer of momentum between far and near wall region as the streaming becomes faster.

TABLE I. Measurements parameters and stage number.

x/λ	f_{ac} (Hz)	U (m/s)	Re_{NL}	Stage
1/25	84	8	3	1
1/25	84	12	8	1
1/25	84	15	13	1/2
1/25	84	22	27	2
1/25	84	27	41	2
1/25	84	32	59	2
1/25	84	36	74	3
1/25	88	9	5	1
1/25	88	15	13	1
1/25	88	22	29	1/2
1/25	88	29	51	2
1/25	88	34	68	2
1/25	88	41	98	3
1/15	113	12	12	2
1/15	113	19	28	2
1/15	113	22	38	2
1/15	113	26	53	2
1/15	113	32	78	3
1/15	113	38	115	3
1/15	113	41	132	3/4
1/15	113	45	159	4
1/15	113	53	211	4
1/10	150	15	22	2
1/10	150	21	47	2
1/10	150	26	71	2
1/10	150	37	137	3
1/10	150	49	247	3/4

Similar tendencies were observed for different positions: $x=\lambda/25$, $x=\lambda/15$, and $x=\lambda/10$ but not all along the axis. For measurements presented in this paper, the dependence between the ratio x/λ and the transition for the different stages could not be clarified.

V. CONCLUSION

The variation of axial streaming velocity dimensioned by the Rayleigh centerline velocity with respect to the transverse coordinate r was measured in an uncontrolled thermal case and for $\lambda/25<x<\lambda/10$. When the Reynolds number is increased until for $Re_{NL}\approx 20$, the maximum normalized velocity of the outer streaming vortices decreases in the center of guide section and increases in the near wall region. For further increasing $Re_{NL}(>20)$, the thickness and the maximum normalized velocity of the outer streaming vortex further decreases. The radial dependence of the axial component of the streaming velocity departs from parabolic Rayleigh theory for $Re_{NL}>20$ and are not in agreement with any available theory.

In the measurements of acoustic streaming reported previously in the literature, the near wall region was not addressed. In the present study, for $x=\lambda/10$, the variation of axial streaming velocity dimensioned by the Rayleigh centerline velocity in the near wall region with respect to the transverse coordinate r was measured. At $Re_{NL}\approx 70$, two

new vortices appear near the guide wall. The thickness of the streaming vortex nearest the wall remains equal and the thickness of second streaming vortex near the wall increases with the Reynolds number.

Similar tendencies were observed for other positions but not all along the tube axis. It would be interesting to repeat the experiments described in the present study for different positions x along the guide and to link the obtained results with the wave harmonics. From a theoretical point of view, there is now need to go further in the development of fast streaming studies such as the one of Menguy and Gilbert,[18] to compare this experimental behavior to theoretical predictions, especially in the near wall region. Also for thermoacoustic devices, it would be interesting to repeat the experiments for narrow tube, although this is a great experimental challenge.

ACKNOWLEDGMENTS

The authors wish to acknowledge the technical support of Laurent Philippon, Patrick Braud, Philippe Szeger, and Daniel Epinoux. They thank the reviewer for help in improving this paper.

[1]M. Hamilton, Y. Ilinski, and E. Zabolotskaya, "Acoustic streaming generated by standing waves in two-dimensional channels of arbitrary width," J. Acoust. Soc. Am. **113**, 153–160 (2003).
[2]J. R. Olson and G. W. Swift, "Acoustic streaming in pulse tube refrigerators: Tapered pulse tubes," Cryogenics **37**, 769–776 (1997).
[3]H. Bailliet, V. Gusev, R. Raspet, and R. A. Hiller, "Acoustic streaming in closed thermoacoustic devices," J. Acoust. Soc. Am. **110**, 1808–1821 (2001).
[4]M. Thompson and A. Atchley, "Measurements of Rayleigh streaming in high-amplitude standing wave," in *Nonlinear Acoustics at the Beginning of the 21st Century*, edited by O. V. Rudenko and O. A. Sapozhnikov (MSU Faculty of Physics, Moscow, 2002), Vol. 1, pp. 183–190.
[5]J. P. Sharpe, C. A. Greated, C. Gray, and D. M. Campbell, "The measurements of acoustic streaming using particle image velocimetry," Acustica **68**, 168–172 (1989).
[6]M. P. Arroyo and C. A. Greated, "Stereoscopic particle image velocimetry," Meas. Sci. Technol. **2**, 1181–1186 (1991).
[7]M. Campbell, J. A. Cosgrove, C. A. Greated, S. Jack, and D. Rockliff, "Review of LDA and PIV applied to the measurement of sound and acoustic streaming," Opt. Laser Technol. **32**, 629–639 (2000).
[8]D. B. Hann and C. A. Greated, "The measurement of flow velocity and acoustic particle velocity using particle-image velocimetry," Meas. Sci. Technol. **8**, 1517–1522 (1997).
[9]M. Thompson and A. Atchley, "Simultaneous measurement of acoustic and streaming velocities in a standing wave using laser doppler anemometry," J. Acoust. Soc. Am. **117**, 1828–1838 (2005).
[10]M. Thompson, A. Atchley, and M. Maccarone, "Influences of a temperature gradient and fluid inertia on acoustic streaming in a standing wave," J. Acoust. Soc. Am. **117**, 1839–1849 (2005).
[11]W. L. Nyborg, "Acoustic streaming," in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1965), Vol. 2B, Chap. 11, pp. 290–295.
[12]M. Faraday, "On a peculiar class of acoustical figures; and on certain forms assumed by groups of particles upon vibrating elastic surfaces," Philos. Trans. R. Soc. London **121**, 299–340 (1831).
[13]L. Rayleigh, "On the circulation of air observed in Kundt's tubes, and on some allied acoustical problems," Philos. Trans. R. Soc. London **175**, 1–21 (1884).
[14]N. Rott, "The influence of heat conduction on acoustic streaming," Z. Angew. Math. Phys. **25**, 417–421 (1974).
[15]Q. Qi, "The effect of compressibility on acoustic streaming near a rigid boundary for a plane traveling wave," J. Acoust. Soc. Am. **94**, 1090–1098 (1993).
[16]R. Waxler, "Stationary velocity and pressure gradients in a thermoacoustic stack," J. Acoust. Soc. Am. **109**, 2739–2750 (2001).
[17]M. Hamilton, Y. Ilinski, and E. Zabolotskaya, "Thermal effects on acoustic streaming in standing waves," J. Acoust. Soc. Am. **114**, 3092–3101 (2003).
[18]L. Menguy and J. Gilbert, "Non-linear acoustic streaming accompanying a plane stationary wave in a guide," Acta Acust. **86**, 249–259 (2000).
[19]The Reynolds number has a different definition in the articles of Menguy and Gilbert (Ref. 18) (to a different definition of δ_ν) and of Thompson and Atchley (Refs. 4, 9, and 10): their Reynolds number corresponds to our Reynolds number defined in Eq. (4) multiplied by two.
[20]S. Moreau, R. Boucheron, J. Valière, and H. Bailliet, "LDV and PIV measurements in the acoustic boundary layers," Proceedings of the 9th french-speaking congress of laser velocimetry, Brussels, Belgium, 14–17 September 2004.
[21]Here, because LDV measurements are brought back on one acoustic period, the relevant parameter for measurements convergence is not the data rate (number of samples divided by time of measurement) as usual but the number of samples.
[22]$\mathrm{Re}_{\mathrm{NL}}=10$ for M. Thompson, A. Atchley, and M. Maccarone (Ref. 10).

Numerical evaluation of tree canopy shape near noise barriers to improve downwind shielding

T. Van Renterghem[a] and D. Botteldooren
Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

(Received 9 August 2007; revised 16 November 2007; accepted 21 November 2007)

The screen-induced refraction of sound by wind results in a reduced noise shielding for downwind receivers. Placing a row of trees behind a highway noise barrier modifies the wind field, and this was proven to be an important curing measure in previous studies. In this paper, the wind field modification by the canopy of trees near noise barriers is numerically predicted by using common quantitative tree properties. A realistic range of pressure resistance coefficients are modeled, for two wind speed profiles. As canopy shape influences vertical gradients in the horizontal component of the wind velocity, three typical shapes are simulated. A triangular crown shape, where the pressure resistance coefficient is at maximum at the bottom of the canopy and decreases linearly toward the top, is the most interesting configuration. A canopy with uniform aerodynamic properties with height behaves similarly at low wind speeds. The third crown shape that was modeled is the ellipse form, which has a worse performance than the first two types, but still gives a significant improvement compared to barriers without trees. With increasing wind speed, the optimum pressure resistance coefficient increases. Coniferous trees are more suited than deciduous trees to increase the downwind noise barrier efficiency.
 [DOI: 10.1121/1.2828052]

PACS number(s): 43.28.Fp, 43.50.Gf, 43.28.Js [VEO] Pages: 648–657

I. INTRODUCTION

The screen-induced refraction of sound by wind is a well-known problem,[1–3] resulting in a reduced shielding efficiency of noise barriers in case of downwind sound propagation. The use of a row of trees behind noise barriers was shown to be an interesting solution to improve noise shielding from highways. In Ref. 4, a wind tunnel study at scale showed that changing the wind field near noise barriers by using synthetic windbreaks limits the negative effects of the wind to an important degree. In a monitoring campaign along a highway,[5] the positive effect of a row of trees behind a noise barrier was proven experimentally. Simultaneous noise recordings were made behind part of a long noise barrier with and without trees. In this way, the reduction in screen-induced refraction of sound was measured directly. The (downwind) microphone behind the trees yielded lower total A-weighted sound pressure levels resulting from traffic noise and this difference in levels increased with increasing wind speed. For a wind speed of 10 m/s at a height of 10 m above the ground, an increased shielding of about 4 dBA was observed due to the presence of the trees.[5] In case of (strong) upwind sound propagation, the measured effect of the presence of the trees was very limited.[5]

In Refs. 6 and 7, a numerical model was developed for this type of sound propagation problems, involving complex wind flows. The model was validated with success for the situations measured in the wind tunnel study.[4] Further, additional calculations were performed to find important parameters in situations where noise barriers and trees are combined. The focus was on noise barriers on either side of the acoustic source. The magnitude of the incident wind speed, the distance between the source and the noise barriers, the location and the height of the wind reducing structures, as well as the influence of the porosity of the windbreaks were studied. The properties of synthetic windscreens were used to simulate the wind field near noise barriers and trees. Measured pressure drops as a function of flow velocity, for screens with different porosities, were used for these calculations.

Practical recommendations concerning the type of trees that should be used behind noise barriers are however hard to derive from this previous study. For flat windscreens, the optical porosity (i.e., the percentage of open space as seen perpendicularly to the windscreen side) is sufficient to describe its aerodynamic properties.[8] In the case of a tree shelterbelt, two shelterbelts with similar optical porosities may have a very different arrangement of plant elements, different vegetative surface areas and volumes, and a different amount of open spaces within their canopies.[8] It can therefore be concluded that optical porosity is not a good measure to describe the wind flow through the canopies of real trees. This means that the results from the numerical predictions made in Ref. 6 cannot be translated directly to practice.

The calculations in Ref. 6 are further based on windscreens with a uniform porosity. In general, a large variation in the aerodynamic properties of the crown of trees with height is possible. Canopy shape was shown to be an important factor when looking at ground deposed particles in air quality modeling.[9] Changing the wind reducing properties of the canopy as a function of height will result in signifi-

[a]Author to whom correspondence should be addressed. Electronic mail: timothy.van.renterghem@intec.ugent.be

cantly altered wind fields. It is therefore interesting to study the influence of canopy shape on the refraction of sound near noise barriers.

This paper is organized as follows. In Sec. II, the simulation of the flow field near noise barriers in combination with trees is considered. Section II A discusses briefly how atmospheric boundary layer flows can be modeled accurately. In Sec. II B, the effect exerted by the canopy of trees on the wind flow is considered, and it is shown how this can be simulated within standard computational fluid dynamics (CFD) software. Section III discusses briefly the numerical model to simulate sound propagation near the noise barrier, and an overview of model parameters is given. In Sec. IV, numerical results are presented and discussed and in Sec. V, conclusions are drawn.

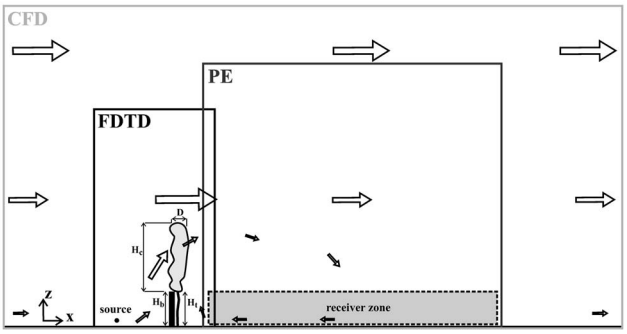


FIG. 1. Overview of the simulation areas, indicating the CFD, FDTD, and PE area. H_b indicates the barrier height, H_t the tree trunk height, H_c the canopy height, and D indicates the width of the canopy. The arrows give an indication of the wind direction and its magnitude near the noise barrier in combination with trees. The receiver zone is shown as well.

II. NUMERICAL SIMULATION OF FLOW FIELD

A. Atmospheric boundary layer flows

The two-dimensional velocity fields near the noise barriers are calculated with the CFD software FLUENT 6.3.[10] The Reynolds-averaged Navier–Stokes equations are solved by applying a standard $k-\varepsilon$ turbulence model. This "turbulence closure" model is widely applied in engineering applications, and is sufficiently accurate for the current application. Turbulent effects are introduced by means of two additional equations to quantify the turbulent kinetic energy and its dissipation rate. Accurate modeling of vertical gradients in the horizontal wind velocity component is of main concern. Predicted values of turbulent parameters are not of interest in the current application.

Vertical profiles of the horizontal wind velocity u_x, turbulent kinetic energy k, and turbulence dissipation rate ε need to be set at the upstream boundary condition. The following equations apply to a neutral, atmospheric boundary layer in equilibrium:[11]

$$u_x = \frac{u_*}{\kappa} \ln\left(1 + \frac{z}{z_0}\right), \tag{1}$$

$$k = \frac{u_*^2}{\sqrt{C_\mu}}, \tag{2}$$

$$\varepsilon = \frac{u_*^3}{\kappa(z + z_0)}, \tag{3}$$

where u_* is the friction velocity, κ is the Von Karman constant (equal to 0.4), z is the height above ground level, and z_0 is the aerodynamic roughness length. C_μ is a model constant of the $k-\varepsilon$ model which is parameterized by measurements. It relates k and ε to the turbulent dynamic viscosity μ_t by the following relation:

$$\mu_t = C_\mu \rho \frac{k^2}{\varepsilon}, \tag{4}$$

where ρ is the mass density of air. The value of C_μ is usually set to 0.09.[11]

The flow simulations are performed for friction velocities of 0.4 and 0.8 m/s. The aerodynamic roughness length equals 0.01 m. These parameters fully define the inflow boundary conditions when using Eqs. (1)–(3). The dimensions of the two-dimensional simulation domain are expressed relative to the total length of the flow disturbing structures L, which is equal to the noise barrier height H_b plus the part of the canopy extending above the noise barrier. In the present simulations, the tree trunk height H_t is equal to H_b, whereas the canopy height H_c equals $2H_t$. As a consequence, L equals $3H_b$ (see Fig. 1). Boundary conditions are imposed at sufficient distances from the flow disturbing structures. The height of the computational grid is $25L$. A region of $9L$ upstream and $34L$ downstream from the noise barrier is modeled. Along the length of the top boundary condition, constant values of horizontal velocity, turbulent kinetic energy, and its dissipation rate are imposed, based on the values of the inlet conditions at this height. The outflow boundary condition of FLUENT[10], assuming that there are no stream-wise gradients, is used at the right-hand side of the grid. The recommendation given in Ref. 12, concerning accurate flow simulations in the atmospheric boundary layer, are followed.

B. Flow field near trees

The presence of a canopy—or any windbreak—has a significant effect on the flow field. Such structures exert a drag force on the wind field, causing a net loss of momentum in the (incompressible) flow. When the permeability of the windbreak decreases, the so-called "bleed flow" through the windbreak decreases and the drag force increases. This is accompanied by a stronger upward deflection of the approach flow.

The airflow through the canopy of trees results in a pressure drop. The pressure resistance coefficient k_r is a commonly used measure to quantify this pressure drop, and is defined as follows:

$$\Delta p = k_r \frac{\rho u_x^2}{2}. \tag{5}$$

The pressure resistance coefficient can be related to physical characteristics of the canopy of trees. When assuming that the aerodynamic drag of the canopy balances the pressure drop, one may write, following Ref. 13:

$$k_r \approx \int_0^D C_d \,\mathrm{LAD}\, dx, \tag{6}$$

where D is the width of the canopy layer in horizontal direction, C_d is the dimensionless drag coefficient of the elements of the canopy, and LAD is the leaf area density. The LAD is defined as the total area of leafs per unit volume of the canopy.

Drag coefficients of trees are independent of the wind speed encountered outdoors near ground level.[14] Information concerning the drag coefficient of different types of trees can be found in literature. Values for individual deciduous trees range from 0.15 to 0.25.[15–18] A value of 0.2 is commonly used in numerical studies of wind flow through forests. In Ref. 14, drag coefficients were measured for single-row deciduous "windbreak" species. Most values are near 0.5. Coniferous types of trees are characterised by somewhat larger values (0.6–1.2).[14]

The LAD or the equivalent needle area density (NAD) for conifers, is a common measure in quantitative plant research. As trees can form extremely diverse crown shapes, the LAD or NAD may depend largely on height. The crown form is not only dependent of the species, but also on the local topography, climate, the availability of nutrients, etc. The difference between isolated trees and trees in a dense tree population can be large as well.[9] A typical value of the maximum LAD of the canopy of deciduous trees is 1 $m^2\ m^{-3}$,[18] however large deviations from that value appear. The NAD is usually larger. In Refs. 19 and 20, maximum values of 2 $m^2\ m^{-3}$ were found.

The canopy width D depends on the height as well, and has a maximum value typically in the order of a few meters for a single row of trees. Assuming that the LAD and C_d are constant in horizontal direction, Eq. (6) becomes:

$$k_r \approx C_d \,\mathrm{LAD}\, D. \tag{7}$$

In the remainder of this paper, k_r will be used as an independent variable, and is representative for a variety of combinations of the drag coefficient, the LAD, and the canopy width, as governed by Eq. (7).

A velocity-dependent pressure drop over a plane can be modeled in FLUENT[10] by using the porous jump boundary condition. The pressure drop over this plane represents the total pressure drop as caused by the air flow through the canopy. The resistance coefficient can be made dependent on height to account for vertical changes of LAD.

Three different crown shapes were considered. The barrier height H_b and the trunk height H_t equal 4 m, and the canopy height H_c equals 8 m. A first canopy type has uniform aerodynamic properties with height, and is further indicated as "uniform." This type of canopy form is representative for, e.g., a dense hedge. A second type has a maximum pressure resistance coefficient near the top of the noise barrier (or at the bottom of the canopy) and a linear decrease towards the top of the canopy. This type is further indicated as "triangle," and is typical for conifers. A third canopy type which is considered has an ellipse-like form, with a maximum k_r near the middle of the canopy, and large gradients in k_r near the top and bottom of the crown. This type is further indicated as "ellipse," and is representative for common deciduous trees. The sum of k_r over the total canopy height is kept the same for these three crown forms. This allows investigating the importance of the distribution of the aerodynamic properties over height. An overview of k_r with height for these crown types is given in Fig. 2, for an average k_r equal to 1.

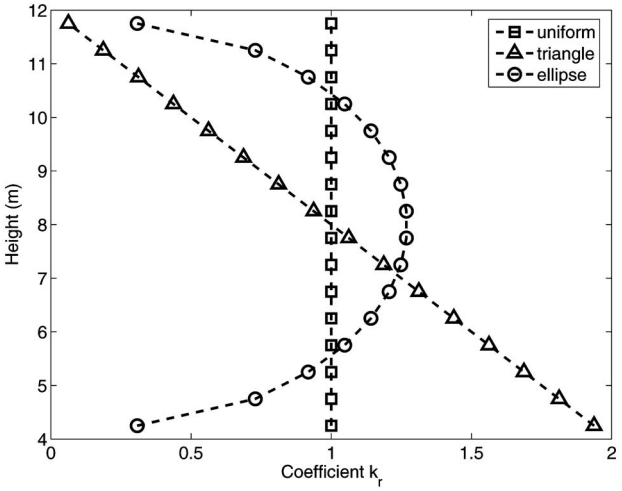


FIG. 2. The pressure resistance coefficient k_r for the three types of crown shapes (uniform, triangle, and ellipse) used for the numerical calculations in this paper. The average values of k_r equal 1.

Numerical calculations are performed for average values of k_r equal to 1, 2, and 4, for the three crown shapes considered. Note that k_r is the product of the drag coefficient, leaf area density, and canopy width. The values used in the calculations cover a wide variety of realistic situations. A specific canopy form is prescribed in a vertical resolution of k_r equal to 0.5 m.

III. SOUND PROPAGATION MODEL

The acoustic calculations are performed with the finite-difference time-domain (FDTD) method, coupled to the parabolic equation (PE) method.[21] Two-dimensional calculations are performed, implying a coherent line source, and infinitely long noise barriers with constant cross sections. Traffic, which is the prominent source when looking at noise barriers, is however more accurately modeled as an incoherent line source. When looking at noise barrier efficiency at individual frequencies, significant differences are observed when comparing calculations made with a coherent and incoherent line source.[22] When averaging to octave bands, as will always be done in this paper, differences become much smaller. An approach such as the one proposed in Ref. 22 cannot be used as the propagation medium is moving. The large number of three-dimensional calculations needed to simulate an incoherent line source more accurately would lead to huge computing times. Nevertheless, the difference between a coherent and incoherent line source on the main quantity used in this paper, namely the tree effect (see Sec. IV for its definition), is expected to be limited.

The FDTD method, solving the moving-medium sound propagation equations,[23,7,24] is used in the direct vicinity of

the noise barrier. An overview of the different regions of the simulation domain, with the corresponding numerical method, is shown in Fig. 1. The stationary flow field as calculated by the CFD software is used as a so-called background flow. This implies that refraction of sound by wind is accounted for accurately, but the acoustic waves do not influence the ambient flow, and generation of sound by wind is not considered. These two latter effects are however not important in the current application. Upward directed flows very close to the barrier and trees are accounted for. Perfectly matched layers are applied at the left, right, and top boundaries of the FDTD computational domain, to simulate an unbounded atmosphere. More information on the numerical schemes is given in the following. In absence of flow, the efficient staggered spatial and staggered temporal grid is used.[25] In a moving medium, staggered-in-space calculations are combined with the prediction-step staggered in time (PSIT) approach.[26] Such a scheme was shown to be an interesting compromise between accuracy, numerical stability, and computational efficiency.[26] Further, flow velocities and propagation distances in the FDTD region are sufficiently low to perform accurate calculations with the PSIT scheme.

The Green's function PE (GFPE) method[27,28] is used to model sound propagation from the source region to the receivers. The GFPE calculations start from a column of complex pressures, derived from the FDTD domain. Refraction is modeled using the effective sound speed approach. A horizontal flow with range-dependent wind speed profiles is assumed. This is a good approximation at sufficient distance behind the noise barrier. On top of the computational PE domain, an absorbing layer was placed.

Combining FDTD in the source region with GFPE downwind from the noise barrier allows modeling the effects of the complex flow field near the source, barrier, and trees accurately, but explores at the same time the efficiency of the GFPE method for the longer distance part of the outdoor sound propagation problem. This hybrid model was shown to be computationally very efficient, without resulting in loss of accuracy. Details concerning the coupling between FDTD and PE can be found in Ref. 21.

The noise barrier height H_b equals 4 m. The noise barrier thickness equals 0.1 m. The source is placed at $2H_b$ upwind from the barrier, at a height of 0.30 m. The PE calculations start at $1H_b$ downwind from the barrier, and continue until $35H_b$. The noise barrier and the ground in the FDTD region are modeled as rigid planes, whereas in the PE region both a rigid ground and grass-covered ground is modeled. For the latter, the common Delany and Bazley model[29] is used, with an effective flow resistivity equal to 200 kPa s/m^2. Downwind sound propagation is considered, for two incident wind speed profiles, characterized by friction velocities of 0.4 and 0.8 m/s (see Sec. II A).

Scattering of sound on tree elements is not considered in this paper as this is mainly a high-frequency phenomenon. Measurements behind a noise barrier with and without deciduous trees, in the absence of wind, showed that below 1.5 kHz, scattering is smaller than 1 dB.[6] At 10 kHz, a difference of 6 dB was measured.[6] Traffic however produces only a small amount of acoustic energy in this high frequency range relative to low frequency bands. So the contribution of this scattered sound to the total A-weighted sound pressure levels is small.[6] This conclusion was confirmed by several authors. In Ref. 30, it was concluded that roadside trees do not significantly influence traffic noise at ground level. Even belts of trees of several tens of meters result in only little attenuation for traffic noise.[31] Martens states that the foliage of trees can be seen as a low-pass filter: The frequencies of the dominant peaks in traffic noise are too low to be amplified or weakened.[32]

The presence of a noise barrier in wind results in an increase in turbulence compared to the amount of turbulence observed over unobstructed ground. This can cause scattering of sound into the shielded area, thereby reducing the noise barrier efficiency. Downwind sound propagation calculations through screen-induced turbulence, for a similar configuration as the one considered in this paper, were performed for a sound frequency of 500 Hz in Ref. 33. It was shown that turbulent scattering results in fluctuations in the noise level in the shielded region up to 3 dB at 250 m from the source. The time-averaged effect, on the other hand, was only of the order of 0.2 dB. Therefore, it was concluded that screen-induced turbulence could be neglected when looking at average noise levels.

The main interest in this study is in shielded traffic noise. Therefore, calculations up to the octave band of 1000 Hz are sufficient. The maximum frequency to be considered is 1405 Hz, and lies within the 1 dB region for scattering. The octave bands with center frequencies 63, 125, 250, 500, and 1000 Hz are used in the analysis. To calculate the energetically average sound pressure level in each octave band, 15 frequencies are considered.

The following computational parameters were used. The FDTD spatial discretization step was 0.02 m in both dimensions. This led to more than ten cells per wavelength for the highest frequency considered. The temporal discretization step was 40 μs, and 5000 time steps were sufficient to build the PE starting fields ranging from ground level until 40 m. The perfectly matched layers at the boundaries of the domain consisted of 40 computational cells. A broadband Gaussian pulse was emitted at the source position. For the PE calculations, ten computational cells per wavelength were used in vertical direction. The horizontal propagation step was equal to a single wavelength, in order to have sufficient spatial resolution when plotting sound pressure fields and to accurately account for the rapid changes of the wind speed profiles in the lee of the barrier. At each propagation step, the wind speed profile was updated. The thickness of the PE damping layer was 150 times the wavelength. To avoid spatial interpolation while calculating octave band values, the spatial parameters of the PE calculation were kept constant for each of the 15 frequencies in the octave band considered and corresponded to the highest frequency in that band.

IV. RESULTS AND DISCUSSION

The zone of interest for quieting (i.e., the receiver zone, see Fig. 1) extends in horizontal direction from $1H_b$ to $35H_b$, and in vertical direction from ground level up to $1H_b$. Dis-

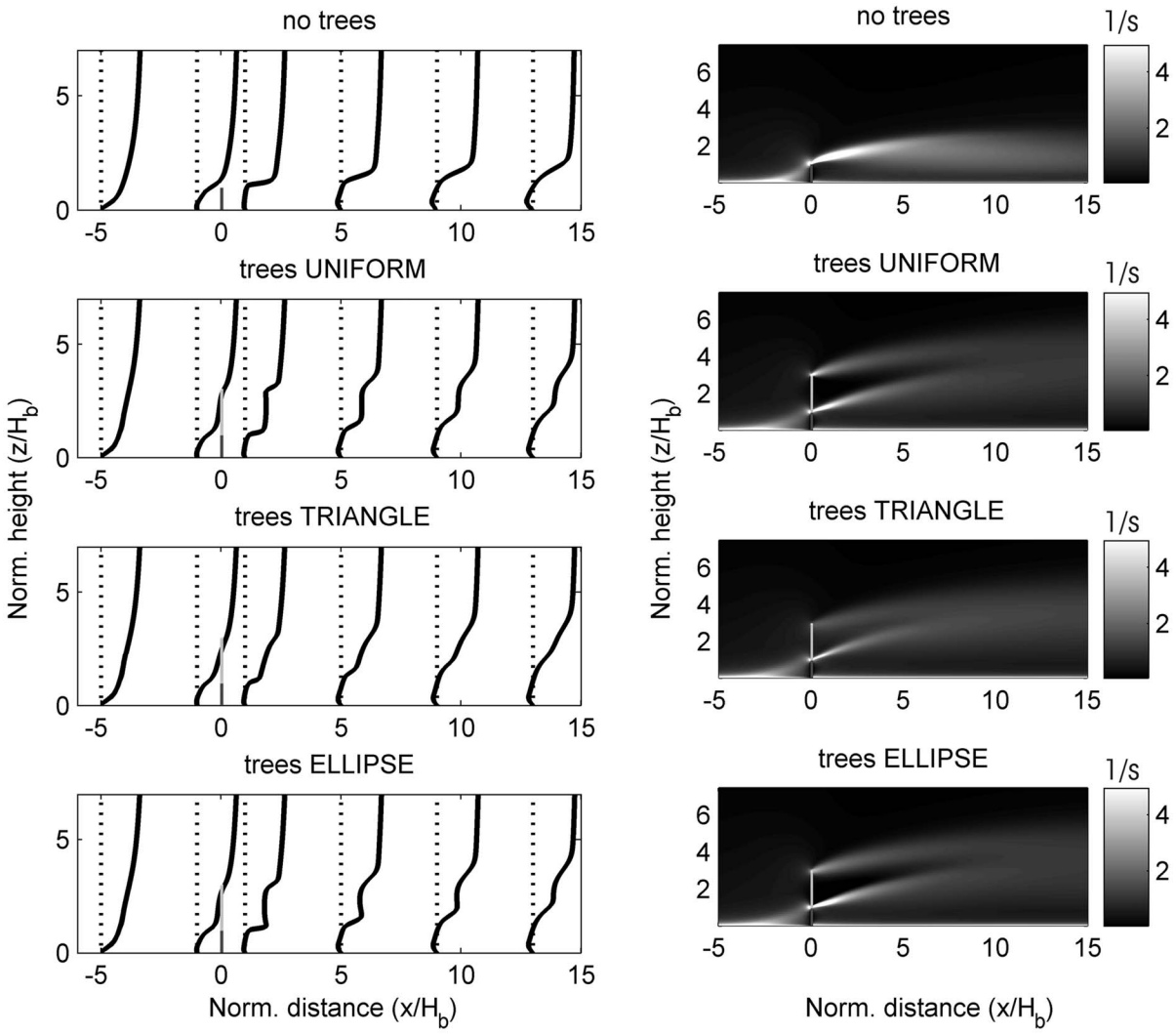


FIG. 3. In the left-hand column, vertical profiles of the horizontal component of the wind speed are shown at selected locations. In the right-hand column, the corresponding fields of vertical gradients in the horizontal wind speed are presented. Only positive gradients are shown. Distances and heights are expressed relative to the noise barrier height H_b, which equals 4 m. The first row of figures shows the fields when no trees are present near the noise barrier. In the second row, the uniform canopy shape, in the third row the triangular form, and in the last row the ellipse shape is considered. The average values of k_r equal 2. The incident friction velocity u_* equals 0.8 m/s.

tances and heights are expressed relative to the noise barrier height, but this does not imply that scaling is possible for the calculations performed in this study. The results will be presented as contour plots of sound levels, as sound levels along a horizontal line at a fixed height, or as histograms indicating the fraction of the area of the receiver zone falling within a certain sound level class. A bin increment of 1 dB will be used in the histograms. Full octave band sound pressure levels are considered. To limit the number of contour plots, only the octave bands with center frequency 125 and 1000 Hz are shown. These frequency bands are representative for respectively the engine noise and tire-road noise peaks in typical traffic spectra.[34]

The insertion loss IL is defined as the sound pressure level in absence of a noise barrier, minus the sound pressure level in presence of a barrier, for the same source receiver configuration. The screen-induced refraction of sound by wind SIROS is the sound pressure level with the noise barrier in the presence of wind, minus the sound pressure level with the noise barrier in absence of wind, for the same source-noise barrier–receiver configuration. A third quantity that will be used is the tree effect TE, which is defined as the sound pressure level in the presence of a noise barrier and wind, minus the sound pressure level in the presence of a noise barrier combined with trees and wind, for the same source–noise barrier–receiver configuration. The following equations give an overview of the definitions of IL, SIROS, and TE:

$$\mathrm{IL} = L_{p,\text{no barrier,no trees,no wind}} - L_{p,\text{barrier,no trees,no wind}}, \quad (8)$$

$$\mathrm{SIROS} = L_{p,\text{barrier,no trees,wind}} - L_{p,\text{barrier,no trees,no wind}}, \quad (9)$$

$$\mathrm{TE} = L_{p,\text{barrier,no trees,wind}} - L_{p,\text{barrier,trees,wind}}. \quad (10)$$

Positive values of IL indicate that the noise barrier is effective in reducing sound pressure levels in absence of wind. Positive values of SIROS indicate that the wind reduces the

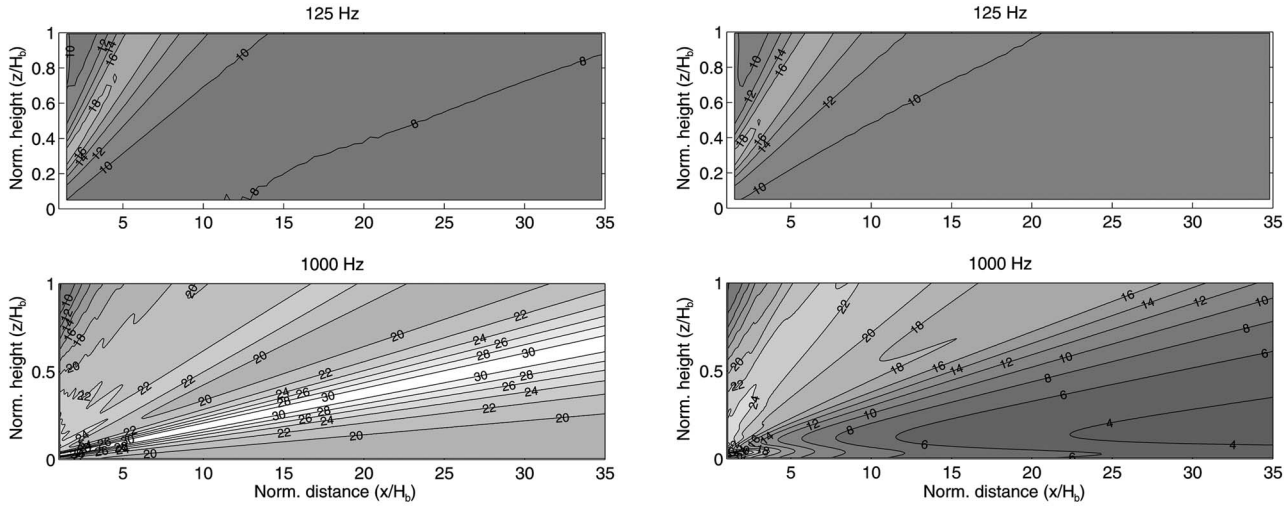


FIG. 4. Contour plots of IL (insertion loss) for the octave bands with center frequencies 125 and 1000 Hz. The noise barrier height H_b equals 4 m. On the left, a rigid ground is assumed downwind from the noise barrier. Plots on the right correspond to a grass-covered ground downwind from the noise barrier.

barrier efficiency. Positive values of TE indicate that the presence of trees increases shielding when there is wind, or alternatively, part of the SIROS is counteracted.

Vertical gradients in the horizontal component of the wind speed determine the magnitude of the screen-induced refraction of sound in the shielded area. In Fig. 3, profiles of the horizontal wind component are shown at selected locations near the noise barrier, and near the noise barrier in combination with the different canopy shapes. Fields plots of positive, vertical gradients in the horizontal wind speed are shown as well. The friction velocity of the inflow boundary condition was 0.8 m/s, and an average value of k_r equal to 2 is used. In absence of trees, large positive gradients in the wind speed are observed starting from the top of the noise barrier, stretching in downwind direction. In the presence of trees, the downwind area with large positive gradients is significantly reduced. Near the top of the trees however, additional gradients appear. Such gradients are most prominent in case of uniform canopy properties with height. The triangular tree shape, on the other hand, induces a smooth transition between the top region of the canopy and the undisturbed region outside the canopy, resulting in smaller gradients at the tree top. The gradients near the top of the barrier are most significantly reduced in case of the triangle crown shape: the maximum values of k_r are found just near the noise barrier top. These barrier-top gradients are somewhat larger for the ellipse form than for the uniform canopy.

In the case of a friction velocity equal to 0.4 m/s, similar conclusions could be drawn. The maximum gradients that are found near the top of the noise barrier and the top of the canopy stay more or less the same. However, the region with positive gradients is much smaller, and only appears close to the barrier and trees.

Contour plots of IL (in absence of wind) in the receiver zone are shown in Fig. 4, for a rigid and grass-covered ground. With increasing octave band center frequency, the IL

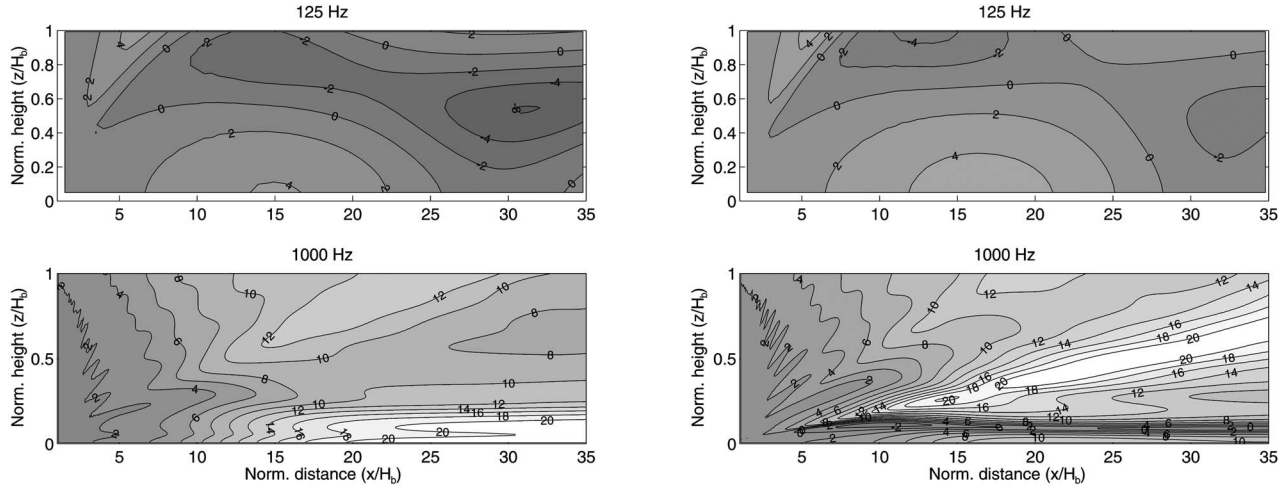


FIG. 5. Contour plots of SIROS (screen-induced refraction of sound by wind) for u_*=0.8 m/s. The octave bands with center frequencies 125 and 1000 Hz are shown. The noise barrier height H_b equals 4 m. On the left, a rigid ground is assumed downwind from the noise barrier. Plots on the right correspond to a grass-covered ground downwind from the noise barrier.

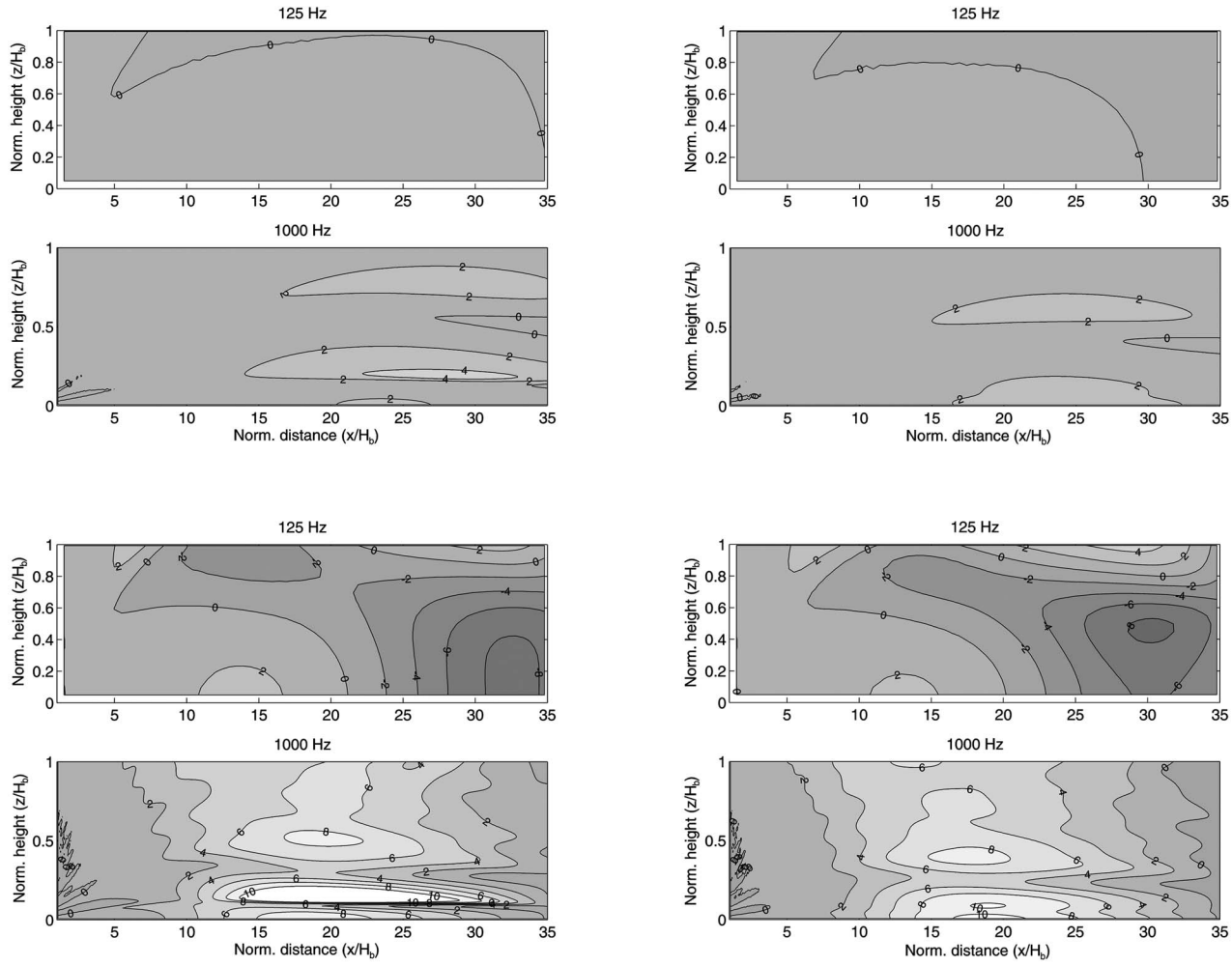


FIG. 6. Contour plots of TE (tree effect). The panels above are for $u_*=0.4$ m/s, the panels below for $u_*=0.8$ m/s. The octave bands with center frequencies 125 and 1000 Hz are shown. The noise barrier height H_b equals 4 m. On the left, a rigid ground is assumed downwind from the noise barrier. Plots on the right correspond to a grass-covered ground downwind from the noise barrier. A triangle crown shape is used, with $k_r=2$.

becomes larger. For the low frequencies, the differences between rigid ground and grass-covered ground are small. For the higher frequencies, the IL in case of a softer ground is smaller, and this is very significant for the octave band with center frequency 1000 Hz: A large zone with values for the IL lower than 5 dB is observed. This is explained by the reduction of the positive influence of soft ground on propagation from this low lying source by the presence of the noise barrier.

Contour plots of SIROS, in case of a friction velocity of 0.8 m/s, are shown in Fig. 5. For the octave band of 63 Hz, refraction is limited. For the octave band of 125 Hz, values of SIROS are negative at some locations, indicating that the wind results in a (limited) decrease of the sound pressure level.

For higher frequencies, values for SIROS are larger and mainly positive. At 1000 Hz, values exceed 10 dB starting from about $15H_b$ downwind from the noise barrier, and maximum values found in the region of interest are larger than 20 dB. In case of the rigid ground, and especially for the octave band of 250 Hz, zones of negative SIROS are still found. This is caused by a shift in the location where conditions for destructive interference are met.

Contour plots of TE are shown in Fig. 6, in case of a friction velocity of 0.4 and 0.8 m/s, and for a rigid ground and grass-covered ground. A triangular crown shape is considered, equivalent to a uniform canopy with $k_r=2$. The influence of the trees is very small at the octave band with center frequency 63 Hz, as screen-induced refraction of sound is limited as well.

For the lower wind speed ($u_*=0.4$ m/s), the wind modification by the trees has a significant effect starting from the octave band of 250 Hz. In case of a rigid ground, maximum effects exceed 5 dB starting from 500 Hz. For the softer ground, maximum effects are somewhat smaller, but zones with negative TE are hardly present.

For the higher wind speed ($u_*=0.8$ m/s), the maximum values, either positive or negative, are much larger. The maximum improvements by the presence of trees in a wind situation now exceed 10 dB starting from 250 Hz. Large zones with positive effects, over the full height of the receiver zone, are found for the octave bands of 500 and 1000 Hz. The region of significant improvement is found roughly between $10H_b$ and $30H_b$ downwind from the noise

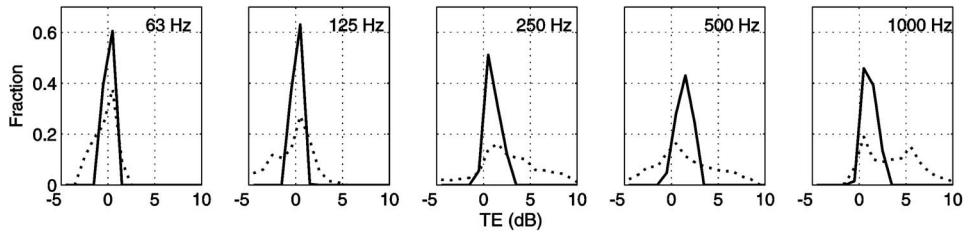


FIG. 7. Distribution of TE over decibel classes in the receiver zone, for a noise barrier height H_b which equals 4 m. Friction velocities of 0.4 m/s (full lines) and 0.8 m/s (dotted lines) are considered. The octave bands with center frequencies 63, 125, 250, 500, and 1000 Hz are shown. A grass-covered ground is assumed downwind from the noise barrier. A triangle crown shape is used, with $k_r=2$.

barrier. The difference between a rigid ground and grass-covered ground becomes small, especially at 500 and 1000 Hz.

The histograms in Fig. 7 clarify the results. The distribution of the area in the receiver zone over TE classes becomes broader with increasing wind speed. At 125 Hz, large zones with negative TE are found, for both ground types. The fraction of the receiver zone with negative values decreases with frequency, and at 1000 Hz only positive effects are found. This is observed for both wind speeds.

In the histograms in Fig. 8, a comparison is made between the TE for the different tree forms, for $u_*=0.4$ and 0.8 m/s. The average values of k_r are equal to 2 in all cases. For a friction velocity of 0.4 m/s, significant differences between the different crown shapes are observed, starting from the octave band of 250 Hz. The uniform and triangle shape give quite similar tree effects, which are better than the tree effect for an ellipse shape.

For a friction velocity of 0.8 m/s, only the 1000 Hz octave band seems to be significantly affected by tree form. This is however the dominant frequency band when looking at rolling noise near highways. A similar conclusion can be drawn when looking at the TE along a horizontal line, at a fixed receiver height of 2 and 4 m (see Fig. 9). At the distance where maximum effects are observed, the triangle shape gives an improvement of about 2–3 dB compared to the ellipse form. The uniform canopy shape lies in between the other types.

It can therefore be concluded that the triangular crown shape is the most interesting one, followed by the uniform one. For lower wind speeds, both types behave similarly. The ellipse form has a somewhat worse performance, but still improves the downwind shielding significantly compared to a noise barrier without trees. Maximum values are a few decibels smaller, while negative TE areas are more frequently observed.

The influence of the magnitude of k_r on TE is shown in Fig. 10 for a uniform crown. A grass-covered ground is considered downwind. Numerical predictions are shown for both the low wind speed and high wind speed.

For the low wind speed, uniform trees with $k_r=2$ are a significant improvement over trees with $k_r=1$, at all frequencies. The differences between $k_r=2$ and $k_r=4$ are less pronounced. Below 250 Hz, the difference between them is very limited. Above 250 Hz, $k_r=4$ results only in a limited additional improvement compared to $k_r=2$. It can therefore be concluded that values of k_r larger than 2 do not increase downwind shielding. For the high wind speed, k_r equal to 4 gives a significant improvement over $k_r=2$. Further, the larger the value of k_r, the smaller the fraction of the area of the receiver zone with negative TE, especially at higher frequencies. An asymptotical value is not found for the range of values of k_r that are modeled at the high wind speed. From this analysis, it is clear that conifers are preferred behind noise barriers to improve the downwind shielding, since their typical needle area densities and canopy element drag coefficients lead to larger pressure resistance coefficients. Further, the typical crown form of coniferous trees is close to the triangular canopy shape. In addition, during winter, there is no loss in biomass.

V. CONCLUSIONS

In this paper, the possibilities of modifying the wind field by the canopy of trees near noise barriers, in order to improve downwind shielding, is numerically investigated.

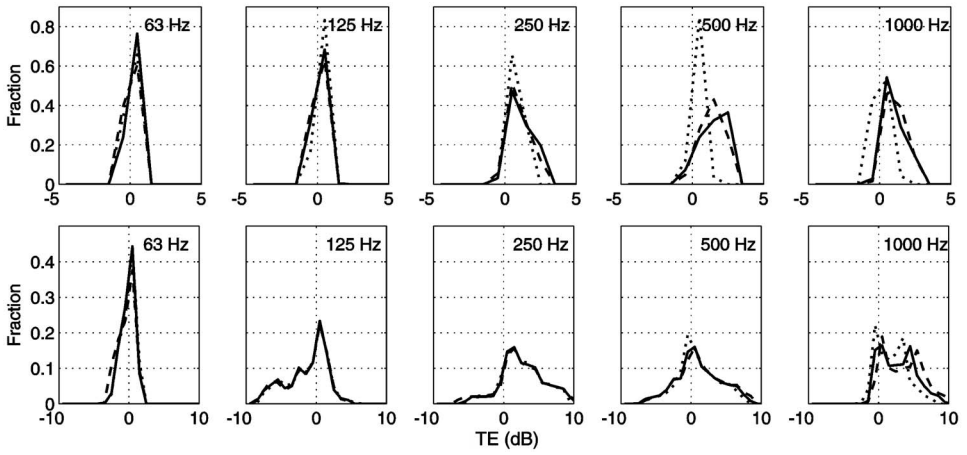


FIG. 8. Distribution of TE over decibel classes in the receiver zone, for a noise barrier height H_b which equals 4 m. The full lines represent uniform canopy, the dashed lines the triangular canopy and dotted lines the ellipse form. Friction velocities of 0.4 m/s (first row) and 0.8 m/s (second row) are considered. A grass-covered ground is assumed downwind. The average values of k_r equal 2.

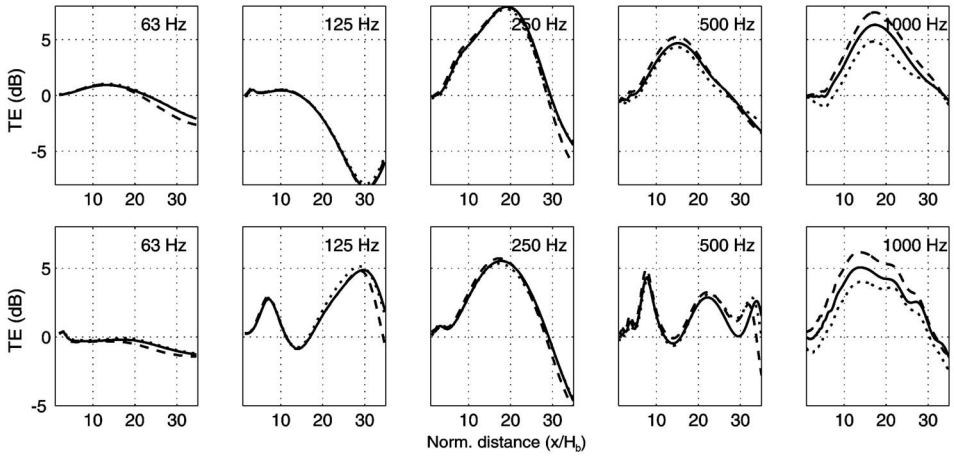


FIG. 9. TE along a horizontal line in the shielded area of the noise barrier, at a fixed height of 2 m (first row) and 4 m (second row). The noise barrier height H_b equals 4 m. The full lines represent uniform canopy, the dashed lines the triangular canopy and dotted lines the ellipse form. A friction velocity of 0.8 m/s is considered, together with an average value of k_r equal to 2. Grass-covered ground is assumed downwind from the noise barrier.

Common (or documented) quantitative tree properties are used to predict wind fields. These are the leaf area density, the canopy element drag coefficient, and the canopy width. The pressure loss coefficient is approximately equal to the product of these three quantities. Scattering on leaves and the effects of turbulence are not taken into account. This choice is justified by considering the rather low frequency interval of importance of shielded traffic noise.

In the configuration under study, a negative effect of wind on the downwind noise barrier shielding efficiency is observed, starting from the octave band with center frequency of 250 Hz. Below 1000 Hz, zones with increased shielding by the action of the wind are possible, because of a shift of the location where conditions for destructive interference are met, especially in case of a rigid ground downwind from the noise barrier. At 1000 Hz, only negative wind effects are found, and their magnitude exceed 20 dB for an incident wind speed profile with a friction velocity of 0.8 m/s.

The triangular crown shape, where the pressure drop is maximum at the bottom of the canopy and which decreases linearly towards the top, is the most interesting configuration. Analysis of the vertical gradients in the horizontal component of the wind speed yielded the smallest values for this configuration, both near the barrier top and the top of the canopy. The second best configuration in this numerical study is a canopy with uniform aerodynamic properties with height. For the low wind speed modeled (friction velocity of 0.4 m/s), both types behave similarly. The ellipse form has a somewhat worse performance, but still improves the downwind shielding significantly compared to a noise barrier without trees. The zone with increased shielding by the presence of trees is located mainly at distances between $10H_b$ and $30H_b$ downwind from the noise barrier (with H_b the noise barrier height). With increasing wind speed, the optimum pressure resistance coefficient increases. For the low wind speed used in this paper, a pressure resistance coefficient equal to 2 is sufficient in case of a uniform canopy. For the high wind speed, a value of 4 gave a significant improvement over a value of 2.

The largest positive and most consistent effects by the presence of the trees in wind are found for the octave band of 1000 Hz. This is the dominant frequency band when looking at noise near highways.

The numerical analysis in this paper leads to the conclusion that coniferous trees are more suited than deciduous trees to improve the wind field near noise barriers. Their typical needle area densities and canopy element drag coefficients lead to larger pressure resistance coefficients. Further, their canopy shape is usually close to the optimal triangular form, and during winter, there is no significant loss in biomass.

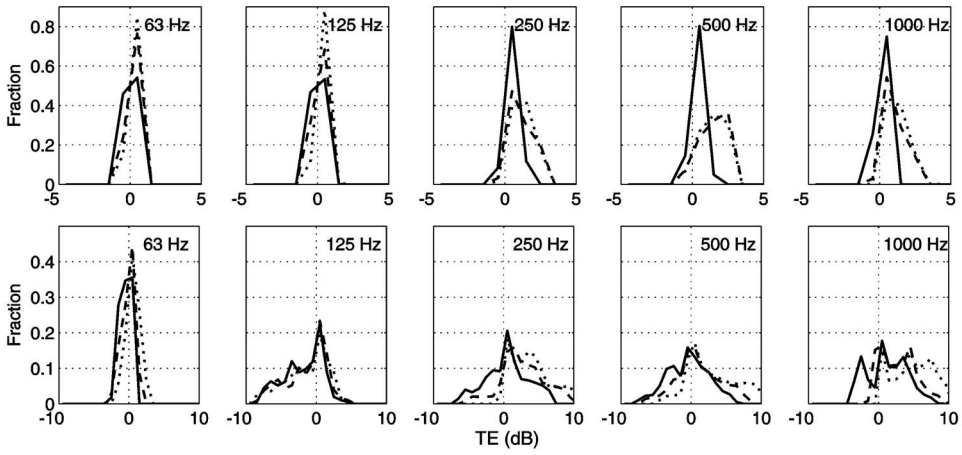


FIG. 10. Distribution of TE over decibel classes, for a friction velocity of 0.4 m/s (first row) and 0.8 m/s (second row). The noise barrier height H_b equals 4 m. The full lines are for $k_r=1$, the dashed lines for $k_r=2$, and the dotted lines for $k_r=4$. A grass-covered ground is assumed. A uniform crown shape is simulated.

[1]R. De Jong and E. Stusnick, "Scale model studies of the effect of wind on acoustic barrier performance," Noise Control Eng. **6**, 101–109 (1976).
[2]K. Rasmussen and M. Arranz, "The insertion loss of screens under the influence of wind," J. Acoust. Soc. Am. **104**, 2692–2698 (1998).
[3]E. Salomons, "Reduction of the performance of a noise screen due to screen-induced wind-speed gradients. Numerical computations and wind tunnel experiments," J. Acoust. Soc. Am. **105**, 2287–2293 (1999).
[4]T. Van Renterghem, D. Botteldooren, W. Cornelis, and D. Gabriels, "Reducing screen-induced refraction of noise barriers in wind by vegetative screens," Acust. Acta Acust. **88**, 231–238 (2002).
[5]T. Van Renterghem and D. Botteldooren, "Effect of a row of trees behind noise barriers in wind," Acust. Acta Acust. **88**, 869–878 (2002).
[6]T. Van Renterghem and D. Botteldooren, "Numerical simulation of the effect of trees on downwind noise barrier performance," Acust. Acta Acust. **89**, 764–778 (2003).
[7]T. Van Renterghem, "The finite-difference time-domain method for simulation of sound propagation in a moving medium," Ph.D. dissertation, University Gent, 2003.
[8]X. Zhou, J. Brandle, C. Mize, and E. Takle, "Three-dimensional aerodynamic structure of a tree shelterbelt: Definition, characterization and working models," Agroforest. Syst. **63**, 133–147 (2005).
[9]J. Donat and B. Ruck, "Simulated ground deposition of fine airborn particles in an array of idealized tree crowns," Boundary-Layer Meteorol. **93**, 469–492 (1999).
[10]FLUENT, *Computational Fluid Dynamics Software*, version 6.3, Fluent, Incorporated, Centerra Resource Park, Lebanon, NH.
[11]P. Richards and R. Hoxey, "Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model," J. Wind. Eng. Ind. Aerodyn. **46–47**, 145–153 (1993).
[12]B. Blocken, T. Stathopoulos, and J. Carmeliet, "CFD simulation of the atmospheric boundary layer—wall function problems," Atmos. Environ. **41**, 238–252 (2007).
[13]J. Wilson, "Numerical studies of flow through a windbreak," J. Wind. Eng. Ind. Aerodyn. **21**, 119–154 (1985).
[14]L. Hagen and E. Skidmore, "Windbreak drag as influenced by porosity," Trans. ASAE **14**, 464–465 (1971).
[15]H.-B. Su, R. Shaw, K. Pawu, C.-H. Moeng, and P. Sullivan, "Turbulent statistics of neutrally stratified flow within and above a sparse forest from large-eddy simulation and field observations," Boundary-Layer Meteorol. **88**, 363–397 (1998).
[16]Ü. Rannik, T. Markkanen, J. Raittila, P. Hari, and T. Vesala, "Turbulence statistics inside and over forest: Influence of footprint prediction," Boundary-Layer Meteorol. **109**, 163–189 (2003).
[17]M. Irvine, B. Gardiner, and M. Hill, "The evolution of turbulence across a forest edge," Boundary-Layer Meteorol. **84**, 467–496 (1997).
[18]J. Lopes da Costa, F. Castro, J. Palma, and P. Stuart, "Computer simulation of atmospheric flows over real forests for wind energy resource evaluation," J. Wind. Eng. Ind. Aerodyn. **94**, 603–620 (2006).
[19]A. Porté, A. Bosc, I. Champion, and D. Loustau, "Estimating the foliage area of Maritime pine (Pinus pinaster Aït.) branches and crowns with application to modelling the foliage area distribution in the crown," Ann. For. Sci. **57**, 73–86 (2000).
[20]Y. Wan, "Crown structure, radiation absorption, photosynthesis and transpiration," Ph.D. dissertation, University of Edinburgh, Edinburgh, U.K., 1988.
[21]T. Van Renterghem, E. Salomons, and D. Botteldooren, "Efficient FDTD-PE model for sound propagation in situations with complex obstacles and wind profiles," Acust. Acta Acust. **91**, 671–679 (2005).
[22]D. Duhamel, "Efficient calculation of the three-dimensional sound pressure field around a noise barrier," J. Sound Vib. **197**, 547–571 (1996).
[23]R. Blumrich and D. Heimann, "A linearized Eulerian sound propagation model for studies of complex meteorological effects," J. Acoust. Soc. Am. **112**, 446–455 (2002).
[24]V. Ostashev, D. Wilson, L. Liu, D. Aldridge, N. Symons, and D. Marlin, "Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation," J. Acoust. Soc. Am. **117**, 503–517 (2005).
[25]D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," J. Acoust. Soc. Am. **98**, 3302–3308 (1995).
[26]T. Van Renterghem and D. Botteldooren, "Prediction-step staggered-in-time FDTD: an efficient numerical scheme to solve the linearised equations of fluid dynamics in outdoor sound propagation," Appl. Acoust. **68**, 201–216 (2007).
[27]K. Gilbert and X. Di, "A fast Green's function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **94**, 2343–2352 (1993).
[28]E. Salomons, "Improved Green's function parabolic equation method for atmospheric sound propagation," J. Acoust. Soc. Am. **104**, 100–111 (1998).
[29]M. Delany and E. Bazley, "Acoustic properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[30]S. Tang and P. Ong, "A Monte Carlo technique to determine the effectiveness of roadside trees for containing traffic noise," Appl. Acoust. **23**, 263–271 (1988).
[31]J. Kragh, "Road traffic noise attenuation by belts of trees," J. Sound Vib. **74**, 235–241 (1981).
[32]M. Martens, "Foliage as a low-pass filter: Experiments with model forests in an anechoic chamber," J. Acoust. Soc. Am. **67**, 66–72 (1980).
[33]D. Heimann and R. Blumrich, "Time-domain simulations of sound propagation through screen-induced turbulence," Appl. Acoust. **65**, 561–582 (2004).
[34]U. Sandberg and J. Ejsmont, *Tyre/Road Noise Reference Book* (Informex, Kisa, Sweden, 2002).

The impact of water column variability on horizontal wave number estimation and mode based geoacoustic inversion results

Kyle M. Becker[a)]
The Pennsylvania State University, Applied Research Laboratory, State College, Pennsylvania 16804-0030

George V. Frisk
Department of Ocean Engineering, Florida Atlantic University, Dania Beach, Florida 33004

(Received 28 August 2007; accepted 13 November 2007)

The influence of water column variability on low-frequency, shallow water geoacoustic inversion results is considered. The data are estimates of modal eigenvalues obtained from measurements of a point source acoustic field using a horizontal aperture array in the water column. The inversion algorithm is based on perturbations to a background waveguide model with seabed properties consistent with the measured eigenvalues. Water column properties in the background model are assumed to be known, as would be obtained from conductivity, temperature, and depth measurements. The scope of this work in addressing the impact of water column variability on inversion is twofold. Range-dependent propagation effects as they pertain to eigenvalue estimation are first considered. It is shown that mode coupling is important even for weak internal waves and can enhance modal eigenvalue estimates. Second, the effect of the choice of background sound speed profile in the water column is considered for its impact on the estimated bottom acoustic properties. It is shown that a range-averaged sound velocity profile yields the best geoacoustic parameter estimates. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821976]

PACS number(s): 43.30.Bp, 43.20.Mv, 43.60.Pt [AIT] Pages: 658–666

I. INTRODUCTION

Although the influence of physical oceanographic processes on acoustic forward and inverse problems in deep water has been studied for decades,[1–4] it is only relatively recently that the role of water column variability has been explored in low-frequency, shallow-water acoustics.[5–7] In contrast to deep-water environments, where the effect of the bottom can be ignored in many applications, shallow-water problems have been considered historically to be ones in which the seabed plays the dominant role in influencing acoustic propagation in the ocean waveguide.[8,9] Shallow-water inversion methods have therefore focused on the determination of the bottom geoacoustic properties in three dimensions, assuming a benign water column with acoustic characteristics that are temporally stable and dependent only on depth.[10] In fact, this model of the water column is now considered to be an oversimplified one in many environments, and both temporally and spatially varying oceanographic features in three dimensions must be incorporated into the overall shallow-water acoustic picture.[11]

Recent work has examined the effects of internal waves on inversion for determining average water column properties using modal travel time estimates as input data.[12] In the present work, it is of particular interest to examine the implications of introducing realistic water column variability into the geoacoustic inversion process.[13–15] This paper focuses on two important aspects of this problem, assuming a perturbative inversion scheme that utilizes estimates of the modal eigenvalues as input data. These eigenvalues are obtained from horizontal array measurements of the acoustic field due to a continuous wave (cw) point source in the water column. First, the effects of range-dependent sound speed variation associated with weak internal wave fluctuations are considered. Second, the choice of the background sound speed profile in a variable ocean and its effect on the inverted geoacoustic properties is considered. The approach used here is based on a frozen ocean, or snapshot in time, of an otherwise dynamic environment. Although, methods for including the dynamics of the environment as they pertain to this work are being investigated,[16] they are beyond the scope of this initial study and will be addressed in future works.

This paper is organized as follows. Section II describes the theoretical framework within which the range-dependent water column variability will be studied. Section III discusses the impact of sound speed variability on the modal eigenvalue estimation process for the specific example of weak internal wave fluctuations. Section IV examines the choice of background model for the sound speed profile in the water column and describes its impact on the geoacoustic inversion results. Section V summarizes our results and conclusions.

II. THEORY

The inversion approach considered here is rooted in the Hankel transform pair relationship between the pressure field

[a)]Author to whom correspondence should be addressed. Electronic mail: kmbecker@psu.edu.

$p(r;z,z_s)$ due to a time harmonic point source ($e^{-i\omega t}$) and the depth-dependent Green's function[8]

$$p(r;z,z_s) = \int_0^\infty g(k_r;z,z_s)J_0(k_r r)k_r dk_r,$$

$$g(k_r;z,z_s) = \int_0^\infty p(r;z,z_s)J_0(k_r r)rdr, \quad (1)$$

where $J_0(k_r r)$ is the zero order Bessel function, k_r is the horizontal wave number, and r is the range between the source (at depth z_s) and receiver (at depth z). For range-independent waveguides, Eq. (1) is exact, and satisfies an inhomogeneous depth-dependent Helmholtz equation with impedance boundary conditions at the surface ($z=0$) and the bottom ($z=h$). The bottom boundary condition is a function of the seabed properties such that a geoacoustic inverse problem can be formulated based on estimates of the Green's function.[17,18] Using asymptotic approximations valid for $k_r r \gg 1$, the depth-dependent Green's function of Eq. (1) can be approximated by the horizontal wave number spectrum[8]

$$g(k_r;z,z_s) \sim \frac{e^{i\pi/4}}{\sqrt{2\pi k_r}} \int_{-\infty}^{\infty} p(r;z,z_s)\sqrt{r}e^{-ik_r r}dr. \quad (2)$$

Experimentally, the depth-dependent Green's function, in this work approximated by the horizontal wave number spectrum, is estimated from pressure data measured as a function of range on a single hydrophone at a constant depth.[19] In shallow water, the resulting spectrum is characterized by a finite number of distinct peaks that correspond to the eigenvalues of the propagating modes in the waveguide. These eigenvalues are then used as input data to a perturbative inversion algorithm to infer properties of the seabed.

The relationship between the depth-dependent Green's function and complex pressure field described in the previous paragraph is strictly valid for range-independent environments. For the range-dependent case, a convenient representation of the pressure field is given by the adiabatic mode sum expressed as[9]

$$p(r,z) \sim \frac{\sqrt{2\pi}e^{i\pi/4}}{\rho(0,z_s)} \sum_{m=1}^{M} \Psi_m(0,z_s)\Psi_m(r,z) \frac{e^{i\int_0^r k_{mr}(r')dr'}}{\sqrt{\int_0^r k_{mr}(r')dr'}}, \quad (3)$$

where the Ψ_m are the local mode functions at the source and receiver locations, k_{mr} are the modal eigenvalues, and ρ is the density at the source location. In this representation, propagating modes and their corresponding eigenvalues adapt to local waveguide properties and any changes that occur with range. A local estimate of the Green's function can be obtained by limiting the aperture of the integral in Eq. (2). The wave number spectrum at range r_i is found by integrating over the region $r_i - L/2 \leq r_i \leq r_i + L/2$, where L is the aperture length, over which the waveguide environment is assumed to be approximately range independent. For a truly adiabatic environment, moving the aperture along in range using short range steps gives a picture of the modal content of the waveguide as it evolves with range.[20] The adiabatic representation provides an intuitive way to understand how modal content evolves with range, but is not a necessary requirement for extracting range-dependent modal information using the Hankel transform. For the synthetic pressure fields generated for this study, the introduction of the internal wave field to the otherwise range-independent waveguide causes energy in the propagating modes to be transferred among each other violating the adiabatic assumption. However, estimates of the local depth-dependent Green's function can still be obtained by integrating the acoustic field over a finite aperture as just described.

III. WATER COLUMN EFFECTS ON WAVE NUMBER ESTIMATION

Wave number estimates were obtained for an environment with range and depth-dependent water column properties. The waveguide environmental model was chosen from a set of models designed for benchmarking shallow water range-dependent acoustic propagation codes.[21] In addition to being vetted by the modeling community, the simplified deterministic internal wave model provided features consistent with measurements of temperature variation in shallow water as measured on thermistor strings[11] or by other means. The chosen model yielded an environment with a modulated displacement of the thermocline, a feature consistent with other simplified models based on solutions to the Korteweg-deVries equation and used in studies of shallow water acoustic propagation.[5,7] Further, the environment chosen was straightforward to incorporate into the chosen acoustic models. As used here, the model yields a single realization, or snapshot, representative of a dynamic ocean environment. The environment was specified with range-independent fluid sediment properties and a constant water/sediment interface depth of 200 m. The seabed had a compressional wave speed of 1700 m/s, a density of 1.5 g/cm^3, and an attenuation of 0.1 dB/λ. Density in the water column was 1 g/cm^3 with no attenuation. The background sound speed profile in the water column was generally downward refracting with a slight duct occurring above 26 m depth and was given by[22]

$$c(z) = 1515 + 0.016z, \quad z < 26 \text{ m}$$

$$c(z) = c_o[1 + a(e^{-b} + b - 1)], \quad z \geq 26 \text{ m}, \quad (4)$$

where z is the depth in meters, $c_o = 1490$ m/s, $a = 0.25$, and $b = (z-200)/500$. Range dependence of the environment was achieved by perturbing the background sound speed profile in the water column as a function of range and depth to simulate an internal wave field. The sound speed perturbation, δc, was given by[22]

$$\delta c(z,r) = 4(z/B)e^{-z/B} \sum_{i=1}^{5} cos(K_i r), \quad (5)$$

with $K_i = 2\pi[2000 - 300(i-1)]^{-1}$ m^{-1} and $B = 25$ m. The sound speed field in range and depth for this model is shown in Fig. 1. The maximum sound speed perturbation was about 7.5 m/s. For this waveguide environment, 100 Hz acoustic

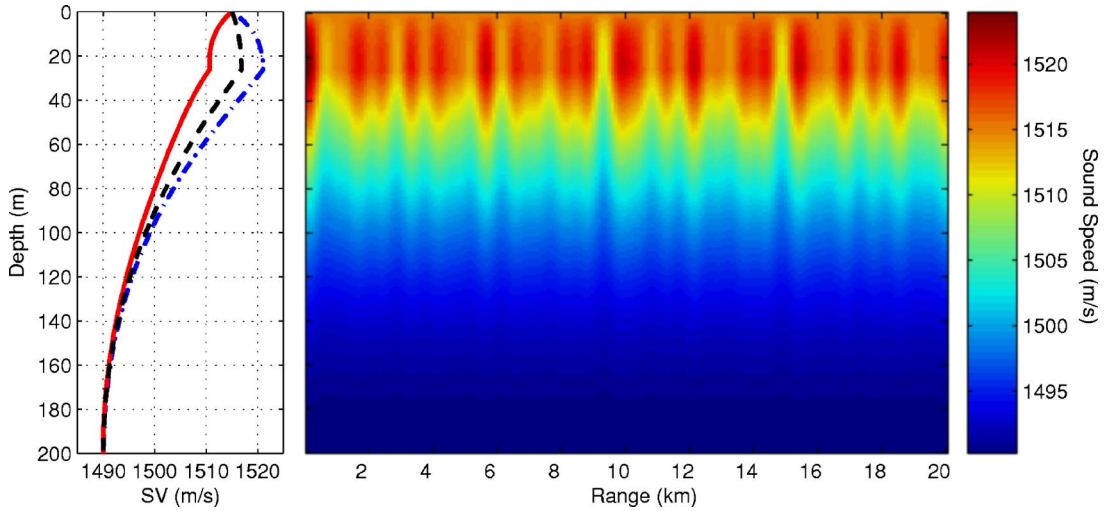


FIG. 1. (Color online) Sound speed field in the water column used for range-dependent acoustic propagation calculations. The background profile of Eq. (4), along with the minimum and maximum variations of all the perturbed profiles, is shown in the left panel.

field data were synthesized for a source depth of 45 m. The field was computed out to 20 km over the entire 200 m depth of the water column. The depth and range grids were specified at 1 and 5 m, respectively. Full-wave synthetic pressure fields were generated using both the parabolic equation (PE) code RAM,[23] and the coupled mode code COUPLE.[24,25] The PE code was used to generate range-dependent pressure field data, from which modal eigenvalues were estimated, while COUPLE was used to generate range-dependent modal eigenvalues and amplitudes as benchmark data to compare with the estimated values. Using the wave numbers obtained from the PE model provided confidence that the estimates were independent of the choice of propagation model. In COUPLE, the waveguide environment was divided into 200, 100-m-long, range-independent segments. Modes and modal eigenvalues were calculated for each of the segments and the full wave field determined for a cw point source at $(0, z_s)$ and matching boundary conditions at the vertical interfaces.[25] Acoustic field data with the same source characteristics were also synthesized on the same depth/range grid for a range-independent environment using the background sound speed profile in the water column. The acoustic intensity for the range-independent field data subtracted from the data synthesized including internal waves are shown in Figs. 2 and 3. Figure 2 shows the intensity differences over all depths between 0 and 5 km from the source, and Fig. 3 shows the differences between 15 and 20 km from the source. Absolute differences in intensity between the two fields can be greater than 25 dB and occur where there is a mismatch between the null depths or locations in the modal interference patterns of the two fields. The impact of propagation through the internal wave field is greater at longer ranges from the source as seen in Fig. 3. Nevertheless, the pressure field synthesized including the internal wave field still exhibits a high degree of coherence with range as evident in Fig. 4 which shows the pressure field magnitude at a fixed depth of 60 m. In particular, the figure indicates a large degree of correlation for the peak and null locations between the fields calculated for range-independent and range-dependent environments. The consistency of modal cycle distances for the two different propagation environments indicates that differences between adjacent modal eigenvalues are preserved.

Using the synthesized acoustic pressure fields, horizontal wave number spectra were estimated from horizontal aperture data at 60 m depth. Range-dependent estimates were obtained from the complex pressure data using a sliding window autoregressive (AR) spectral estimator.[26] Subapertures of 3 km were used and stepped every 200 m to span the entire 20 km. Model order for the AR estimator was chosen as $1/3$ the number of data points in each subaperture. Note that the subaperture length L is much larger than the range segmentation of the waveguide, so that multiple range-independent regions are integrated over to obtain spectra. This integration process results in broader spectral peaks and smoothed wave number estimates with range. The horizontal wave number spectrum estimates as a function of range with-

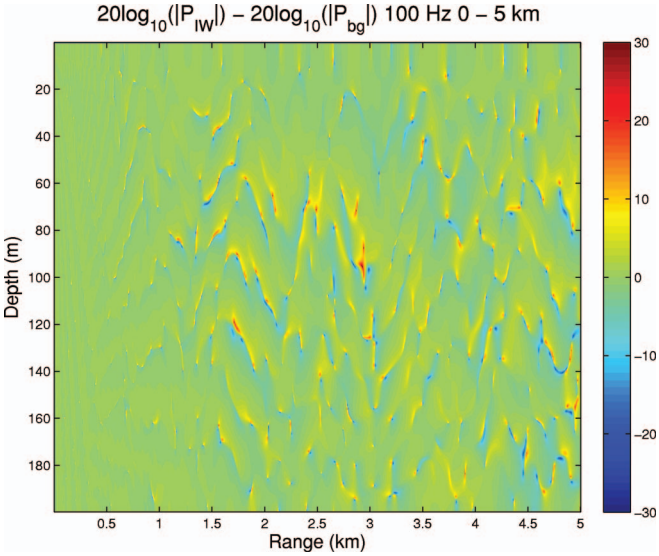


FIG. 2. Difference in acoustic field intensity with and without internal waves between 0 and 5 km. Source frequency is 100 Hz and source depth is 45 m. Color scale is in dB.

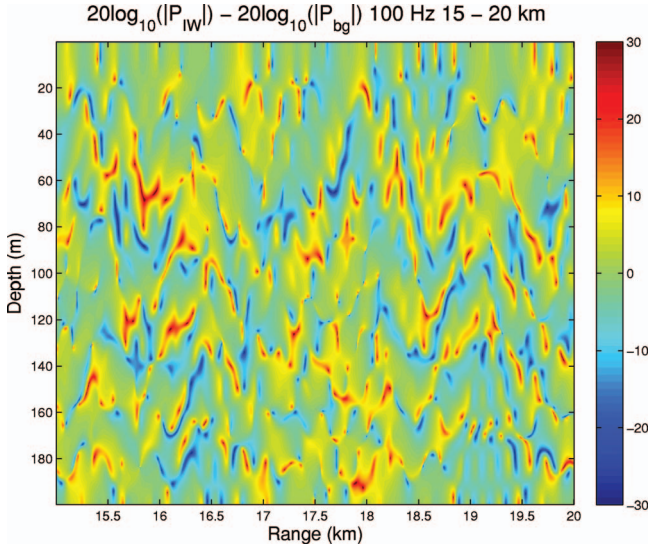


FIG. 3. Difference in acoustic field intensity with and without internal waves between 15 and 20 km. Source frequency is 100 Hz and source depth is 45 m. Color scale is in dB.

out and including the effects of internal waves are shown in Figs. 5 and 6, respectively. The figures present spectra normalized by the largest spectral peak amplitude estimated at every range step. Spectral levels greater than −80 dB are shown in the plots as black with strong spectral lines indicating horizontal wave numbers of propagating modes. For the background model, or unperturbed case, Fig. 5 indicates the presence of six strong modes for the given source/receiver depths. For the internal wave case, Fig. 6, eight strong modes are indicated, six of which are common to the unperturbed case. To compare with the spectral estimation results, modes for the unperturbed case, or background sound speed profile, were calculated using COUPLE. The modal eigenvalues determined from COUPLE are plotted as dashed lines overlaid on the estimated spectral lines. There are 13 real modal eigenvalues predicted for the range-independent case, whereas Figs. 5 and 6 indicate that the number of strong spectral lines is less than 13 in either case. This result is not unexpected as source and/or receiver depths can play a large role in modal excitation amplitudes with low amplitudes occurring for source/receiver depths close to the mode nulls. However, although not all modes were estimated from the data, the strong spectral lines in the range/wave number plots do show good agreement with values calculated using COUPLE. Further, and most interesting, when including internal waves in the propagation environment, two additional strong spectral lines are observed. These two lines and the original six all show good agreement with the values calculated for the background model. The implication of these results is that for this example of weak internal waves in the water column, estimated modal eigenvalues are consistent with the waveguide boundary conditions and suitable for use in an inversion algorithm to recover seabed properties. This observation is consistent with the results of Preisig and Duda when examining modal propagation through internal solitary waves.[5] They concluded that bottom variability is a secondary effect relative to water column variability with regard to modal excitation amplitudes and coupling. Further, they showed that after passing through a solitary wave event, the modes become uncoupled, satisfying the original boundary conditions and medium properties.

To gain further understanding of the previous results, modal amplitudes calculated using COUPLE at the receiver depth were examined as a function of range for the range-dependent and range-independent waveguides. As stated previously, modal eigenvalues had to be calculated independently for each range-independent segment of the

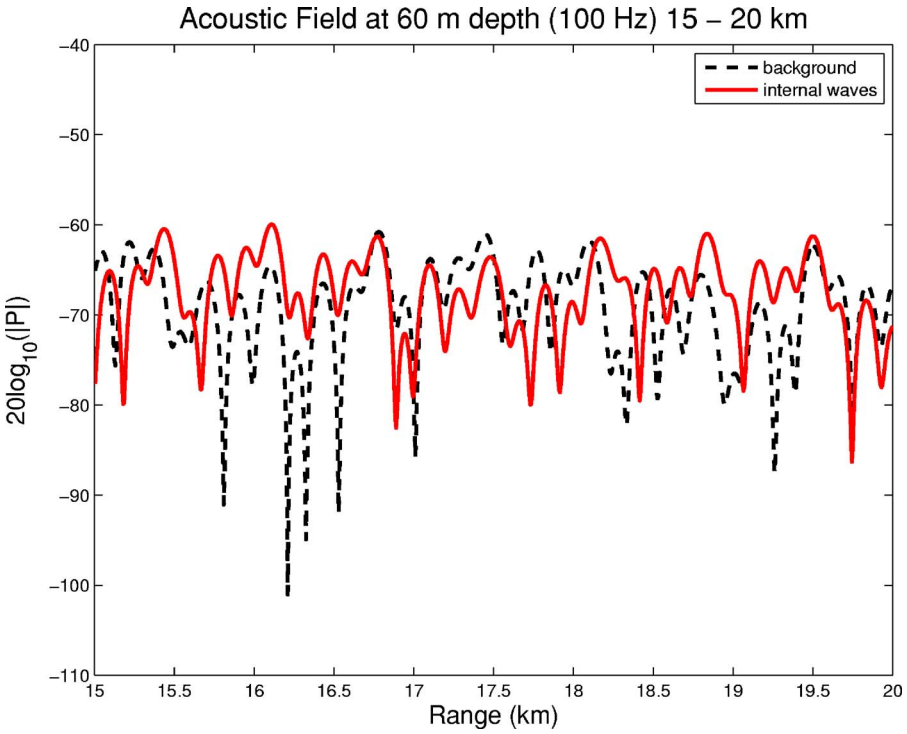


FIG. 4. (Color online) Magnitude squared of acoustic pressure with and without internal waves 10–15 km from the source at measured at 60 m. Source frequency is 100 Hz and source depth is 45 m.

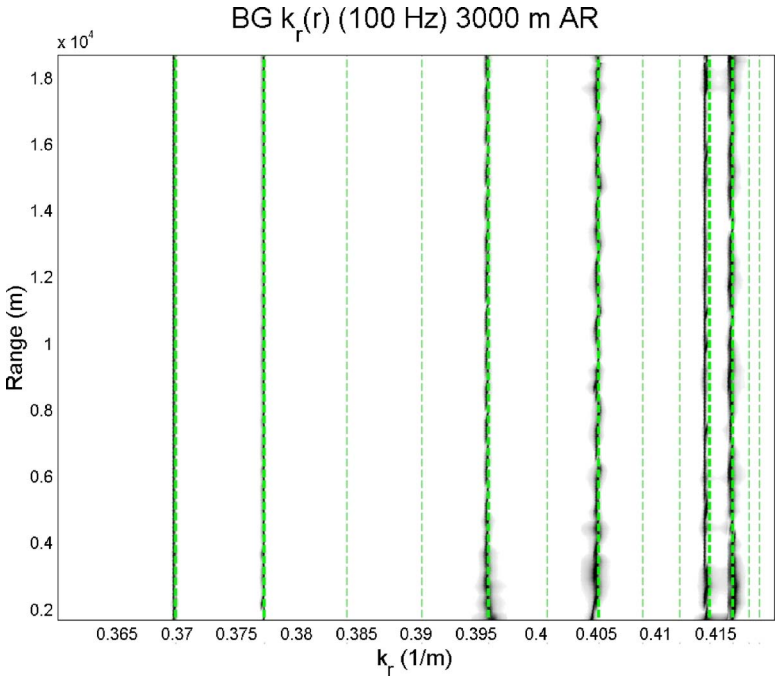


FIG. 5. (Color online) Wave number estimation with range with no internal waves. Dark lines are spectral estimation results, dashed lines are wave numbers returned from normal mode program for range-independent background sound speed profile.

waveguide. Modal amplitudes as a function of range were determined by solving the coupling matrices at the range interfaces.[9] For the range-independent case, coupling does not occur and individual mode amplitudes fall off with range according to cylindrical spreading. The cylindrical spreading behavior of the first ten modes of the range-independent case is shown by the solid line in Fig. 7. The dashed lines are the modal amplitudes calculated when including the internal wave field. The thin line on the plot at −80 dB represents the threshold above which the strong spectral lines in the range/wave number plots were observed. Higher modal amplitudes relative to this line indicate higher signal level, and greater contribution to the total field, for that particular mode. From Fig. 7, much more variability is observed in the amplitudes of the low signal level modes. Consistent with the spectral estimates, the modal amplitude calculations reveal amplitudes of modes 6 and 10 for the internal wave case to be much greater than for the background case. The additional energy observed in modes 6 and 10 indicates the coupling of energy from other modes into these modes and is responsible for the strong spectral lines observed for these modes in Fig. 6. From these results, and at least for this particular example, the selective amplification of certain modes at the receiver depth due to the sound speed variability in the water column allowed for their determination.

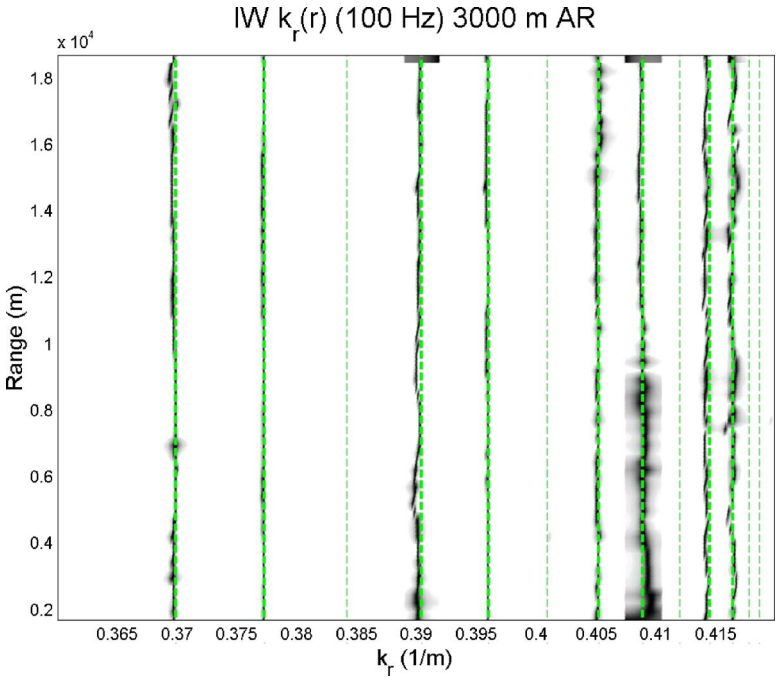


FIG. 6. (Color online) Wave number estimation with range including internal waves. Dark lines are spectral estimation results, dashed lines are wave numbers returned from normal mode program for range-independent background sound speed profile.

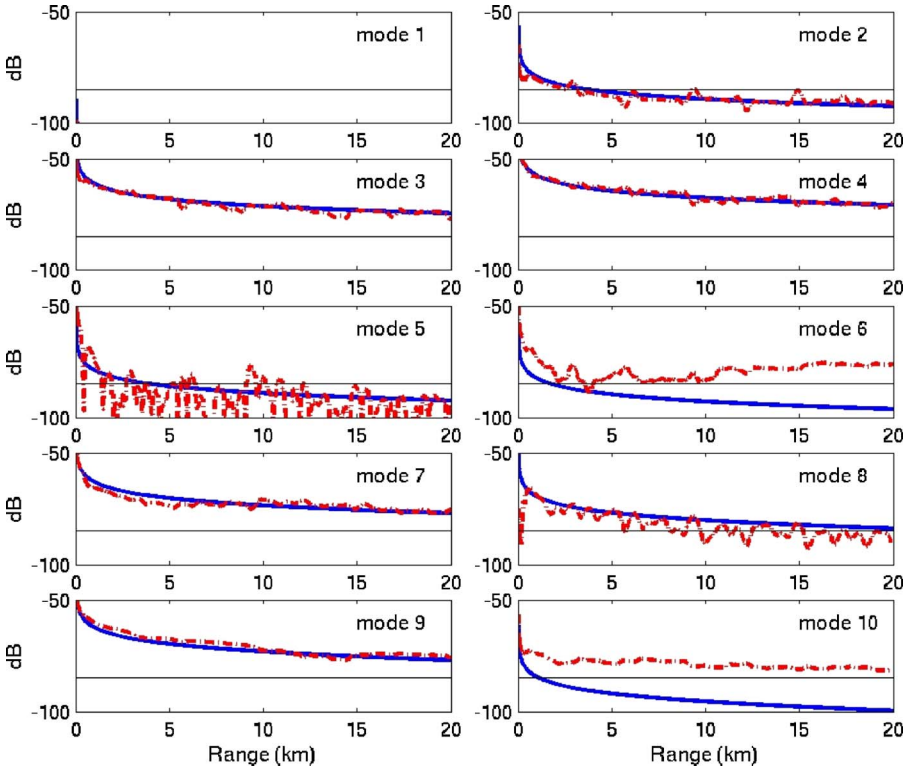


FIG. 7. (Color online) Amplitudes of the first 10 modes ($20\log_{10}|\Psi_m(r,z)|$) as a function of range without and including internal waves in the water column. Solid line is for range-independent waveguide. Thin line at −80 dB is threshold above which strong spectral lines in the range/wave number plots were observed.

IV. INFLUENCE OF WATER COLUMN ON INVERSION RESULTS

The preceding section illustrated some effects of internal waves on estimating modal eigenvalues for use as input data to a perturbative inversion algorithm. In this section, the background sound speed profile for the water column used in this inversion method is examined for its role in affecting the geoacoustic parameter estimates. The inverse method used in this work is based on a linearized relationship between sound speed and modal eigenvalues. The relationship is defined by an integral equation, which is solved over all depths in the domain of the problem, given by[18]

$$\Delta k_m = \tilde{k}_{rm} - \mathring{k}_{rm} = \int_0^\infty \mathring{\rho}^{-1} |\mathring{\Psi}_m(z)|^2 \mathring{k}^2(z) \frac{\Delta c(z)}{\mathring{c}(z)} dz. \tag{6}$$

where the accent $\mathring{}$ indicates properties, wave numbers, mode functions and eigenvalues of a background, or starting, environmental model to be perturbed. Measurement data for the inversion are Δk_m obtained by taking the difference between the measured modal eigenvalues, $\tilde{k}_{rm}$, and those calculated for the starting model $\mathring{k}_{rm}$. The solution to the integral equation follows by a discrete parametrization of the waveguide model in depth resulting in a matrix equation.[18] The inversion algorithm proceeds by assuming a starting environmental model for the environment from which background mode functions and eigenvalues are calculated using a mode code. The starting model is then perturbed via the inversion algorithm, yielding an updated model. The process is iterative and stopped when the modal eigenvalues of the updated model suitably match the measured eigenvalues, or when Δk_{rm} is minimized. In order of magnitude, the number of iterations is typically ten. Further details of the inversion algorithm are given in the literature.[18,27]

With the inverse problem defined, this section seeks to address the effects of incomplete knowledge or incorrect assumptions about the sound speed profile in the water column on the geoacoustic inversion results for the seabed using modal eigenvalues as input data. Water column properties are often assumed to be known based on measurements obtained from conductivity, temperature, and depth (CTD) data collected during acoustic measurements. The known water column properties are used in the starting environmental model for calculating background mode functions. However, often the water column data collected during an experiment does not fully characterize the propagation environment because it is either spatially or temporally undersampled relative to the acoustic data.

To understand the effect of an incorrectly selected background profile, the inversion approach was applied independently to a number of different range-independent waveguide realizations. The measurement modal eigenvalues, $\tilde{k}_{rm}$, used in the inversion algorithm for each independent run were the same, and determined by the mean values of the wave numbers returned from COUPLE for all the range-independent waveguide realizations. Water column sound speed profiles used in the different realizations were based on the internal wave field described by Eqs. (4) and (5) and illustrated in Fig. 1. The mean values were determined for 200 realizations obtained by incrementing the range every 100 m. All other properties of the waveguide, including bottom geoacoustic properties, were equal to those described in Sec. III. The intent of averaging eigenvalues this way was to provide a tractable way to include the effect of integration inherent in modal eigenvalue estimation for the input data. For a wave-

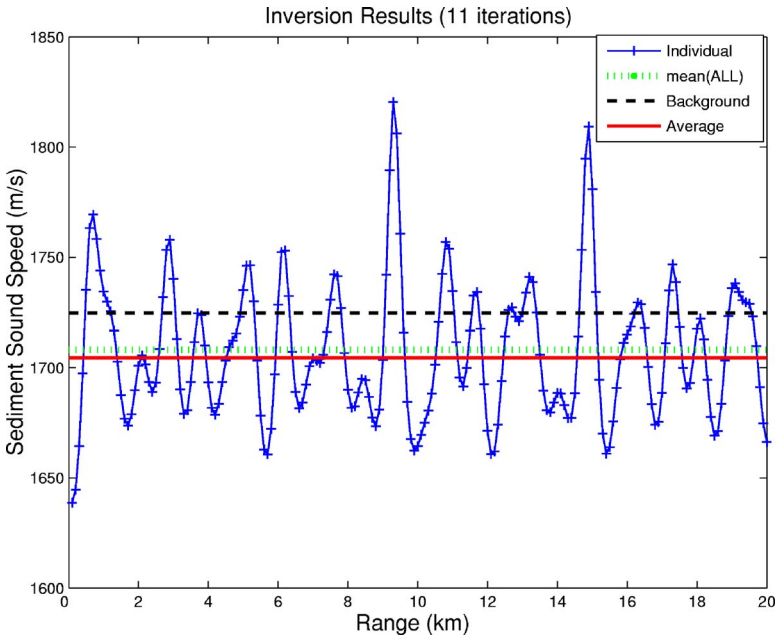


FIG. 8. (Color online) Range-dependent inversion results for bottom sound speed. Plus signs indicate results obtained using different background sound speed profiles at each range step, the lighter vertical dashed line is the mean sound speed value of the range varying results, the dashed line is the inversion result using the background sound speed profile, and the solid line is the result obtained using a range-averaged sound speed profile in the starting model.

guide with range-independent sediment properties, the averaging approach is consistent with the modal eigenvalue estimates that would be obtained from range/wave number displays such as in Figs. 5 and 6.

The starting models and input data, Δk_{rm}, used in the parametrized integral equation for each inversion calculation, were determined using different realizations of the water column sound speed profile in the waveguide. The sound speed profile realizations were those described in the previous paragraph. Each waveguide realization was assumed to be range independent with background mode functions and modal eigenvalues determined using a frequency of 50 Hz. In the previous section, 100 Hz was used in the examples to demonstrate modal eigenvalue estimation, however, for the given source and receiver locations, not all modes were detected. At 50 Hz, the number of modal eigenvalues estimated from synthetic full-wave data was consistent with the number predicted. Further, for a 3-km (strike long) aperture, and noise free synthetic data, the mean square error between predicted and estimated eigenvalues would be less than 10^{-7} m^{-1}.[26] Consequently, full-wave pressure data were not calculated for each waveguide realization, and background modal eigenvalue predictions, determined using COUPLE, were used directly in the inversion algorithm based on Eq. (6).[27] The inversion algorithm was devised to output a single value representing the sound speed in a sediment half space for comparison with the sediment sound speed of 1700 m/s in the waveguide model.

The number of independent inversions performed was 200, equal to the number of range-independent segments making up the range-dependent environment. With a starting model based on the local sound speed profile, any particular estimate of geoacoustic sediment properties would be equivalent to that achieved by making a CTD cast at that particular range and assuming it represented the entire waveguide. Individual geoacoustic parameter estimates obtained using the different sound speed profile data were compared to inversion results obtained using a starting model comprised of the background sound speed profile given by Eq. (4). The individual estimates were also compared to results obtained using an average sound speed profile obtained by taking the mean sound speed value at each depth for the 200 waveguide realizations. This approach is equivalent to using a range-averaged sound speed profile, representing the full 20 km waveguide, in the starting model. Additionally, the mean value of the individual parameter estimates is determined for comparison with the other results. Inversion results as a function of range are shown in Fig. 8. The closest result to the true value of 1700 m/s was achieved using a range-averaged sound velocity profile in the background starting model. The variation in results obtained for the individual sound speed profiles at each range indicates the types of biases that could be expected by using a single sound speed profile to characterize the environment. Depending on the sound speed profile used in the starting model, the results can vary by as much as 150 m/s. The result obtained using the range-averaged profile was within 5 m/s of the expected value. The nature of the variability in the range-dependent case is illustrated by Fig. 9. At each range step the modal problem for a given sound speed profile is non-linear and is solved numerically to determine the mode functions and eigenvalues. The resulting information used in the linear inverse problem with the input data Δk_m is biased by the choice of starting background model. Although the kernel, based on the background modal functions is also variable, this is a secondary effect. The relationship between the sound speed deviation and the eigenvalues of mode 1 is shown in the figure and highly correlated. The sound speed deviation is shown at both mid-water column depth and near the bottom. As there is little variation in sound speed near the bottom, the variability in the eigenvalues is due to variability at mid-depths.

V. CONCLUSIONS

The impact of water column variability on modal inversion results was examined. For a weak internal wave field,

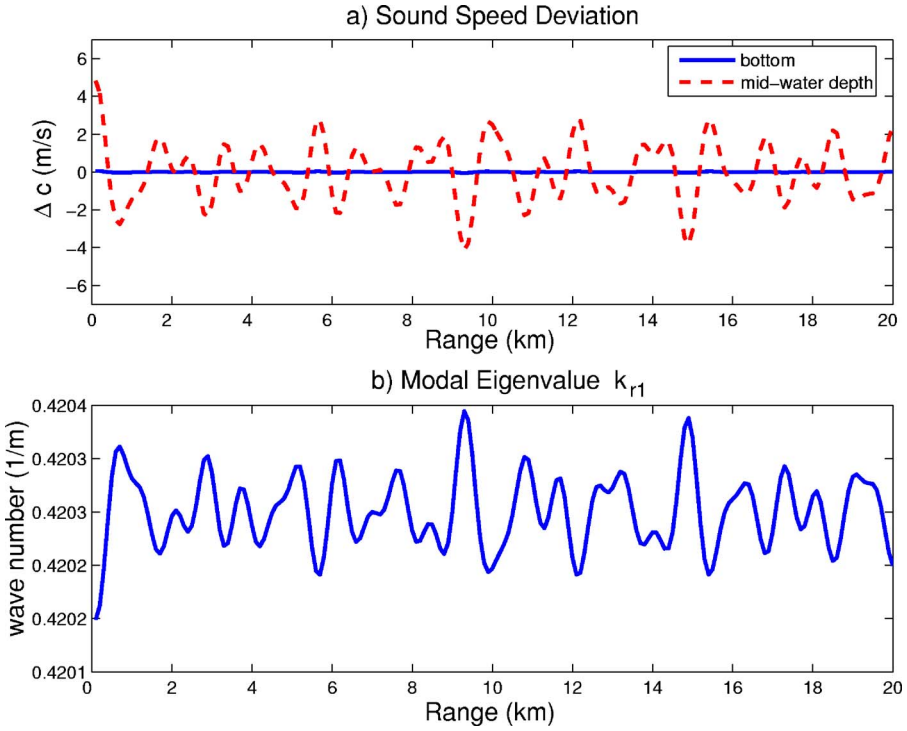


FIG. 9. (Color online) Comparison of sound speed variations in water column with horizontal wave number for first mode. The wave numbers track the sound speed variations. (a) Sound speed difference over range at 55 m (dashed) and 100 m (solid) depths. (b) Modal eigenvalues for mode 1 as a function of range.

mode coupling was observed, but spectral wave number estimates were consistent with modal eigenvalues satisfying the boundary conditions for the underlying range-independent sediment. Further, mode coupling enhanced the ability to estimate higher order modes. A conclusion of this study is that eigenvalue estimation for use in modal inversion methods to determine bottom properties is robust in the presence of weak internal wave fields. The integration process inherent in wave number spectral estimation over a given aperture in horizontal range suppresses much of the variation caused by the water column fluctuations, particularly for the higher order modes. As higher order modes are known to be more sensitive to changes in bottom properties, the results shown here suggest a way to discriminate between effects of water column variability and seabed variability on shallow water acoustic measurements. The inversion results for the range-dependent waveguide yielded biased results that depended on the water column sound speed profile used in the background model. The bias was removed by using a range-averaged sound speed profile. In conclusion, for range-variable water column properties the best inversion results would be obtained by measuring the spatial sound velocity field and taking the range-averaged depth profile. Although not trivial, measurement techniques for doing so are available including towed CTD chains[28] and instrumented tow cables.[29]

ACKNOWLEDGMENTS

Special thanks to Dr. Richard B. Evans of SAIC for his help with the COUPLE normal mode propagation model. The work was supported by an Office of Naval Research Special Research Awards in Underwater Acoustics Entry-Level Faculty Award, Grant No. N00014-04-1-0248 and also ONR Grant No. N00014-04-1-0296.

[1]S. M. Flatte, R. Dashen, W. H. Munk, K. M. Watson, and R. Zachariasen, *Sound Transmisson Through a Fluctuating Ocean*, edited by S. M. Flatte (Cambridge University Press, Cambridge, 1979).
[2]S. M. Flatte and G. Rovner, "Calculations of internal-wave-induced fluctuations in ocean-acoustic propagation," J. Acoust. Soc. Am. **108**(2), 526–534 (2000).
[3]W. Munk and F. Zachariasen, "Sound propagation through a fluctuating stratified ocean: Theory and observation," J. Acoust. Soc. Am. **59**(4), 264–291 (1976).
[4]A. R. Robinson and D. Lee, *Oceanography and Acoustics: Prediction and Propagation Models* (AIP Press, Woodbury, NY, 1994).
[5]J. C. Preisig and T. F. Duda, "Coupled acoustic mode propagation through continental-shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269 (1997).
[6]C. J. Higham and C. T. Tindle, "Coupled perturbed modes and internal solitary waves," J. Acoust. Soc. Am. **113**(5), 2515–2522 (2003).
[7]T. F. Duda, "Acoustic mode coupling by nonlinear internal wave packets in a shelfbreak front area," IEEE J. Ocean. Eng. **29**(1), 118–125 (2004).
[8]G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (PTR Prentice–Hall, Englewood Cliffs, NJ, 1994).
[9]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994).
[10]N. R. Chapman, S. Chin-Bing, D. King, and R. B. Evans, "Benchmarking geoacoustic inversion methods for range-dependent waveguides," IEEE J. Ocean. Eng. **28**, 320–330 (2003).
[11]J. R. Apel, M. Badiey, C. S. Chiu, S. Finette, R. Headrick, J. Kemp, J. F. Lynch, A. Newhall, M. H. Orr, B. H. Pasewark, D. Tielbuerger, A. Turgut, K. von der Heydt, and S. Wolf, "An overview of the SWARM shallow water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**(3), 465–500 (1997).
[12]G. Bouchage and M. I. Taroudakis, "Fluctuations of the modal arrival times due to linear internal waves: Application to inversion," J. Comp. Acs. **14**(4), 469–487 (2006).
[13]K. M. Becker and G. V. Frisk, "Effects of sound speed fluctuations due to internal waves in shallow water on horizontal wave number estimates," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, Dordrecht, 2002).
[14]R. Field, J. Newcomb, J. Showalter, J. George, and Z. Hallock, "Acoustic fluctuations and their harmonic structure," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, Dordrecht, 2002).
[15]Y. T. Lin, "An equivalent transform method for evaluating the effect of water-column mismatch on geoacoustic inversion," IEEE J. Ocean. Eng.

31(2), 284–298 (2006).
[16]K. M. Becker, "Modeling sound propagation in shallow water including dynamic internal wave fields," J. Acoust. Soc. Am. **119**(5), Part 2, 3226–3227 (2006).
[17]G. V. Frisk and J. F. Lynch, "Shallow water waveguide characterization using the Hankel transform," J. Acoust. Soc. Am. **76**(1), 205–216 (1984).
[18]S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic properties in shallow water," J. Acoust. Soc. Am. **82**(3), 998–1017 (1987).
[19]G. V. Frisk, J. F. Lynch, and S. D. Rajan, "Determination of compressional wave speed profiles using modal inverse techniques in a range-dependent environment in Nantucket Sound," J. Acoust. Soc. Am. **86**(5), 1928–1939 (1989).
[20]K. Ohta and G. V. Frisk, "Model evolution and inversion for seabed geoacoustic properties in weakly range-dependent shallow-water waveguides," IEEE J. Ocean. Eng. **22**, 501–521 (1997).
[21]Various authors, "Acoustical oceanography and underwater acoustics: Benchmarking range dependent numerical models," chaired by K. B. Smith and A. I. Tolstoy, J. Acoust. Soc. Am. **109**, 2332–2335 (2001).
[22]K. B. Smith, "Benchmarking shallow water range-dependent acoustic propagation modeling," Test Case III: Internal Waves. (December 20, 2000), http://web.nps.navy.mil/~kbsmith/ChicagoASA/iws.html (last accessed August 27, 2007).
[23]M. D. Collins, "A split-step Pade solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[24]R. B. Evans, "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," J. Acoust. Soc. Am. **74**, 188–195 (1983).
[25]R. B. Evans, COUPLE, 1997 Version (November 20 1997). http://oalib.hlsresearch.com/Modes/couple/ (last accessed August 27, 2007).
[26]K. M. Becker and G. V. Frisk, "Evaluation of an autoregressive spectral estimator for modal wave number estimation in range-dependent shallow water waveguides," J. Acoust. Soc. Am. **120**, 1423–1434 (2006).
[27]K. M. Becker, S. D. Rajan, and G. V. Frisk, "Results from the geoacoustic inversion techniques workshop using modal inverse methods," IEEE J. Ocean. Eng. **28**, 331–341 (2003).
[28]F. S. Henyey, K. Williams, K. M. Becker, J. Lyons, M. S. Ballard, H. J. Camin, P. Gabel, I. Koszalka, and J. Beitel, "Internal wave measurements with a towed CTD chain on the New Jersey Shelf," EOS Trans. Am. Geophys. Union **87**(52), Fall Meeting Suppl., Abstract No. OS23C-02 (2006).
[29]A. A. Ruffa and M. T. Sundvik, "Instrumented tow cable measurements of temperature variability in the water column," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, Dordrecht, 2002).

Geoacoustic inversion by mode amplitude perturbation

Travis L. Poole[a)]
Woods Hole Oceanographic Institution, Gigelow 209, MS#12, Woods Hole, Massachusetts 02543

George V. Frisk[b)]
Department of Ocean Engineering, Florida Atlantic University, SeaTech Campus, 101 North Beach Raod, Dania Beach, Florida 33004

James F. Lynch[c)]
Woods Hole Oceanographic Institution, Bigelow 209, MS#12, Woods Hole, Massachusetts 02543

Allan D. Pierce[d)]
Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215

(Received 13 August 2007; accepted 13 November 2007)

This paper introduces a perturbative inversion algorithm for determining sea floor acoustic properties, which uses modal amplitudes as input data. Perturbative inverse methods have been used in the past to estimate bottom acoustic properties in sediments, but up to this point these methods have used only the modal eigenvalues as input data. As with previous perturbative inversion methods, the one developed in this paper solves the nonlinear inverse problem using a series of approximate, linear steps. Examples of the method applied to synthetic and experimental data are provided to demonstrate the method's feasibility. Finally, it is shown that modal eigenvalue *and* amplitude perturbation can be combined into a single inversion algorithm that uses all of the potentially available modal data. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821975]

PACS number(s): 43.30.Bp, 43.30.Ma [AIT] Pages: 667–678

I. INTRODUCTION

In 1987 Rajan *et al.*[1] described a perturbative inversion method for obtaining the geoacoustic properties of the seabed from measurements of modal eigenvalues. Since the bottom influences the modal amplitudes as well, it seems natural to ask if the modal amplitudes can also be used to estimate the bottom properties. Furthermore, perhaps both the modal eigenvalue and amplitude data can be used together in a single inversion algorithm to achieve an improved result. This paper addresses both of these issues. Its main objective is to introduce a perturbative inversion scheme analogous to that in Ref. 1 using modal amplitudes, instead of modal eigenvalues, as the input data. We then show that modal eigenvalue *and* amplitude perturbation can be combined into a single algorithm, thereby combining the advantages of both methods.

The method assumes that measurements of the acoustic pressure field due to a cw point source have been made at one or more low frequencies ($<$500 Hz) in the water column. The quantity of primary interest is the compressional wave speed as a function of depth, but the density and attenuation profiles can also be sought, and we derive the necessary equations for determining them as well. The effects of shear waves are ignored in this development, but they could be included in a more general formalism.

Unlike Rajan *et al.*,[1] who made explicit use of the Hankel transform method for estimating the modal eigenvalues, we intentionally avoid a detailed discussion of the manner in which the data are collected and processed to obtain the required modal information. Though this is a critical step along the path to determining the bottom parameters, it is a separate issue from the modal amplitude inversion itself. We therefore assume knowledge of the modal amplitudes (preferably with an estimate of the uncertainty in those data) as well as information about the water column sound speed profiles at the locations where the bottom properties are to be determined. Whether these estimates are the output of a Hankel transform of horizontal line array (HLA) data collected over range as in Ref. 1, or from filtering the modes on a vertical line array (VLA) over depth as suggested by Ref. 2, or by some other method (for example, Ref. 3), the methodology of the modal amplitude inversion will remain the same. Thus, we avoid in-depth discussion of how the data are collected, and focus instead on the subsequent use of the data to estimate the bottom properties. Our preferred method of obtaining the modal amplitude data from measurements of the point source acoustic field will be the subject of a future publication.

Perturbative inversion methods approach the solution of the nonlinear problem of determining the bottom parameters from measurements of the acoustic field via a series of small, linear steps. During the first step of the algorithm a user-provided background model is used to calculate the parameters of the acoustic field (such as the eigenvalues or modal

[a)]Present address: Grant Institute of Earth Science, University of Edinburgh, Kings Buildings, Edinburgh, EH9 3JW, United Kingdom. Electronic mail: travis.poole@ed.ac.uk
[b)]Electronic mail: gfrisk@seatech.fau.edu
[c)]Electronic mail: jlynch@whoi.edu
[d)]Electronic mail: adp@bu.edu

amplitudes) using a forward model. Using this background model, the first-order derivatives of the acoustic parameters with respect to the bottom parameters are also computed. Next, the computed acoustic parameters are compared to the measured values, and we compute the change to the background model needed to correct (in a least-squares sense) our estimates based on an assumption of linearity. Since this linear assumption is not strictly correct, our correction will not give us the true bottom properties. If the background model is sufficiently similar to the true bottom, however, the correction will give us an improved estimate of the geoacoustic properties which can be used as a new background model, and the process can be repeated. In theory, each iteration should bring us closer to the true set of bottom parameters, and we can accept the result to which the process converges to be the best estimate of those parameters. This iterative sequence of steps, which typically involves at most a few tens of iterations, is to be contrasted with matched field inversion methods which search the parameter space by computing thousands or tens of thousands of forward models.

The paper is organized as follows: In Sec. II we derive the perturbation results we use in the method and set up the linear equations to be inverted. Section III discusses the singular value decomposition method of performing the inversion, and discusses the errors associated with such an inversion. In Sec. IV we apply the mode amplitude perturbation algorithm to two data sets to demonstrate its performance. The first is a simple synthetic case in which the true answer is known, and the second is a real data set from the LWAD 99-1 Experiment.[4] In Sec. V we combine the eigenvalue and mode amplitude perturbation methods into a single algorithm and apply this combined method to a synthetic data set to investigate its capabilities. The last section summarizes our results and conclusions.

II. DERIVATION OF THE MODAL AMPLITUDE PERTURBATIVE INVERSION ALGORITHM

In this section we derive the modal amplitude perturbative inversion algorithm, starting with the application of perturbation theory to the basic normal mode equations. We then discuss various approaches to bottom parametrization, and end the section by describing the inversion process itself. Our derivation of the perturbation result for the mode amplitudes follows the work of Tindle *et al.*[5]

We start with the depth-separated normal mode equation (cf. Refs. 6–8)

$$\frac{1}{\rho(z)}\frac{d^2Z_n}{dz^2}+\frac{d}{dz}\left(\frac{1}{\rho(z)}\right)\frac{dZ_n}{dz}+\frac{1}{\rho(z)}[q(z)-k_n^2]Z_n=0, \quad (1)$$

where z is the depth coordinate, positive downwards (with the air-water interface at $z=0$), $\rho(z)$ is the density, $Z_n(z)$ is the nth mode function, k_n is the modal eigenvalue associated with $Z_n(z)$, and $q(z)=k^2(z)=\omega^2/[c^2(z)]$, where $c(z)$ is the sound speed, and ω is the angular frequency of the time-harmonic source signal. We propose a perturbation to the waveguide such that $q(z)\rightarrow q(z)+\Delta q(z)$, which causes perturbations in the other terms: $Z_n(z)\rightarrow Z_n(z)+\Delta Z_n(z)$, and $k_n \rightarrow k_n+\Delta k_n$.

If we collect the unperturbed terms we get the unperturbed equation. If we collect terms with first-order perturbations we find that

$$\frac{1}{\rho(z)}\frac{d^2\Delta Z_n}{dz^2}+\frac{d}{dz}\left(\frac{1}{\rho(z)}\right)\frac{d\Delta Z_n}{dz}+\frac{1}{\rho(z)}[q(z)-k_n^2]\Delta Z_n + \frac{1}{\rho(z)}[\Delta q(z)-2k_n\Delta k_n]Z_n=0. \quad (2)$$

We assume that the perturbations are small, so terms of higher than first order are neglected.

Because the unperturbed normal modes form a complete set, we can expand any function of z in terms of them. We propose an expansion of ΔZ_n of the form $\Delta Z_n(z)=\Sigma_j a_{nj}Z_j(z)$. Substituting this into Eq. (2) we find that

$$\sum_j a_{nj}\left[\frac{1}{\rho}\frac{d^2Z_j}{dz^2}+\frac{d}{dz}\left(\frac{1}{\rho}\right)\frac{dZ_j}{dz}\right]+\frac{1}{\rho}(q(z)-k_n^2)\sum_j a_{nj}Z_j + \frac{1}{\rho}(\Delta q(z)-2k_n\Delta k_n)Z_n=0. \quad (3)$$

The term in square brackets can be replaced using Eq. (1):

$$\sum_j a_{nj}\left[-\frac{1}{\rho}(q(z)-k_j^2)Z_j\right]+\frac{1}{\rho}[q(z)-k_n^2]\sum_j a_{nj}Z_j + \frac{1}{\rho}[\Delta q(z)-2k_n\Delta k_n]Z_n=0. \quad (4)$$

Some of the terms containing $q(z)$ cancel, leaving us with

$$\sum_j a_{nj}(k_j^2-k_n^2)\frac{Z_j}{\rho}+\frac{1}{\rho}(\Delta q(z)-2k_n\Delta k_n)Z_n=0. \quad (5)$$

From this point we can make use of the orthonormality property of the normal modes. If we apply the operator $\int_0^D(\bullet)Z_n(z)dz$ to the equation, we are left with

$$\int_0^D\frac{\Delta q(z)Z_n^2(z)}{\rho(z)}dz-2k_n\Delta k_n = 0 \Rightarrow \Delta k_n=\frac{1}{2k_n}\int_0^D\frac{\Delta q(z)Z_n^2(z)}{\rho(z)}dz, \quad (6)$$

which is the result in Ref. 1. If we instead apply the operator $\int_0^D(\bullet)Z_m(z)dz$ (note the change of subscript from n to m) we get

$$a_{nm}(k_m^2-k_n^2)+\int_0^D\frac{\Delta q(z)Z_m(z)Z_n(z)}{\rho(z)}dz = 0 \Rightarrow a_{nm}=\frac{1}{k_n^2-k_m^2}\int_0^D\frac{\Delta q(z)Z_n(z)Z_m(z)}{\rho(z)}dz, \quad (7)$$

which is valid for $m\neq n$. This result, along with the above-used expansion, allows us to express the change in a mode function $\Delta Z_n(z)$ due to a change in the profile, $\Delta q(z)$. Since the quantity that is actually changing in the bottom is the sound speed, $c(z)$, we can use the perturbation result that $\Delta q(z)=-2\Delta c(z)\omega^2/c^3(z)$ to put the expression in terms of the sound speed perturbation. It should be noted that Ref. 5 lists

the $m=n$ term as being $a_{nn}=-1/2\Sigma_{m\neq n}a_{nm}^2$ based on the idea that the mode function must retain its normalization. However, this result is inconsistent with the earlier neglect of terms of second and higher order. It can be shown, however, that the actual value for the $n=m$ term must be zero, since the change of the mode function must be orthogonal to the mode function itself:

$$\int_0^D \frac{Z_n^2(z)}{\rho(z)}dz = 1 \Rightarrow \frac{\partial}{\partial X}\int_0^D \frac{Z_n^2(z)}{\rho(z)}dz = \frac{\partial}{\partial X}1$$

$$\Rightarrow \int_0^D \frac{2Z_n(z)\,\partial Z_n(z)/\partial X}{\rho(z)}dz = 0 \tag{8}$$

for any parameter X that would cause a change in the mode function. In order to properly make use of the normalization as in Ref. 5, one would have to retain higher order terms in the original perturbation.

The sound speed, $c(z)$, is not the only bottom parameter that affects the mode functions and eigenvalues. The density profile $\rho(z)$ is also a factor. Analogous to the results for sound speed in Ref. 5, we can use perturbation theory to determine the effects of perturbations to a density parameter. If we define the function $\beta(z)=1/\rho(z)$ and denote differentiation with respect to z with primes, our modal equation becomes

$$\beta(z)Z_n'' + \beta'(z)Z_n' + \beta(z)[k^2(z) - k_n^2]Z_n = 0. \tag{9}$$

Just as we did for sound speed, we can introduce a perturbation, collect first-order terms, expand the perturbed mode functions in terms of the unperturbed mode functions, and apply the orthonormality property of the unperturbed modes. When we do so, we obtain two perturbation results for the density parameter:

$$\Delta k_n = \frac{1}{2k_n}\int_0^D Z_n(z)Z_n'(z)\left(\Delta\beta\frac{\beta'}{\beta} + \Delta\beta'\right)dz, \tag{10}$$

which is the eigenvalue perturbation result for density, and

$$a_{nm} = \frac{1}{k_n^2 - k_m^2}\int_0^D Z_m(z)Z_n'(z)\left(\Delta\beta\frac{\beta'}{\beta} + \Delta\beta'\right)dz, \tag{11}$$

which is the result for the modal expansion coefficients, analogous to Eq. (7). Note that because both the equation for the eigenvalue perturbation, and the expansion of the mode function perturbation involve z derivatives, discontinuities must be handled with particular care.

Another bottom parameter of interest is the attenuation profile. Frisk[6] and other texts show that if we introduce an imaginary component to the sound speed profile by making $k(z)$ complex, the eigenvalue becomes complex as well. While the change in the mode function itself is usually negligible when attenuation is added, the complex portion of the eigenvalue creates the appearance that the entire mode function has been reduced. If we make the small change $k(z) \rightarrow k(z)+i\alpha(z)$, the modal eigenvalue will be changed as well: $k_n \rightarrow k_n + i\delta_n$, where the modal attenuation is given by

$$\delta_n = \frac{1}{k_n}\int_0^D \frac{\alpha(z)}{\rho(z)}k(z)Z_n^2(z)dz. \tag{12}$$

This leads to the apparent change in the mode function $Z_n \rightarrow Z_n e^{-\delta_n r}$, or $\Delta Z_n = -Z_n(1-e^{-\delta_n r})$. For many purposes it is often best to look at the apparent change in the mode function over a range step due to the attenuation within that range step. In such a case, the apparent change in the mode function from the beginning of the step to the end will be $Z_n \rightarrow Z_n e^{-\delta_n \Delta r}$, where Δr is the length of the range step. If $\delta_n \Delta r \ll 1$ we can use the Taylor expansion for the exponential, and keep only the linear terms, giving $Z_n \rightarrow Z_n(1-\delta_n\Delta r)$ or $\Delta Z_n \approx Z_n\delta_n\Delta r$. Because of its dependence on range, attenuation can seriously complicate our inversion. Fortunately, in cases where the measurements are taken at sufficiently short ranges from the source, attenuation can be neglected, or estimated by other means.[9]

Further bottom parameters, such as shear speeds and shear attenuations, may also be sought, in which case perturbation results for those parameters would also be needed. For the applications to low-frequency sediment acoustics of interest in this paper, the results for sound speed, density, and attenuation perturbations are considered sufficient. In fact, we will focus on the sound speed perturbation results, since they tend to have the strongest effect on the modal parameters.

The next step in the derivation of the inversion algorithm is to discretize our representation of the bottom. We do this so that we can solve for a finite number of unknowns rather than functions of the continuous variable z. We start by making the assumption that the unknown functions $c(z)$, $\rho(z)$, and $\alpha(z)$ can be written in terms of a weighted expansion of some known depth functions with unknown coefficients. For example, $c(z)$ can be expanded as $c(z)=c_0(z)+\Sigma_i X_i c_i(z)$. Here $c_0(z)$ is a hypothesized, or background, model for the sound speed profile. The functions $c_i(z)$ are arbitrary, user-defined functions that should be selected so as to be able to capture the important features of the sound speed profile. The unknown scalar coefficients X_i are what we seek, since once we have them we can reconstruct the sound speed profile.

The number and complexity of the $c_i(z)$ functions that should be used depends on how detailed a profile is required, and what kinds of bottom features are considered possible. Ideally one should use the smallest number of parameters that fully capture the bottom features. However, since one does not usually know beforehand what bottom features are present, it will often be necessary to use more than the ideal number of parameters.

Once the unknown functions have been parametrized we can compute the derivatives of the mode functions with respect to the unknown scalars X_i using the perturbation results we derived earlier. To do this we make the substitution $\Delta c(z)=X_i c_i(z)$ for the sound speed equations, $\Delta\beta(z)=B_i\beta_i(z)$ for the density, and $\alpha(z)=A_i\alpha_i(z)$ for the attenuation, and use the fact that $\Delta Z_n/X_i \approx \partial Z_n/\partial X_i$ for sufficiently small values of X_i. The results of this procedure are

$$\frac{\partial Z_n(z)}{\partial X_i} = \sum_m a_{nm}^{X_i} Z_m(z), \tag{13}$$

where

$$a_{nm}^{X_i} = \frac{2\omega^2}{k_m^2 - k_n^2} \int_0^D \frac{c_i(z) Z_n(z) Z_m(z)}{c_0^3(z)\rho(z)} dz, \tag{14}$$

$$\frac{\partial k_n}{\partial X_i} = \frac{-1}{k_N} \int_0^D \frac{c_i(z) Z_n^2(z)}{c_o^3(z)\rho(z)} dz, \tag{15}$$

$$\frac{\partial Z_n(z)}{\partial B_i} = \sum_m a_{nm}^{B_i} Z_m(z), \tag{16}$$

where

$$a_{nm}^{B_i} = \frac{1}{k_n^2 - k_m^2} \int_0^D Z_m(z) Z_n'(z) \left(\beta_i \frac{\beta_0'}{\beta_0} + \beta_i' \right) dz, \tag{17}$$

$$\frac{\partial k_n}{\partial B_i} = \frac{1}{2k_n} \int_0^D Z_n(z) Z_n'(z) \left(\beta_i \frac{\beta_0'}{\beta_0} + \beta_i' \right) dz, \tag{18}$$

$$\frac{\partial Z_n(z)}{\partial A_i} \approx \frac{Z_n(z)\Delta r}{k_n} \int_0^D \frac{\alpha_i(z)}{\rho(z)} k(z) Z_n^2(z) dz. \tag{19}$$

With these derivatives, we are nearly ready to carry out the inversion. Given our background mode functions, and measurements of the actual mode functions, we could move on to the actual inversion. However, in the real world, where source levels are often not known precisely, and where precise instrument calibration can be an issue, using the mode functions themselves can be problematic. Further, since the expression for the field always contains the product of the mode functions at two depths, it is usually not possible to measure the mode function by itself. Because of these issues we will actually use the ratio of each mode function to the first mode function. This eliminates source level and calibration concerns and only slightly complicates our calculation. If we define the quantity

$$m_n(z, z_s) \equiv \frac{Z_n(z) Z_n(z_s)}{Z_1(z) Z_1(z_s)}, \tag{20}$$

we can compute the derivative using the product and quotient rules, and our earlier results. Note that when using the adiabatic approximation,[10] $Z_n(z)$ is computed at the receiver location, and $Z_n(z_s)$ at the source location. The derivative of the ratio with respect to some parameter γ can be written as

$$\begin{aligned} \frac{\partial m_n}{\partial \gamma} = & \frac{\partial Z_n(z_s)/\partial\gamma Z_n(z) Z_1(z_s) Z_1(z)}{Z_1^2(z_s) Z_1^2(z)} \\ & + \frac{\partial Z_n(z)/\partial\gamma Z_n(z_s) Z_1(z_s) Z_1(z)}{Z_1^2(z_s) Z_1^2(z)} \\ & - \frac{\partial Z_1(z_s)/\partial\gamma Z_n(z_s) Z_1(z_s) Z_n(z)}{Z_1^2(z_s) Z_1^2(z)} \\ & - \frac{\partial Z_1(z)/\partial\gamma Z_n(z_s) Z_1(z_s) Z_n(z)}{Z_1^2(z_s) Z_1^2(z)}. \end{aligned} \tag{21}$$

Note that the first mode has been selected somewhat arbitrarily here. Any other mode can be used in the denominator if it is more convenient. However, the first mode is usually the best choice, since it is nonzero everywhere in the water column and is usually less affected by changes in the bottom than the other modes.

It should also be pointed out that by dealing with the mode ratios we are implicitly assuming that there is no mode coupling. This method is intended for use in cases where the adiabatic approximation[10] is valid. In cases where this assumption is violated the transfer of energy between modes can give rise to apparent changes in the bottom properties which do not actually reflect the true bottom. The method can still be used as long as the true mode ratios, as defined earlier, are used as input. However, measuring those quantities when there is mode coupling is a nontrivial problem.

At this point we are ready to set up the inversion. We create a matrix, $[\partial m/\partial X]$, each row of which contains the partial derivatives of one mode function at one depth with respect to each of the parameters sought. For example, if the parameters sought are X_1, X_2, B_1, A_1, the first row of $[\partial m/\partial X]$ is

$$\left[\frac{m_2(z_1, z_s)}{\partial X_1} \frac{m_2(z_1, z_s)}{\partial X_2} \frac{m_2(z_1, z_s)}{\partial B_1} \frac{m_2(z_1, z_s)}{\partial A_1} \right], \tag{22}$$

where z_1 is the first measurement depth. We start with m_2 because m_1 is always equal to 1. The next few rows of $[\partial m/\partial X]$ may be the same derivatives but evaluated at different measurement depths. It should be pointed out that using multiple measurement depths does not necessarily give any additional, independent information. However, by using more depths one gets more robustness to measurement noise, which is especially useful when one depth is near a null of one of the modes.

There are two other quantities that are needed before we can carry out the inversion: the column vector $\mathbf{X}$, which contains all the parameters we seek, and the column vector $\Delta \boldsymbol{m}$, which contains the difference between the mode ratios at each depth and mode number. Some might find it more intuitive to think of the $\mathbf{X}$ vector as a $\Delta \boldsymbol{X}$ vector, which contains the *differences* between the desired parameters in the background model and the true case. This notation becomes a bit cumbersome, however, when partial derivatives become involved, so we have opted for calling the vector just $\mathbf{X}$.

The various components having been collected, we now have

$$\left[\frac{\partial m}{\partial X} \right] \mathbf{X} = \Delta \boldsymbol{m}. \tag{23}$$

We can solve this equation in the least-squared error sense using the pseudoinverse matrix:[11–13]

$$\mathbf{X}_{\mathrm{LS}} = \left[\frac{\partial \tilde{m}^{\#}}{\partial X} \right] \Delta \boldsymbol{m}, \tag{24}$$

where $[\partial \tilde{m}^{\#}/\partial X]$ is the pseudoinverse matrix of $[\partial m/\partial X]$.

It must be noted that we have worked so far under an approximation of a linear relationship between the variables when performing this inversion, when in reality the relation-

ship is nonlinear. Therefore we should not expect this single-step inversion to give us the correct bottom parameters. However, if our background model resembles reality, then our answer should at least give a correction to our background, which can be used to generate a new background, and the process can be repeated. After a few (usually ~30 or less for cases we have examined) iterations, the algorithm should converge to the values that give the best possible fit (in a least-squares sense) to the input data, given our parametrization. It must be kept in mind, however, convergence is not guaranteed. An initial background model too different from the true bottom, or a parametrization of the bottom which does not capture all important features can result in divergence.

III. SOURCES OF ERROR

To properly interpret the output of inversion algorithm, it is necessary to understand the errors that may be present in it. In this section we will discuss several sources of errors which can contribute to the overall uncertainty in the final estimate of the geoacoustic parameters. Even when the errors in the input vector are small, careless use of the pseudoinverse can lead to large errors in the estimates of the bottom parameters. Avoiding this possibility requires an understanding of how singular value decomposition is used to obtain the pseudoinverse matrix. As singular value decomposition is a well-known technique, we will not describe it here, but instead direct the reader to any standard text on matrix algebra, such as Ref. 14.

In cases where the pseudoinverse is a true inverse matrix, there are three types of errors to consider: (1) Errors in the measurement vector, (2) errors due to overly simple parametrization of the bottom, and (3) errors due to convergence to an incorrect local minimum. If the iteration process converges to the correct answer, and the parametrization of the bottom is sufficient to match reality, then the error in the estimate is linearly related to the error in the measurement. The bias of the input data should be zero. If it is not, it should be subtracted from the input data before the inversion. For completeness, however, we state that if there is bias in the input, the bias in the parameter estimation is equal to the input bias times the inversion matrix:

$$\langle \boldsymbol{X}_{\mathrm{LS}} \rangle = \left[\frac{\partial \tilde{m}^{\#}}{\partial X} \right] \langle \Delta \boldsymbol{m} \rangle. \tag{25}$$

Similarly, the variance of the estimate is also dependent on the variance of the input:

$$\mathrm{cov}(\boldsymbol{X}_{\mathrm{LS}}) = \left[\frac{\partial \tilde{m}^{\#}}{\partial X} \right] \mathrm{cov}(\Delta \boldsymbol{m}) \left[\frac{\partial \tilde{m}}{\partial X} \right]^{T}. \tag{26}$$

This expression allows us to compute the covariance of our inversion results based on the covariance of our input data, and thus quantify the uncertainty in our results.

It must be kept in mind, however, that implicit in this expression are the assumptions that the algorithm has converged to the correct minimum, that the errors in the input data are small enough that they propagate linearly through the process to the inversion results, and that the parametrization is sufficient to capture all features of the bottom. Errors in the measurement vector are beyond our control. If they are so large as to violate a linear approximation, there is little hope for the algorithm.

The other two sources of error are in competition to some degree. The more parameters used to describe the bottom, the less likely the algorithm is to miss important bottom features. However, the more parameters that are used, the more the problem becomes underdetermined. This results in a greater reliance on the smoothness and minimum norm assumptions used to compute the pseudoinverse matrix, which may not be valid.

Further understanding can be gained by examining the resolution matrix. This is especially informative when the problem is underdetermined, as is often the case in such problems. The resolution matrix, $[R]$, is defined:

$$[R] = [V][V]^{T}, \tag{27}$$

where $[V]$ is the matrix composed of the eigenvectors of $[\partial m/\partial X]^{T}[\partial m/\partial X]$. When the rank, p, is equal to the number of unknowns, N, then $[R]$ will equal the identity matrix. Otherwise, $[R]$ will be such that its elements minimize the quantity

$$\sum_{i} \sum_{j} ([R]_{ij} - \delta_{ij})^{2}, \tag{28}$$

where δ_{ij} is the Kronecker delta function. The degree to which $[R]$ resembles the identity matrix gives the user an idea of the resolution of the inversion result, in that off-diagonal terms will represent "smearing" of one bottom parameter into the others. If the $[R]_{ij}$ is nonzero, that indicates that a change in the ith parameter will show up also as a change in the jth parameter in the inversion result. The resolution matrix is of particular usefulness when using c-at-each-z type bottom parametrizations, because parametrizations that use a small number of parameters are less likely to be underdetermined. Backus and Gilbert[15] address the concept of resolution in much greater detail.

It must also be pointed out that the pseudoinverse matrix may very well be unstable if all singular values are used to compute it. It is usually best to ignore very small singular values when computing the pseudoinverse to make the inversion more robust to error. Doing so reduces the resolution of the inversion, but will also reduce the variance in the answer. Deciding the minimum size of singular values which can be tolerated can be difficult, and some trial-and-error may be necessary to strike the right balance.

IV. EXAMPLES OF MODE AMPLITUDE INVERSION

In this section we demonstrate the feasibility of the mode amplitude perturbation algorithm by applying it to two example cases. The first is a simple synthetic data case without input errors, which serves as a basic proof of concept, while the second makes use of experimental data. For now, we keep the mode amplitude perturbation separate from the already-proven eigenvalue perturbation method in Ref. 1, so

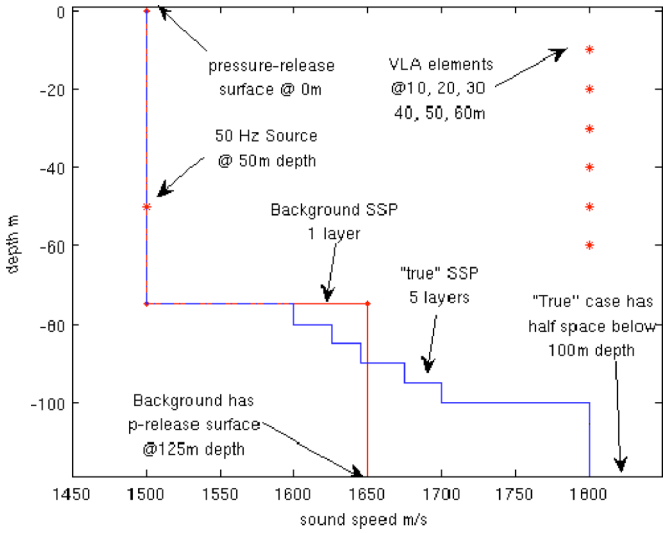


FIG. 1. (Color online) The sound speed profile and experimental geometry for the first example.

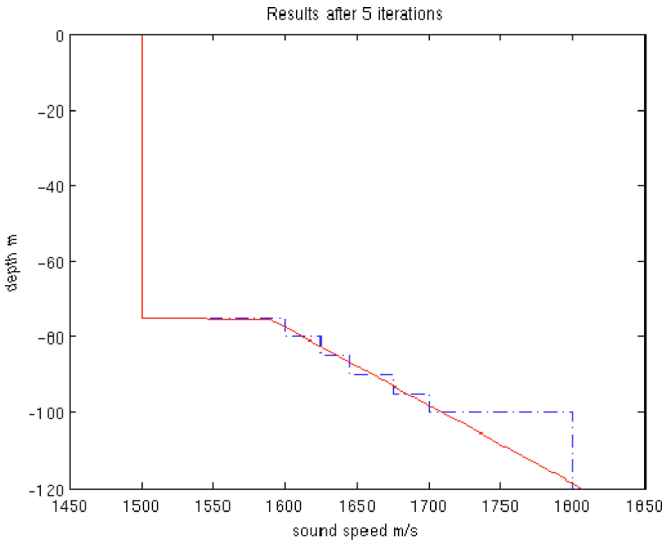


FIG. 2. (Color online) The results of the geoacoustic inversion algorithm after five iterations. The dotted curve shows the "true" sound speed profile, while the solid curve indicates the inversion result.

as to make sure our new method works as expected. In the next section, we discuss how the two methods can be combined.

The first test of the method is a very simple, two-parameter, range-independent example with perfect input data. We hypothesize a "true" waveguide of 75 m of isovelocity water overlying 25 m of layered sediment. Below this, we add a half-space which continues to infinite depth. We excite a pressure field by adding an acoustic point source at 50 m depth, emitting a pure tone at 50 Hz. We assume a constant, known bottom density of 1800 kg/m^3 and zero attenuation. We use the normal mode program KRAKEN[16] to generate the mode functions that will be excited in this waveguide. For this first test, we will use the mode functions directly, rather than computing the field and trying to estimate them. This separates out any potential problems in the inversion method from problems in the mode function estimation algorithm. Since we are using the mode functions directly, and the problem is range independent, we do not need to specify a range for the VLA. We do, however, need to specify the depths of the receivers, and in this case, we will use a six-element VLA with 10 m spacing. The general setup is shown in Fig. 1.

The information given so far is sufficient for solving the forward problem. However, to perform the inversion we also need to specify a background model, and parametrize the bottom. We choose a simple two-parameter bottom model which describes the bottom using only the sound speed at the water-bottom interface, and the sound speed gradient. This model is appealing because the low number of parameters allows us to solve a fully determined problem as long as we have at least three propagating modes, and because it is often capable of describing the true bottom as a reasonable first approximation. While there are layers in the true bottom, the rate of increase of the sound speed of the layers is nearly linear with depth, so a gradient can do well to approximate the sound speed profile.

For the background model, we use a Pekeris-like waveguide with a sound speed of 1650 m/s throughout the bottom. However, unlike the standard Pekeris model, we include a pressure-release surface at 125 m depth. The reason for this is that the expression for the derivative of the mode function includes a sum over all modes. In order to have a complete set, we must have an infinite number of modes, which is only possible if we include the nonpropagating modes that have complex eigenvalues. Solving for complex eigenvalues is a difficult task, and greatly increases the computation time needed for solving the problem. By adding a pressure-release false bottom, we can maintain a proper Sturm–Liouville problem, and the eigenvalues of the nonpropagating modes become purely imaginary, and thus much easier to compute. The error introduced by our use of this false bottom should be small if it is placed at a depth below the turning point of the highest propagating mode.[17]

With all of the necessary components in place, we can now execute the inversion process. The mode functions are computed for the background model, and the mode ratios computed for each mode, and each receiver depth. These are compared to the mode ratios formed using the output from KRAKEN. The derivatives of the ratios are computed following the results of Sec. II. The linear set of equations is solved using the pseudoinverse, giving the necessary changes in the two parameters. These changes in the parameters are incorporated into the background model, and the process repeated again. After five iterations of this process, the algorithm converges to the result shown in Fig. 2, which agrees well with the true profile.

There is nontrivial disagreement below 100 m depth, but using our simple two-parameter model, this is unavoidable. For the first 25 m of the bottom, the fit is as good as our two-parameter model allows, showing that given perfect input data, the method will provide a good estimate of the bottom sound speed profile (SSP). Also, since this result was obtained using a background model containing a pressure-release false bottom, we gain confidence that using such false bottoms in a proper fashion in our background model will not corrupt the inversion results.

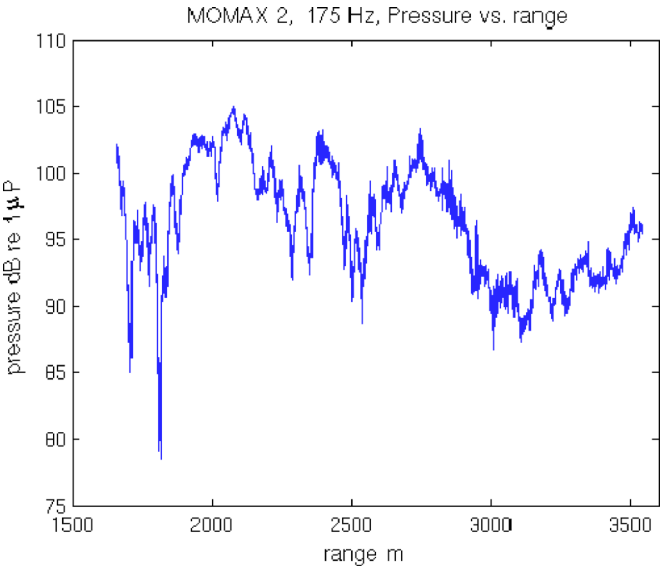


FIG. 3. (Color online) The pressure magnitude vs range measured during MOMAX 2, at 175 Hz.

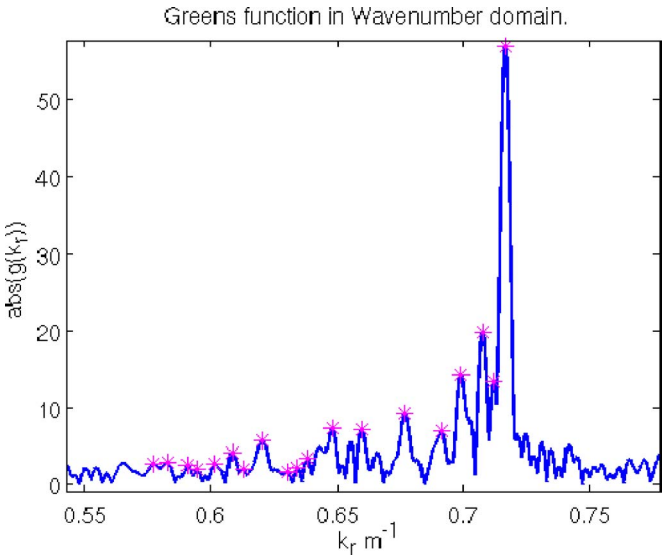


FIG. 4. (Color online) The Hankel transform of the pressure vs range data shown in Fig. 3. Mode peaks used in the inversion are indicated with asterisks.

For our second example we use data from the LWAD 99-1/MOMAX 2 Experiment, which was carried out in the Gulf of Mexico in February 1999.[4] During this experiment the mode amplitudes were estimated by taking a Hankel transform of pressure versus range data, a procedure described in Ref. 9 and used in Ref. 1. A single hydrophone suspended from a drifting buoy moved through the sound pressure field created by a single 175 Hz pure tone source, suspended from the moored R/V Gyre. A synthetic aperture HLA was formed as the receiver drifted, providing the input data for the Hankel transform. This was possible because the waveguide was very nearly range independent over the section of data used, and the 70-m-deep water column was stable during the time period over which the synthetic aperture was formed.

Figure 3 shows the pressure magnitude as a function of range measured by the drifting receiver, after the raw pressure measurement was demodulated for the source frequency of 175 Hz. The magnitude and phase data were then Hankel transformed to obtain an estimate of the depth-dependent Green's function versus horizontal wave number, which is shown in Fig. 4.

The 19 modal peaks used as input data to the inversion algorithm are indicated with asterisks. It may seem unreasonable to assume that all of these peaks are actual modal peaks rather than sidelobes of the finite-aperture transform, but each peak lined up surprisingly well with the expected eigenvalue locations based on prior estimates[18] of the geoacoustic parameters for this location. Further, the widths of the peaks also seem to indicate main lobes rather than sidelobes, as can be seen by observing the narrow sidelobes to the right of the first (largest eigenvalue) mode. It is possible that some of the modes have been misidentified, but we assume the large majority of the modal peaks are correctly interpreted. The possibility of misidentifying modal peaks is a common problem when dealing with experimental data, and is a problem any modal inversion method must be able to handle.

In order to reduce the errors introduced by a few misinterpreted peaks, we use a simple two-parameter model for the sea floor. By solving an overconstrained problem we tend to reject unbiased errors in our input, whereas if we solve an underdetermined problem the algorithm will introduce erroneous bottom properties in order to fit the errors introduced by the measurement. The cost of this, however, is that we are limited to inverting very simple sound speed profiles in the bottom. Fortunately, the precruise estimate in Ref. 18 predicts a bottom that could be described reasonably well with a two-parameter model. If this were not the case, we would have to use more parameters to describe the bottom.

For a background model, we again choose a simple Pekeris-like model. While we have a better *a priori* estimate of the bottom SSP in this case, one usually will not have such a good starting point, so it is a better test of the method to start with a less accurate background model. We use our background model to compute the mode ratios at the receiver depth, and compare those to our measurements. We compute the derivatives of the ratios with respect to our two parameters, and solve the linear system for the necessary changes in the parameters, as described in Sec. III. We repeat this process until we have convergence, and the results are shown in Fig. 5.

It is clear that the final inversion result matches the precruise estimate well. That 25 iterations were required before convergence may be a sign that some modes were misidentified, or that our bottom parametrization failed to capture some of the features of the true sound speed profile. It is also possible that some of the parameters which were treated as known, such as the density profile, the attenuation, and the water column SSP, were not accurate. Despite all these possibilities, however, we achieve a result which compares well with our expectation, so we can be confident the mode amplitude perturbation algorithm works.

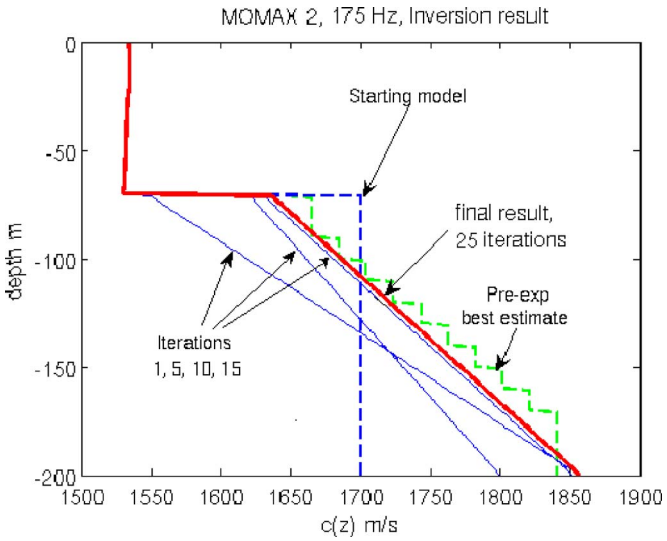


FIG. 5. (Color online) The results of the geoacoustic inversion of the 175 Hz data from MOMAX 2. The vertical dashed line shows the starting model of the sound speed profile, and the solid curves show iterations 1, 5, 10, and 15. The bold solid curve is the final result after convergence at 25 iterations. The step-like dashed curve is the best estimate of the true bottom profile available before the experiment.

V. COMBINING MODE AMPLITUDE AND EIGENVALUE PERTURBATION METHODS

Having seen that mode amplitude perturbation alone can produce a suitable inversion, we now compare the quality of the inversion it produces to that of eigenvalue perturbation method, and then consider the possibility of combining the two methods to provide a better result than either produces on their own. To compare the two methods, we require some metric of performance. The two most common metrics of performance of an inversion algorithm are the bias and variance of the error in the inverted parameters.

As stated earlier, assuming the algorithm converges to the correct minimum, and that the measurement errors are small enough that they affect the estimate linearly, we can compute the bias and variance of our inversion from the bias and variance of the input data using Eqs. (25) and (26). With these expressions for error covariance in hand, it is possible to compare the performance of the two inversion methods. The "better" algorithm is the one with the lower error covariance. However, it is often the case that one method will perform better at estimating one parameter, while the other performs better at estimating another.

For example, eigenvalue perturbation tends to provide a lower-variance estimate of subbottom features, whereas mode amplitude perturbation methods do better at estimating shallower parameters such as interface speeds.[19] This is not unexpected, since the mode functions contain little energy at large depths, and thus are not much affected by the subbottom, whereas the eigenvalues are determined by the boundary conditions, and thus are highly affected by subbottom parameters. Thus, which method is preferred will depend, in part, on which parameters the user is most concerned with estimating accurately.

Further, which method is better will depend on the quality of the input data. If the eigenvalues are estimated very precisely, then eigenvalue perturbation will be superior. On the other hand, if the mode amplitudes are estimated well, but the eigenvalues have significant error, mode amplitude perturbation should be used. Which inversion method is more effective will also depend somewhat on the environment, as this will change the inversion matrix. In short, in order to determine the preferred method for a given scenario, one must perform the inversions and compare the computed error variances.

Since both methods can provide lower-variance estimates for different parameters, it is logical to ask if we can combine the two methods for even better estimates. At first it might seem best to simply take the values of each parameter from the inversion with the lower error variance. However, the estimates of the various parameters are not independent, and mixing and matching them can lead to problems. A better solution is to use all the information available to both methods at once. To do this we can combine the two derivative matrices into one, and the two data vectors into one, forming a new equation:

$$\begin{bmatrix} \frac{\partial Q_{MA}}{\partial X} \\ - \\ \frac{\partial Q_E}{\partial X} \end{bmatrix} \mathbf{X} = \begin{vmatrix} \Delta \boldsymbol{Q}_{MA} \\ - \\ \Delta \boldsymbol{Q}_E \end{vmatrix}. \tag{29}$$

This equation is essentially the same as the one used by either of the methods, but we cannot solve it in quite the same way (i.e., using the Morse–Penrose pseudoinverse), because the two types of input data differ significantly in size. Eigenvalue perturbations are typically on the order of 0.001, whereas mode ratio amplitude perturbations are closer to order 0.1 or even 1. Using the pseudoinverse to find a least-squares solution to Eq. (29) would essentially result in the same answer as if we just used the mode amplitude equation. The eigenvalue perturbations are just too small to affect the answer significantly. Thus, we must take into account the relative sizes of the perturbations rather than the absolute sizes. One way of doing this is to weight each equation (i.e., multiply each row of the above-presented matrix and the corresponding entry in the data vector) by the standard deviation of the perturbation, and then compute the pseudoinverse.

An even better method is to make use of the stochastic inverse (SI).[20,21] This method for creating the inverse matrix not only accounts for the relative sizes of the different perturbations but also for the relative sizes of the changes in the parameters. For example, an interface speed can easily change by 10 m/s over an experiment, whereas changes in a gradient are likely to be much smaller. We will leave the derivation of the stochastic inverse to the references, and state here only the expression for the inverse. For

$$\boldsymbol{d} = [G]\boldsymbol{q} + \boldsymbol{e}, \tag{30}$$

where $\boldsymbol{d}$ is the measurement vector, $\boldsymbol{e}$ a zero-mean noise vector with covariance R_e, and $\boldsymbol{q}$ the unknown, zero-mean parameter vector with covariance R_q, the stochastic inverse solution is

$$\hat{q}_{\mathrm{SI}} = [\tilde{G}^{\mathrm{SI}}]\boldsymbol{d} = [R_q][G]^T[[G][R_q][G]^T + [R_e]]^{-1}\boldsymbol{d}. \quad (31)$$

This inversion minimizes the errors squared, but normalizes by the uncertainty in the quantity with the error. For example, an error in a quantity with variance 0.0001 should be much less than an error in a quantity with a variance of 0.1. The inversion also tries to minimize the q vector, again weighted by the variance of the elements in the vector.

Though the stochastic inverse requires that the user supply considerably more information, which may have to be estimated, it allows one to combine the eigenvalue and mode amplitude perturbation results and find a solution that takes into account all the data available in the pressure field. By looking at the covariance of our estimate, we can determine which of the three methods (mode amplitude, eigenvalue, or the combination) gives the best estimate of the bottom parameters.

Another advantage of combining the two methods is a decreased chance of converging to an incorrect local minimum. While there is little danger of this happening when only a small set of parameters is used to describe the bottom, it is quite possible when a large parameter set is used. The more input data there are, the more "ways out" of the local minima there are, and the less likely the algorithm is to get trapped. The possibility of convergence to an incorrect local minimum is a problem that the user must keep in mind. If the algorithm converges to a profile that does not seem realistic, it is likely that it has reached a local minimum, and the user should try starting with a different background model, to see if the same result is obtained.

Even if the profile produced does look reasonable and reproduces the measurements well, it is still possible (though unlikely) that the profile does not match the true bottom! Unfortunately, this is inescapable due to the underdetermined nature of the problem. That said, using more input data does tend to reduce the risk of local minima. This is only true, of course, if the additional input data contain at least some independent information. While much of the information contained in the mode amplitudes is redundant to that in the eigenvalues, some of the information is independent. Using both sources of information should reduce the chances of there being reasonable-looking profiles at incorrect local minima.

In order to illustrate the combined eigenvalue and mode amplitude perturbation method we examine a synthetic example. For this case we will posit a waveguide 200 m deep and 4000 m in length. The water column is 50 m deep and is treated as isovelocity (1500 m/s, 1000 kg/m^3). Below this is a 70 m deep, range-dependent sediment layer. Two parameters (interface speed and gradient) are sufficient to describe this layer at any given range. Density is treated as constant in the layer (1600 kg/m^3). Below the sediment layer is a range-independent subbottom (1800 m/s, 1800 kg/m^3). This waveguide is shown in Fig. 6.

A synthetic field was generated for the waveguide, simulating a VLA at range 0, and a 100 Hz source moving from 200 to 4000 m from the VLA, at a constant depth of 30 m. Seven propagating modes were used to generate the synthetic field. The VLA had nine elements, with 5 m spacing,

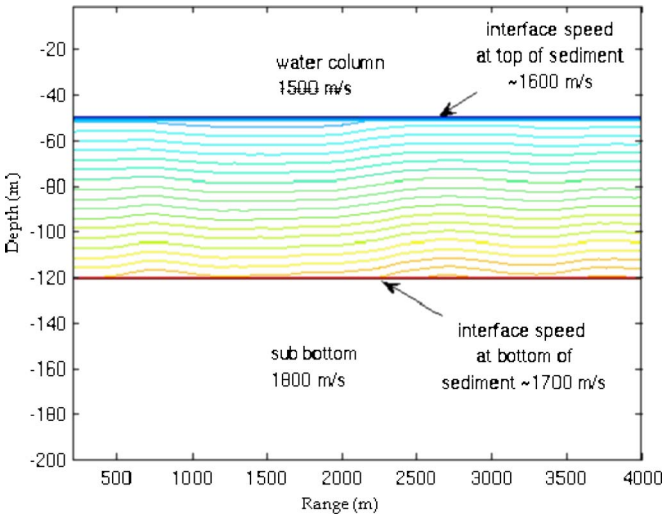


FIG. 6. (Color online) Sound speed contours of the synthetic waveguide. The contours indicate lines of constant sound speed. The water column and subbottom are treated as isovelocity and range independent, whereas the sound speed in the sediment layer varies both in depth and range.

going from 5 m depth to 45 m. Zero mean, Gaussian random noise was added to the field. The modal parameters were then estimated from the VLA measurements. The method by which this was done will be described in a future paper, and it is outside the scope of this paper. All that is necessary for the inversion is that we have an estimate of the parameters, and an estimate of the covariance of our parameter estimates. Figures 7 and 8 show the estimated and true values of the modal parameters as a function of range.

With the estimates of the mode amplitudes and eigenvalues and their variances in hand, we can estimate the bottom parameters. For the inversions we treat the water and subbottom as known, and invert for the interface speed and gradient of the sediment layer. We also treat density within the sediment layer as known and ignore attenuation. As with the first

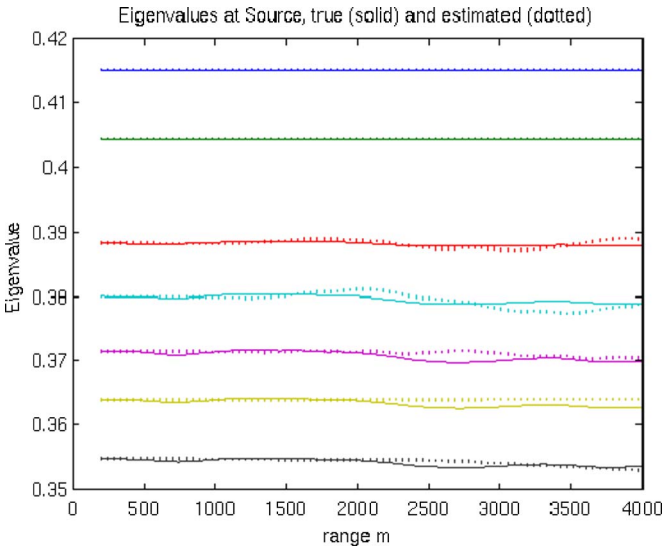


FIG. 7. (Color online) Eigenvalues estimates used to test the combination of eigenvalue and amplitude perturbation methods. The values of the seven eigenvalues at the source location as a function of range are shown. The solid lines indicate the true values, while the estimates are shown with dotted lines.

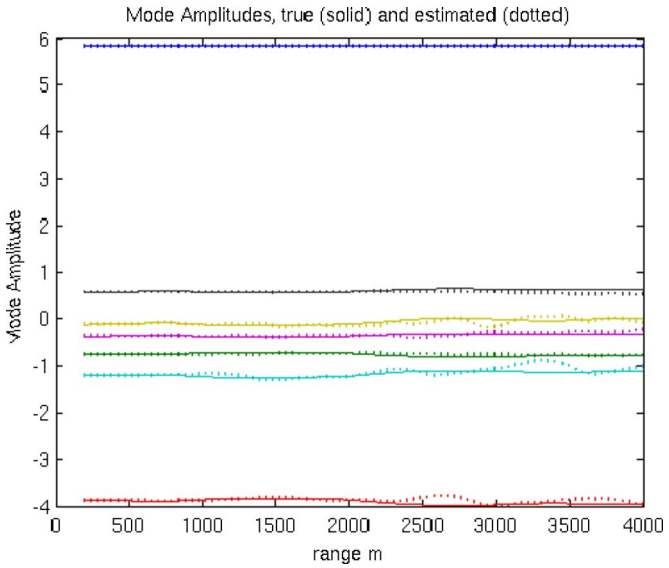


FIG. 8. (Color online) Mode amplitude estimates used to test the combination of eigenvalue and amplitude perturbation methods. The mode amplitudes at 25 m depth on the VLA as a function of source-receiver range are shown. The true values are depicted with solid lines, and estimated values are shown as dotted lines.

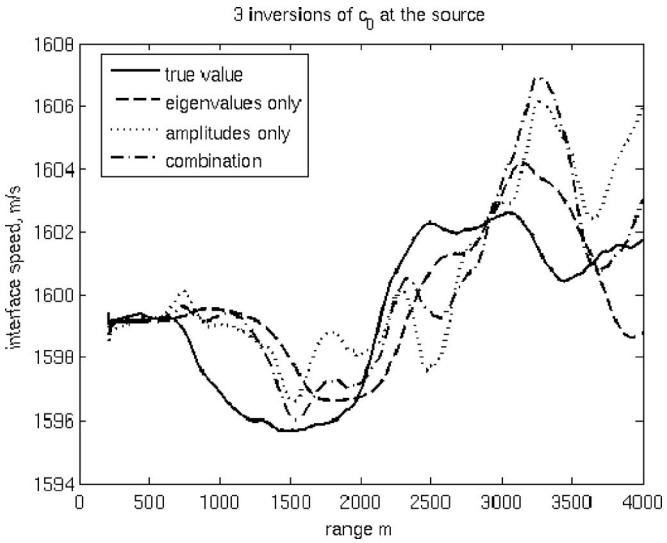


FIG. 10. Interface sound speed at the source, real and inverted. The true interface speed (solid) as a function of range for the synthetic experiment is shown. The three inverted values are also shown. The eigenvalue perturbation result is shown as dashed lines, the mode amplitude inversion as dotted lines, and the combination method result in dash-dotted lines. For most of the experiment, the combination method tends to track whichever of the two inversion results is closer to the true value.

test case, this is done to keep our example simple, and is not a requirement of the method. In order to compare the mode amplitude perturbation method with the eigenvalue perturbation method we do separate inversions using each, and compute the error variances using Eq. (26).

The error variance of the four estimated parameters is shown in Fig. 9. In this case, the error variances favor the mode amplitude perturbation for some of the parameters and eigenvalue perturbation for others. For example, the interface speed at the source location as estimated by mode amplitude perturbation has a lower error variance than it does when estimated by eigenvalue perturbation. But the opposite is true

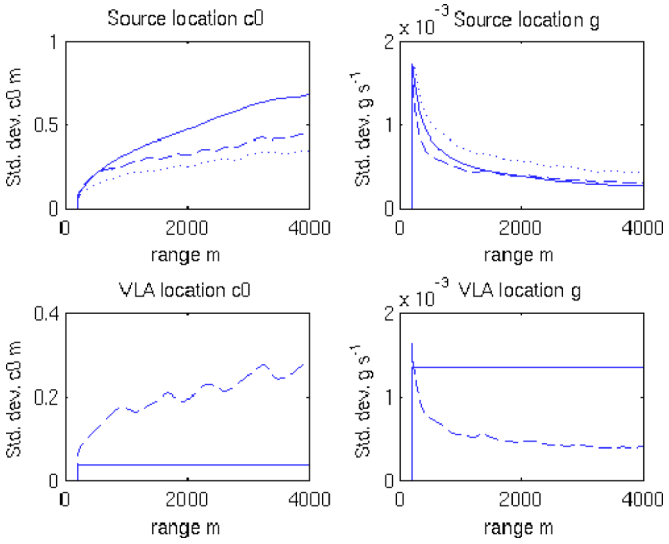


FIG. 9. (Color online) Error variance of the four inverted parameters. The eigenvalue perturbation result is shown as solid lines, the mode amplitude result in dashed lines, and the combined method shown in dotted lines. For the bottom two plots, the curve for the combined method overlays that of the eigenvalue method.

of the receiver-position estimate of the interface speed. Estimated error variance of the source-position gradient is similar with both methods, but the VLA-position estimate of the gradient appears to favor mode amplitude perturbation. This, however, is an error on the part of the parameter estimation algorithm, which has supplied an overconfident estimate of the variance. It is informative to track how such an underestimate of the variance in our estimate of a parameter propagates through our algorithm and affects the inversion result.

Using the stochastic inverse, the estimated covariance of the errors in our measured acoustic parameters, and an estimated matrix of the unknown parameter covariance (the parameters were considered uncorrelated, but the true variances of each parameter were used), we can solve the combined mode amplitude/eigenvalue perturbation equation. And just as we did with the pseudoinverse, we can use the stochastic inverse to compute the error variance of our inversion. When we do so for our synthetic example, we get the best of both methods, which is shown in Fig. 9. As can be seen, the estimate of the source location interface speed is improved over either method, and the variance of the estimate of the VLA position parameters matches that of the eigenvalue method. On the other hand, there is a slight decrease in the quality of the estimate of the gradient at the source position. This is due to our overconfidence in the mode-amplitude method's estimate of this parameter.

Now that we have looked at the predicted accuracy of our inversion, let us examine the actual accuracy. Figures 10–13 show the true values of the parameters along with the inverted values from the three methods. Figure 10 shows the inversion results for the interface speed at the source position. While all three methods are fairly comparable in quality, for most of the experiment the combination result tracks the better of the two other inversions. The differences are not

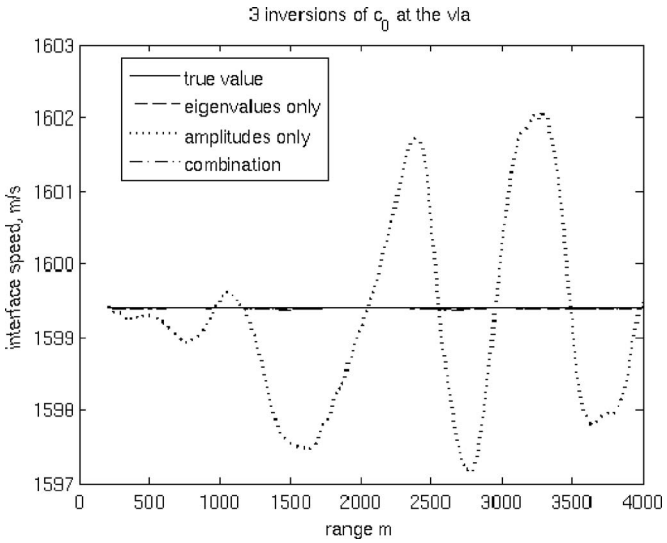


FIG. 11. Interface speed at the VLA, real and inverted. The eigenvalue perturbation result overlays the constant true value of the VLA-position interface speed. As the estimate of the VLA position eigenvalue changes very little, the inverted value of the parameter changes very little as well. The mode amplitudes, however, can change due to changes at the source, or at the VLA, and thus the mode amplitude perturbation result varies with range, and has much larger error. The combination inversion method tracks the result of the eigenvalue perturbation well, and thus remains at the true value.

particularly dramatic in this case, but over the course of the experiment, the combination result does appear to be the best of the three.

Figure 11 shows the inverted values of the interface speed at the VLA position. Here the improvements of the combination method are clear. The eigenvalue perturbation

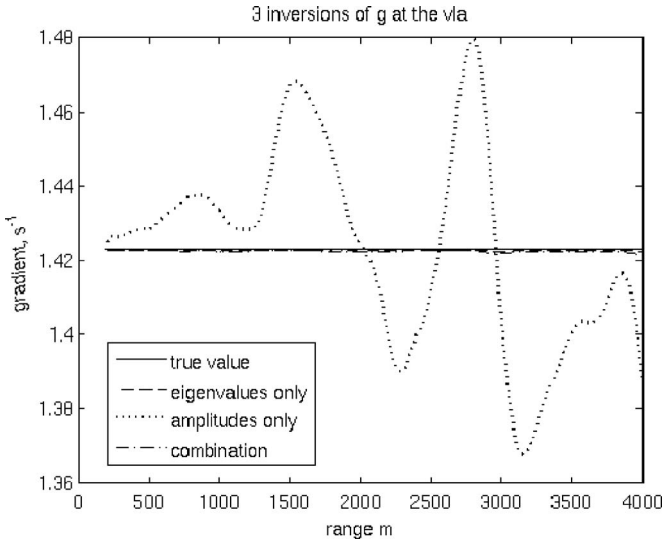


FIG. 12. Gradient at the VLA, real and inverted. As in Fig. 11, the true value is constant, and the unchanging local eigenvalue estimate gives a good result for the eigenvalue perturbation method. The mode amplitude perturbation method, however, has trouble distinguishing between local and source-position changes, and thus provides a poorer estimate of the parameter. The combination result tracks the eigenvalue perturbation result, and the true value, well. Note that the level of error in the mode amplitude perturbation result is at odds with the predicted error variance for this parameter shown in Fig. 9. This is due to an overconfidence (i.e., too low a value for the estimated variance) in some of the estimates used as input data.

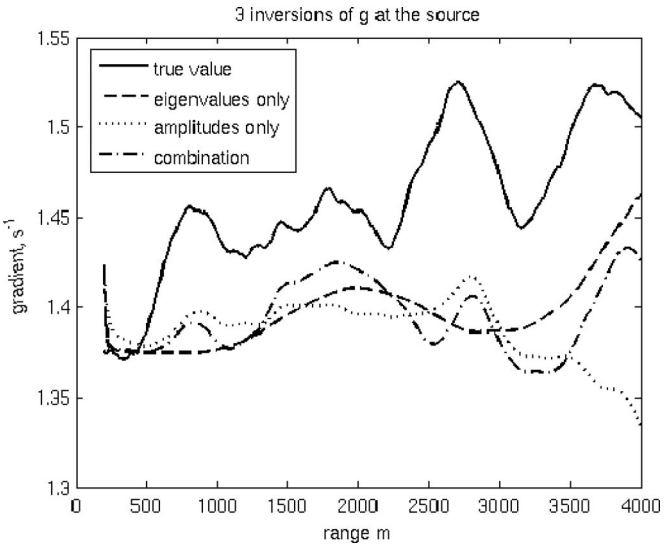


FIG. 13. Gradient at the source, real and inverted. The source-position gradient as a function of source-receiver range is shown. The true value is shown with a solid line. All three estimates are comparable, and show similar amounts of error. However, the combination method does the best at tracking the features of the parameter with range.

method, using an estimate of the unchanging local eigenvalues, does a good job of tracking the true value of the parameter. The mode amplitude method, however, uses mode ratios which are affected by changes at both the source and receiver position, and because of this struggles to distinguish between changes at the source, and changes at the VLA. Thus, its estimate of the parameter is nonconstant, and has significantly larger errors than the eigenvalue perturbation method. The combination method is able to use the local eigenvalue estimate, and thus tracks the true value well. Figure 12 is similar. Again the mode amplitude perturbation method struggles to distinguish between local and source-position changes, and thus gives a nonconstant estimate with larger errors. The methods that make use of the estimate of the local eigenvalue, however, do a good job at tracking the parameter. Note that Fig. 9 would lead us to believe that the mode amplitude estimate would be more precise for this parameter, and yet clearly the eigenvalue estimate is superior. The combined method tracks the eigenvalue result, as we would hope, despite the overconfidence of our variance estimate.

Figure 13 shows the estimates of the gradient at the source position. Similar to the source-position interface speed estimates, all three methods have similar levels of bias, though the mode amplitude and combination methods provide better estimates of this parameter over the last kilometer. Additionally, the combination method does significantly better than the other two methods at tracking the features of the parameter in range. The maxima and minima of the parameter are clearly visible in the combination method estimate, though the values of the estimate are slightly off.

Overall, while the differences between the various methods were not drastic, the combination method tended to do slightly better than the other two methods. Its results tended to resemble those of whichever method produced the better estimate of each parameter. As stated earlier, which method

is best will depend on the quality of the estimates of the input data and the parametrization of the bottom selected. Since the combination method tends to track the better of the two estimates in each case, it can be expected to provide results similar to the better method, whichever method that happens to be in a give case.

VI. CONCLUSIONS

In this paper we have derived the results necessary to carry out a mode amplitude perturbative inversion method for determining the geoacoustic properties of the seabed. We have applied this algorithm to two data sets: One a simple synthetic data case with a known solution, and the other a LWAD 99-1 experimental data set, where a previous estimate of the bottom sound speed profile was available. In both cases the algorithm was able to successfully estimate the bottom geoacoustic parameters. The modal amplitude inversion algorithm is analogous to the modal eigenvalue inversion algorithm described in Ref. 1, the major difference being the type of input data that is used. As a consequence, it is natural to compare the two methods and to combine them so as to gain the benefits of each. Using the stochastic inverse matrix approach, we demonstrated the effectiveness of this hybrid inversion algorithm.

ACKNOWLEDGMENTS

Funding for the research presented here was provided by the Office of Naval Research, and the WHOI Academic Programs Office.

[1]S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).
[2]T. B. Neilsen and E. K. Westwood, "Extraction of acoustic normal mode depth functions using vertical line array data," J. Acoust. Soc. Am. **111**, 748–756 (2002).
[3]P. Hursky, W. S. Hodgkiss, and W. A. Kuperman, "Matched field processing with data-derived modes," J. Acoust. Soc. Am. **109**, 1355–1366 (2001).
[4]G. V. Frisk, K. M. Becker, and J. A. Doutt, "Modal mapping in shallow water using synthetic aperture horizontal arrays," OCEANS 2000 MTS/IEEE Conference and Exhibition, 11–14 September 2000, Providence, RI, Vol. 1, pp. 185–188.
[5]C. T. Tindle, L. M. O'Driscoll, and C. J. Higham, "Coupled mode perturbation theory of range dependence," J. Acoust. Soc. Am. **108**, 76–83 (2000).
[6]G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (Prentice Hall, Englewood Cliffs, NJ, 1994).
[7]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994).
[8]C. L. Pekeris, "Theory of propagation of explosive sound in shallow water," Mem.-Geol. Soc. Am. **27**, 1–117 (1948).
[9]S. D. Rajan, G. V. Frisk, and J. F. Lynch, "On the determination of modal attenuation coefficients and compressional wave attenuation profiles in a range-dependent environment in Nantucket sound," IEEE J. Ocean. Eng. **17**, 118–128 (1992).
[10]A. D. Pierce, "Extension of the method of normal modes to sound propagation in an almost-stratified medium," J. Acoust. Soc. Am. **37**, 19–27 (1965).
[11]E. H. Moore, "On the reciprocal of the general algebraic matrix," Bull. Am. Math. Soc. **26**, 394–395 (1920).
[12]R. Penrose, "A generalized inverse for matrices," Proc. Cambridge Philos. Soc. **51**, 406–413 (1955).
[13]R. Penrose, "On best approximate solution of linear matrix equations," Proc. Cambridge Philos. Soc. **52**, 17–19 (1956).
[14]D. C. Clay, *Linear Algebra and its Applications* (Addison-Wesley, New York, 1994).
[15]G. Backus and F. Gilbert, "Uniqueness in the inversion of inaccurate gross earth data," Proc. R. Soc. London, Ser. A **266**, 187–269 (1970).
[16]M. B. Porter, "The KRAKEN normal mode program," Technical Report, SACLANT Undersea Research Center, La Spezia, Italy, 1991.
[17]A. O. Williams, Jr., "Hidden depths: Acceptable ignorance about ocean bottoms," J. Acoust. Soc. Am. **59**, 1175-1179 (1976).
[18]B. R. Gomes, J. K. Fulford, and R. Nero, "Preliminary pre-experiment environmental characterization for the Littoral Warfare Advanced Development (LWAD) 99-1 SE Gulf of Mexico," Naval Research Laboratory, Stennis Space Center, MS, 1998.
[19]T. L. Poole, "Geoacoustic inversion by mode amplitude perturbation," Ph.D. thesis, MIT/WHOI Joint Graduate Program in Oceanography/Oceanographic Engineering, Cambridge and Woods Hole, MA, 2007.
[20]J. N. Franklin, "Well-posed stochastic extensions of ill-posed linear problems," J. Math. Anal. Appl. **31**, 682–716 (1970).
[21]L. L. Souza, "Inversion for subbottom sound velocity profiles in the deep and shallow ocean," Ph.D. thesis, MIT/WHOI Joint Graduate Program in Oceanography/Oceanographic Engineering, Cambridge and Woods Hole, MA, 2005.

Robustness and constraints of ambient noise inversion

Juan I. Arvelo, Jr.[a)]
The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Rd., Laurel, Maryland 20723

(Received 1 September 2007; revised 4 December 2007; accepted 4 December 2007)

One of the most dominant sources of error in the estimation of sonar performance in shallow water is the geoacoustic description of the sea floor. As reviewed in this paper, various investigators have studied the possible use of ambient noise to infer some key parameters such as the critical angle, geoacoustic properties, or bottom loss. A simple measurement approach to infer the bottom loss from ambient noise measurement on a vertical line array (VLA) is very attractive from environmental and operational perspectives. This paper presents a sensitivity study conducted with simulations and measurements that demonstrates mitigating factors to maximize the accuracy of estimated bottom loss. This paper quantifies the robustness and operational constraints of this measurement approach using an ambient noise model that accounts for wind, shipping, and thermal noise. Also demonstrated are the effects of unaccounted water absorption, array tilt, nearby ship interference, flow noise, calibration error, and array deformation on sonar performance estimation. VLA measurements collected during the Asian Seas International Acoustics Experiment in May-June 2001 were also processed to validate the approach via comparisons with measured bottom loss and transmission loss. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2828205]

PACS number(s): 43.30.Pc, 43.30.Nb [AIT] Pages: 679–686

I. INTRODUCTION

Simulating the propagation of undersea sound in shallow waters requires some form of description of the sea floor and sub-bottom sediments. The research community usually describes the bottom sediments by geoacoustic parameters consisting of the density and compressional/shear sound speeds and attenuation coefficients. However, direct measurement of these parameters is difficult, and geoacoustic inversion is a nontrivial process riddled with assumptions and uncertainties. Developing a simple, indirect measurement that does not require a sound propagation model may prove to be very useful for many research and operational applications.

Buckingham and Jones[1] formulated a possible means to estimate the bottom's critical angle from the measured noise distribution across a vertical array. The critical angle, which is directly linked to the compressional sound speed of the ocean floor, can be used to infer the bottom type. The simplicity of the measurement is the key benefit of this critical angle estimation approach. However, complex bottom structures with sound speed gradients and discontinuities increase the uncertainty in the estimated critical angle. Carbone *et al.*[2] included the measured vertical coherence across frequency to further estimate the sediment shear speed. However, they made use of a simplistic model of the sub-bottom sediments to curve fit the measured data. A straightforward and robust measurement of a bottom descriptor is needed that is not based on any simplification or assumption.

Some propagation models are not strict in their requirement of bottom description. Most ray-based models, for example, could easily use bottom loss as a function of frequency and grazing angles. Harrison and Simons[3] demonstrated the possibility of estimating the bottom loss across frequency and grazing angles by beamforming a vertical array and subtracting the up and down beam noise levels. What is missing in the literature is a sensitivity study to yield operational constraints of the measurement approach and to assess its robustness under harsh environmental conditions. This paper is aimed to conduct a sensitivity study of this approach that addresses multiple sources of error and interference. It also validates the estimated bottom loss and transmission loss against direct measurements collected during the Asian Seas International Acoustics Experiment (ASIAEX) in May-June 2001.

Section II describes the theory supporting this noise inversion approach; Sec. III presents the model-based sensitivity study; and the final section compares measured transmission losses with those from a ray-based model with ambient noise inverted bottom loss as a validation tool.

II. THEORY

Assuming an ocean waveguide with uniform sound speed, the closed-form solution for the spatial coherence between two point receivers due to wind-driven ambient noise, derived by Harrison,[4] may be generalized as

$$\rho(d,\gamma) = 2\pi \int_0^{\pi/2} \times \frac{e^{ikd \sin\phi_r \sin\gamma e^{-\alpha s_p}} + R_b e^{ikd \sin\phi_r \sin\gamma e^{-\alpha(s_c - s_p)}}}{[1 - R_s R_b e^{-\alpha s_p}]} \times J_o(kd\cos\phi_r \cos\gamma) S(\phi_s) \cos\phi_r d\phi_r, \quad (1)$$

where the elevation angles from the surface and bottom to the receiver are given by ϕ_s, ϕ_b, ϕ_r, respectively; γ represents the orientation between the two point receivers as

[a)]Electronic mail: juan.arvelo@jhuapl.edu

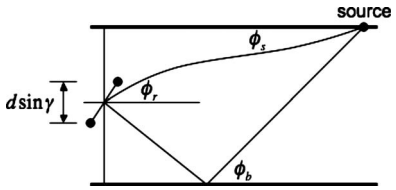


FIG. 1. Illustration defining the scenario variables used to develop an analytic model of the noise coherence and vertical anisotropy.

shown in Fig. 1; α is the absorption in the water column; S is the surface source function; d is the element spacing; and R_s and R_b are the surface and bottom reflection coefficients, respectively. Because sea conditions near the receiver dominate wind-generated noise, straight-line propagation yields the following estimated partial and complete cycle distances:

$$s_p = z/\sin((\phi_r + \phi_s)/2) \tag{2}$$

and

$$s_c = (2H - z)/\sin((\phi_b + \phi_s)/2), \tag{3}$$

where z and H are the receiver and bottom depths, respectively. These formulas differ slightly from those in Harrison,[4] because it is important to account for the receiver's depth in a shallow water column. The surface source function ranges from a dipole pattern for low wind speeds or frequencies to a monopole pattern at high wind speeds or frequencies.[5,6]

The spatial coherence becomes the omnidirectional ambient noise power when the element spacing is set to zero, appearing as

$$\rho(d=0) = 2\pi \int_0^{\pi/2} \frac{1 + R_b e^{-\alpha(s_c - s_p)}}{[1 - R_s R_b e^{-\alpha s_p}]} S(\phi_s)\cos\phi_r d\phi_r, \tag{4}$$

where the integrand represents the ambient noise vertical distribution. This integrand is the summation of two components. The left component is the noise distribution for elevations towards the surface. The right component corresponds to elevations towards the seafloor. A division of these two components is equivalent to the subtraction of the beam noise levels. Therefore, the difference between the vertical beam noise levels becomes

$$\Delta VN = 10\log(R_b e^{-\alpha(s_c - s_p)}), \tag{5}$$

which indicates two conditions that must be satisfied for this value to be equivalent to the bottom loss. The first condition is that the water absorption must be of a relatively negligible amount, which may be satisfied at lower frequencies. The second condition is that at higher frequencies, it is also possible to infer the bottom loss if the receiver is close to the ocean floor. By setting $z=H$ in Eqs. (2) and (3), note that s_c-s_p=0. However, even if these conditions are not satisfied, it may still be possible to account for the correct water absorption to estimate the bottom reflection coefficient.

Note that this formulation does not account for the beamformer's response pattern. Instead, Eq. (5) is the result of a rectangle function (i.e., cookie cutter) vertical response. However, the actual beam response is expected mainly to smooth the inferred bottom loss. A numerical solution can account for the beamformer's effect, including possible energy leakage through the sidelobes. A ray-based passive sonar model[7] was developed to predict the vertical wind-driven noise distribution and compute the cross-spectral density matrices of the source and individual ships in a random population realization, accounting for the receiver's two-dimensional (in azimuth and elevation) conventional or adaptive beam response.

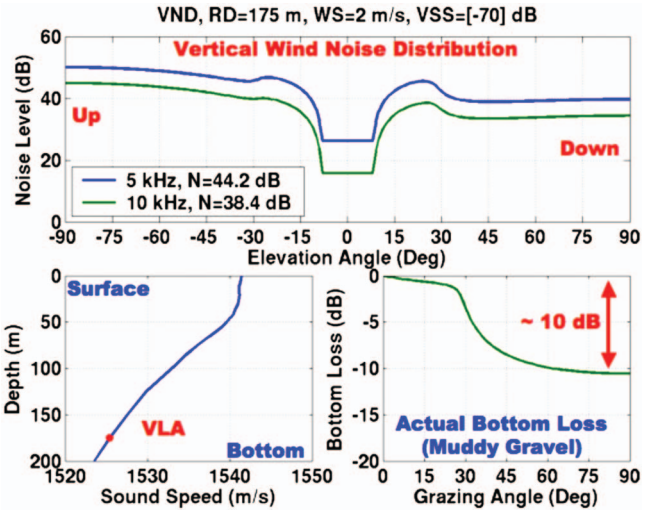


FIG. 2. Canonical East China Sea summer environment and scenario for simulation-based sensitivity analysis of ambient noise inverted bottom loss.

The formulation in this section assumes an isospeed water column, which is adequate for illustration purposes only. The passive sonar model already accounts for refraction and bathymetric variability. It was recently enhanced to predict the array and beamformer's sensitivity to various sources of errors. These errors include the effect of not compensating for the water absorption in Eq. (5), array tilt, nearby discrete ship interference, electronic or flow noise, array shading, calibration errors, and errors due to array shape distortion. In Sec. III, this model is implemented to conduct a sensitivity study of this noise inversion approach.

III. SIMULATION

Because ASIAEX 2001 was conducted in the East China Sea, a representative environmental condition is being used to conduct the model-based sensitivity analysis. The two bottom panels in Fig. 2 show the selected sound speed profile and reflection loss over grazing angles for a muddy-gravel bottom type at 200 m depth. The top panel is the estimated vertical distribution of the wind noise from the integrand of Eq. (4) for the surface function in Ref. 6, wind speed of 2 m/s, and a 175 m receiver depth. In this case, we will assume that the noise notch about the horizontal is filled by forward-scattered energy from volume inhomogeneities uniformly distributed across the entire water column. The model and formulation to account for volume scattering are explained in detail in Ferat and Arvelo.[7] The forward-scattered volume scattering strength is arbitrarily set to −70 dB. For simplicity and without loss of generality, we will assume that the ocean waveguide is flat. This noise distribution is convolved with the array's beam response to yield the estimated

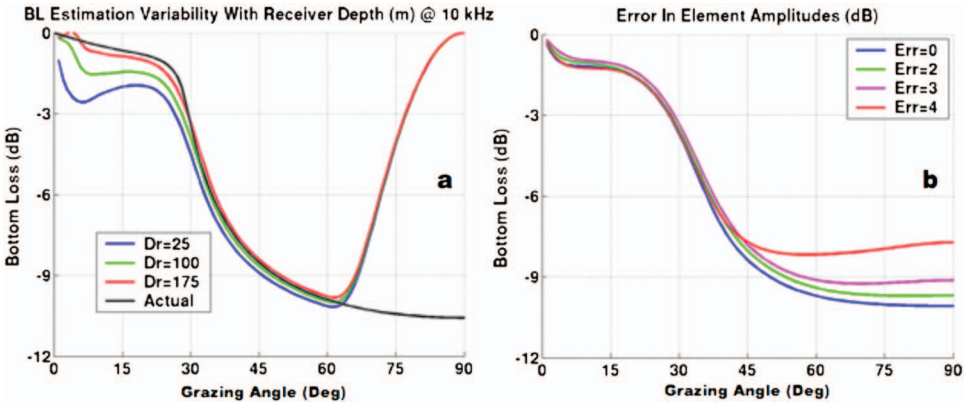


FIG. 3. Ambient noise inverted bottom loss at the array's design frequency (10 kHz) showing (a) the effect of water absorption, beamformer's grating lobes, and (b) array calibration errors below the array's design frequency (5 kHz).

beam noise level. Finally, the up and down beam noise levels are subtracted and compared with the actual bottom loss in Fig. 2 for a selection of error values.

The synthetic vertical array consists of 20 omnidirectional hydrophones uniformly spaced to a design frequency of 10 kHz, which would correspond to an element spacing of 7.5 cm or an array length of 1.425 m. To mitigate energy leakage through the sidelobes of the array's beam response in such an anisotropic noise field, the Dolph–Chebyshev shading is applied to reduce the peak sidelobe levels to 30 dB below the maximum response axis.

At the design frequency, Fig. 3(a) shows the predicted bottom loss for three array depths and water absorption from Thorp's empirical formula. The water absorption at 10 kHz is large enough to produce some error in the estimated bottom loss if the correction in Eq. (5) is not made. As expected, the correction is not necessary if the array is located near the bottom. The divergence to zero bottom loss for grazing angles above 60° is attributed to the grating lobe introduced by the array when beamformed at its design frequency. As the array is beamformed towards either end-fire directions, its grating lobe will be pointed in the opposite direction, causing the estimated bottom loss to be zero. To avoid the grating lobe, the rest of the calculations in this section are safely conducted at a frequency of 5 kHz.

Random amplitude errors with a number of standard deviations have been added to the actual element amplitudes, which would emulate a calibration error. Figure 3(b) shows the effect of calibration error on the estimated bottom loss. To keep the bottom loss error below 1 dB, the amplitude error must be kept below 3 dB. It appears that the subtraction operation to infer the bottom loss is canceling some of the calibration error. A 3 dB calibration error is not considered a very demanding constraint because current calibration errors tend to fall below 1 dB. Sensitivity to phase errors may also be conducted. However, these are equivalent to array element location or time delay errors. Therefore, only array element location errors are explicitly covered here. Time delay and phase errors may be inferred from the array distortion results.

Errors in array element locations are known to affect a beamformer's performance severely. To simulate array deformation effects, the elements of the array are shifted in all directions by a realization of the uncertainty with a normal distribution of a range of standard deviations. Figure 4(a) shows the estimated bottom loss for a handful of standard deviations of the error in element locations. If a maximum bottom loss error is set to 1 dB, the array elements must be known within about 2 cm. Because the acoustic wavelength at 5 kHz is 30 cm, the constraint may be generalized to an error below one-fifteenth of an acoustic wavelength. This important constraint is applied in Sec. IV, where measurements are processed for validation purposes. The equivalent phase and time-delay sensitivities are $\Delta\varphi = 2\pi/15$ and $\Delta\tau = 1/15f$, respectively.

The array tilt does not appear to degrade the bottom loss estimation as significantly as originally predicted in Harrison and Simons.[3] Figure 4(b) shows the influence of array tilt on the predicted bottom loss at 5 kHz, which is safely below the array's design frequency. The general effect of the array tilt is additional smoothing of the estimated bottom loss curve; this is attributed to the conical shape of the beam pattern.

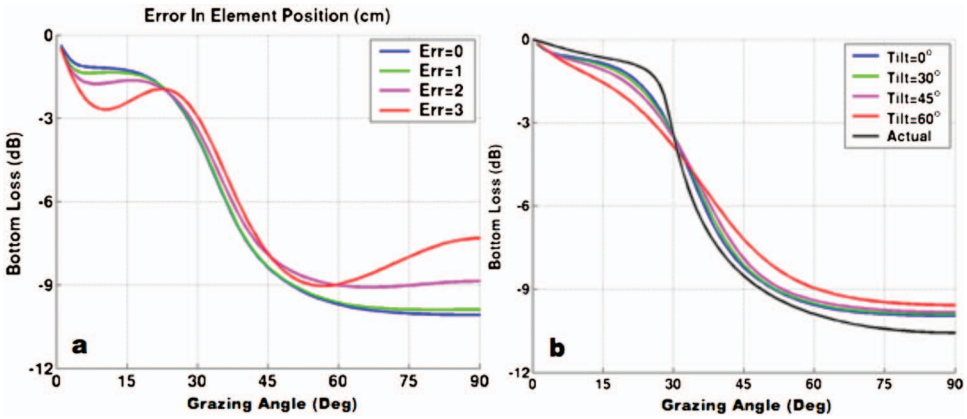


FIG. 4. Sensitivity of ambient noise inverted bottom loss to array element location error (a) and array tilt (b).

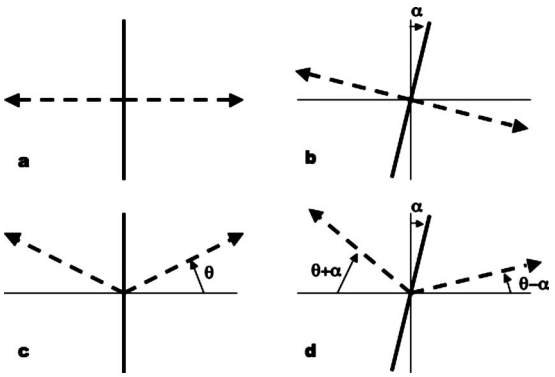


FIG. 5. Illustration of vertical array (solid) and maximum response axis (dash) when the array is straight and tilted. Due to the conical shape of the beam response, the array that is tilted by a small angle α, and steered at an elevation θ, would yield an average beam noise level between $\theta-\alpha$ and $\theta+\alpha$ resulting in a smoother bottom loss curve.

When the array is straight, all azimuthal directions of the maximum-response beam are steered at the same elevation θ, as shown in Figs. 5(a) and 5(c). When tilted by an angle α, the cone covers a span of elevations from $\theta-\alpha$ to $\theta+\alpha$, as shown in Fig. 5(d), which would result in the smoothing of the inferred bottom loss curve. If steered at broadside, the tilted array covers positive and negative angles, as shown in Fig. 5(b), which would result in zero bottom loss. However, because the actual bottom loss at zero grazing angle should always be zero, the tilt does not adversely affect the estimated value. The array's beamwidth at 5 kHz also contributes to this robustness to array tilt. The Dolph–Chebyshev weighted beamwidth is 12° across the −3 dB points and is 30° across the −30 dB points of the peak sidelobe level. If increased robustness is needed for harsher environmental conditions, the array elements may be replaced with vector or gradient sensors. Such directional elements may be steered in the direction of the tilt plane to reduce its smoothing effect.

Usually, the main cause of azimuthal noise anisotropy is the discrete shipping population. Particularly, nearby and loud ships can significantly degrade the accuracy of the estimated bottom loss. To illustrate this effect, a random shipping population realization was generated using East China Sea population densities for fishing boats, merchant ships, and tankers out to 100 km from the vertical array. Due to the high frequency, the low source level and high transmission loss of distant ships does not allow them to significantly affect the noise field, with the exception of one relatively loud ship at 4 km. To qualitatively assess the impact of nearby and distant shipping, ships are deleted when they are located at ranges shorter than a specified minimum range. If this minimum range is zero, all ships in the population are included in the calculations. Figure 6(a) displays the estimated bottom loss for three values of this minimum range. The energy from the nearby ship leaks through the beamformer's sidelobes regardless of their low peak levels, causing the subtraction of the beam noise levels to skew towards zero bottom loss. Distant shipping noise mainly dominates at elevations near the horizontal and hence affects the estimated bottom loss near zero grazing angles.

The effect of shipping noise is expected to be worse due to higher source level and lower transmission loss. However, higher frequencies are not immune to interference effects. As the ship and wind-driven noise decrease with increasing frequency, thermal, electronic, and flow noise become increasingly dominating isotropic and incoherent noise sources that contain no useful environmental information. Figure 6(b) displays estimated bottom-loss curves with various levels of added incoherent noise. Because the omnidirectional wind-driven noise level is about 44 dB, an incoherent noise level of 45 dB obviously yields grossly erroneous bottom loss estimation. All sources of incoherent noise must be below about 10 dB from the omnidirectional wind-driven noise level to ensure reasonable estimation of the bottom loss.

This concludes the simulation-based sensitivity study of noise inverted bottom loss estimation. In the real world, all these sources of error occur simultaneously, causing occasional instances of significant cumulative errors. In Sec. IV, vertical line array (VLA) data are processed to address ways to mitigate these errors, including long integration times and intensity averaging snapshots of bottom loss estimations.

IV. MEASUREMENT

Under the auspices of the Office of Naval Research, the East China Sea boundary interaction study was conducted as part of the ASIAEX 2001. The environment, defined by a box from 28°N to 30°N and 126.5°E to 128°E, crosses the continental slope and is characterized by (1) ocean currents across depths that reach 0.8 m/s, (2) bathymetric variability

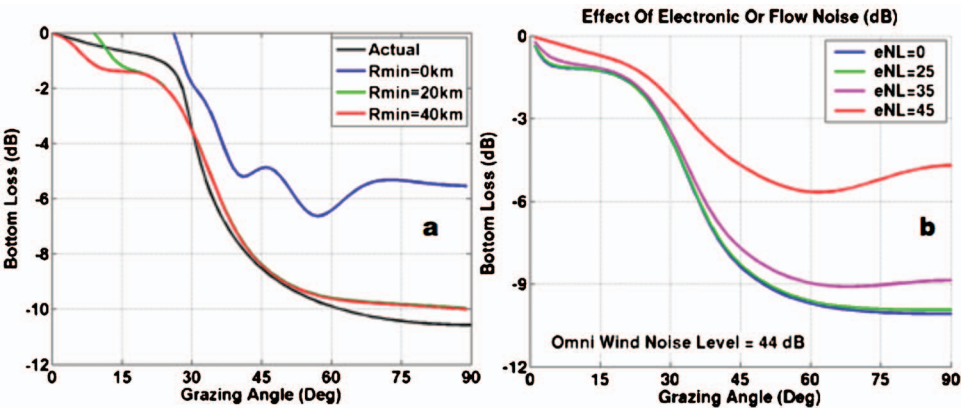


FIG. 6. Effect of near and distant shipping noise (a) and electronic or flow noise (b) on the accuracy of the estimated bottom loss.

TABLE I. Geoacoustic model of sub-bottom structure in the East China Sea.

Sediment layer	Thickness (m)	Constant density (g/cc)	Attenuation coefficient (dB/m kHz)	Frequency exponent	Sound speed (m/s)	Critical angle (deg)
Top	6.7	1.75	0.56	2.0	1605.0 (top)	18.7
					1621.5 (bottom)	20.4
Deeper	17.0	1.6	0.27	1.0	1811.8 (top)	33.0
					1864.8 (bottom)	35.4
Half-space Basement	Semi-infinite	2.3	0.4	2.1	2400	50.7

from 100 m at the shelf to 2 km at the basin, and (3) strong internal waves with temperature fluctuations of up to 3 °C.[8]

The environment had a 30 m mixed layer and a bottom depth of about 105 m at 29° 40.67' N and 126° 49.39' E. One of the measurements collected made use of a 31-element VLA with 21.43 cm spacing. The lowest hydrophone of the array was 7.5 m above the sea floor. The data were sampled at 12 kHz after a low-pass filter at 5 kHz. The array was deployed from the Research Vessel (R/V) *Melville* and had a 3.5 kHz transducer mounted below the array for bottom scattering measurements.

Rouseff and Tang[9] used time snippets between the pings to conduct measurements of ambient noise vertical anisotropy. They integrated over 140 consecutive snapshots—each of which was 0.5 s long—to obtain a smooth ambient noise vertical structure across frequency. The resulting vertical noise distribution exhibited two peaks at elevations near $\pm 10°$ with a notch of low noise towards the horizontal. Rouseff and Tang hypothesized that the noise level inside this notch may be attributed to sound propagation through a waveguide with a significant internal wave field. This hypothesis, however, must still be proven by ruling out other known physical mechanisms such as array distortions due to the strong currents or volume scattering from the forward-scattered ambient noise and the back-scattered source waveforms.

Another useful measurement conducted during the same test was octave-averaged transmission loss from explosive sources at 18 m depth to a sonobuoy sensor at 27 m depth in a 119 m water column. Broadband measurements centered at 25, 50, 100, 200, 400, and 800 Hz were geoacoustically inverted by Knobles *et al.*[10] to yield a model of the sub-bottom structure consisting of two sediments over a semi-infinite half space. Table I summarizes their final model parameters. The top sediment is 6.7 m thick with a constant density of 1.75 g/cm^3, an attenuation coefficient of 0.56 dB/m kHz, a frequency exponent of 2.0, and a sound speed that varies linearly from 1605.0 m/s (critical angle=18.7°) at the top to 1621.5 m/s (critical angle=20.4°) at the bottom. The deeper sediment is 17 m thick with a density of 1.6 g/cm^3, an attenuation of 0.27 dB/m kHz, a frequency exponent of 1.0, and a sound speed that increases from 1811.8 m/s (critical angle=33°) to 1864.8 m/s (critical angle=35.4°). The half-space basement has a density of 2.3 g/cm^3, attenuation coefficient of 0.4 dB/m kHz, a frequency exponent of 2.1, and a uniform sound speed of 2400 m/s (critical angle=50.7°).

The critical angles were computed for a water sound speed of 1520 m/s. All these bottom sediments are assumed to have negligible elasticity. This bottom model was shown to yield an element-level time series that satisfactorily matches the measurements at multiple receiver depths and ranges.

This geoacoustic model of the stratified seafloor is useful in the computation of the equivalent bottom loss, which may serve to validate noise-inverted estimations. To this end, the ASIAEX 2001 data used in Ref. 9 were beamformed to infer the bottom loss at 200, 400, and 800 Hz. Because the array is 6.4 m long, 200 Hz stresses the lower limit, as the array would be around one wavelength long and the beamwidth about 60° wide after Dolph–Chebyshev shading to reduce sidelobe peaks to 20 dB below the maximum response axis. This shading was implemented to mitigate the possible leakage from the noise arriving from above the array when beamforming towards lower elevations. However, further sidelobe reduction was avoided because it would also cause the negative side effect of a wider main lobe. This beamwidth reduces to about 30° at 400 Hz and 15° at 800 Hz. As indicated before, this beamwidth controls the smoothness of the estimated bottom loss and inversion sensitivity to array tilt and distortions.

Unfortunately, measurements of array tilt and shape were not collected during this event despite the strong ocean currents. However, the simulations in the previous section indicate that a maximum random error in element location of less than one-fifteenth of an acoustic wavelength is required to keep bottom loss errors below 1 dB. This corresponds to an error of 50 cm at 200 Hz, 25 cm at 400 Hz, and 12.5 cm at 800 Hz. The demanding constraint in element location decreases the confidence level with increasing frequency. Therefore, the 800 Hz results may be considered the high-end limit for ambient noise inversion of this particular dataset.

It is also important to appreciate the role of the integration time on the bottom loss estimation. The minimum integration time may be estimated by the array length and number of elements. The minimum snapshot duration is the array

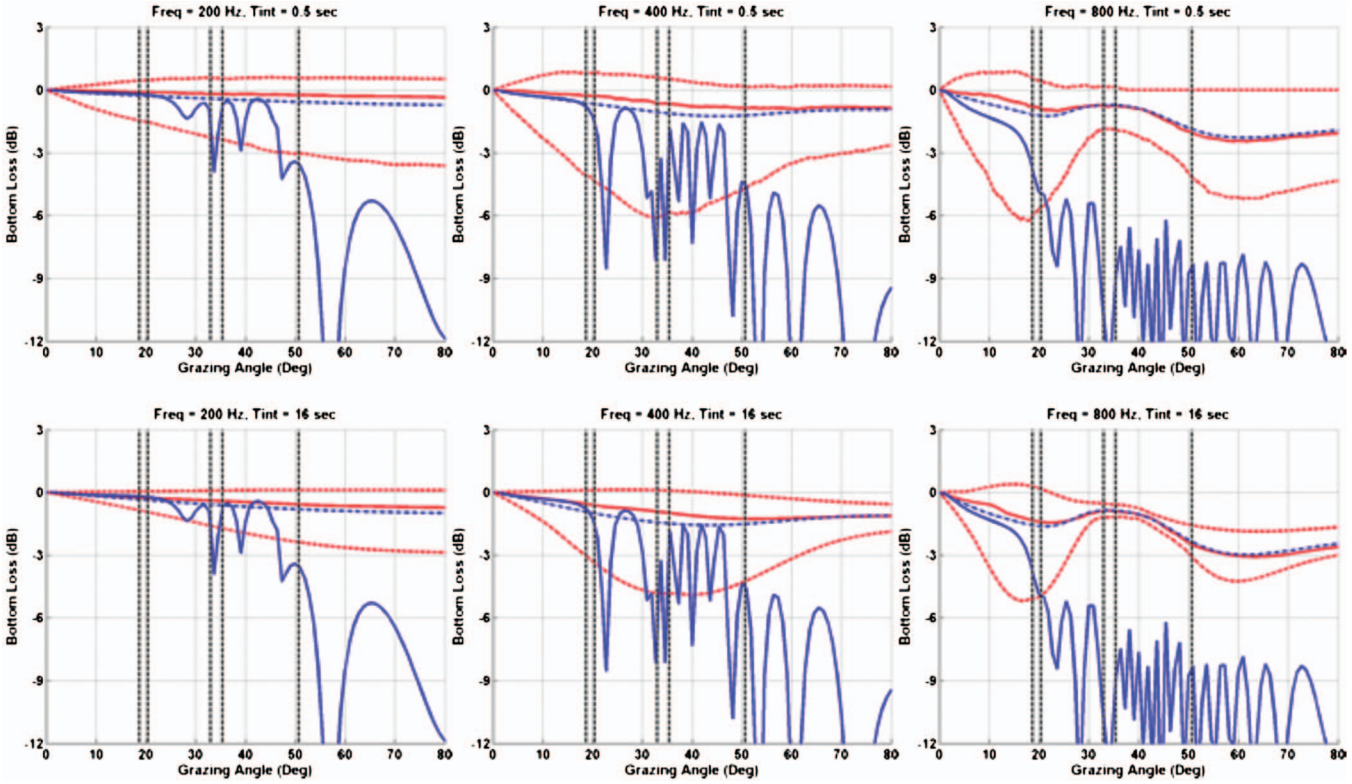


FIG. 7. Estimated bottom loss at three frequencies and two integration times with geoacoustic inversion parameters (blue solid curves), incoherent-mean noise inversion (blue dash curves), median noise inversion (red solid curves) and 10 and 90 percentiles noise inversion (red dash curves). The vertical black dash lines represent the critical angles at the sub-bottom sediment interfaces.

length (6.4 m) over the group velocity (1500 m/s). This yields a minimum of 4.3 ms. Because each array element has 3 degrees of freedom, the integration time required for full resolution is 3 degrees of freedom times the 31 elements times the minimum snapshot duration. This case would yield a minimum integration time of 0.4 s. Therefore, the bottom loss was estimated for integration times ranging from 0.5 to 16 s to observe the effect of longer integration times. It is also worth mentioning that, at frequencies much lower than the array's design frequency, the actual number of degrees of freedom is smaller than the number of elements, and the minimum integration time is actually much less than 0.4 s. For frequencies below 800 Hz, this value is closer to 0.1 s.

The 10, 50, and 90 percentile curves of multiple bottom loss estimations were computed to observe the variability and influence of the integration time. The incoherent-mean bottom loss[1] was also calculated to compare with the median bottom loss curve. For "ground truth" bottom loss curves, a wave number integration algorithm (OASR)[11] was used with the geoacoustic inversion parameters from the Knobles *et al.* study[10] at the three frequencies. The bottom loss results are displayed as the blue solid curves in all the panels of Fig. 7. This figure displays bottom loss curves for two integration times and three frequencies as indicated above each panel. The solid red curves represent the median of all the noise-inverted bottom loss curves, and the dashed red curves are the 10 and 90 percentiles. Finally, the dashed blue curves next to the solid red curves are the incoherent-mean values, and the vertical black dash lines represent the critical angles previously computed for each sediment sound speed limit. The lowest critical angles are for the topmost sediment and the highest critical angle is for the semi-infinite basement. At the highest frequency, the reflected energy is most affected by the uppermost sediment, while the influence of the deeper sediments increases with decreasing frequency.

At first inspection, it is obvious that none of the noise-inverted curves match the ground truth at most grazing angles. Matching all of the details of the blue curves is not expected due to the width of the beamformed main lobe, which causes the curves to be much smoother. Mismatch at the steeper angles is worse than below the critical angle. However, the bottom loss below the critical angle is much more important for long-range propagation predictions. Knobles *et al.*[10] conducted an excellent geoacoustic inversion that matched the measured transmission loss at a wide frequency range. However, because the geoacoustic inversion makes use of long-range propagation to infer the bottom parameters, the resulting bottom loss at angles above critical have a much lower confidence level. The geoacoustic inversion is expected to yield closer bottom loss to the actual ground truth at smaller grazing angles. On the other hand, the ambient noise inverted bottom loss at steep grazing angles is driven by surface noise sources that are much closer in range. In summary, the ambient noise inverted bottom loss curves are closer to the geoacoustic inverted ground truth, where it matters most—below the critical angles.

This critical angle varies across the three frequencies because low-frequency energy penetrates deeper into the seafloor sediments where the sound speed is larger. Below the

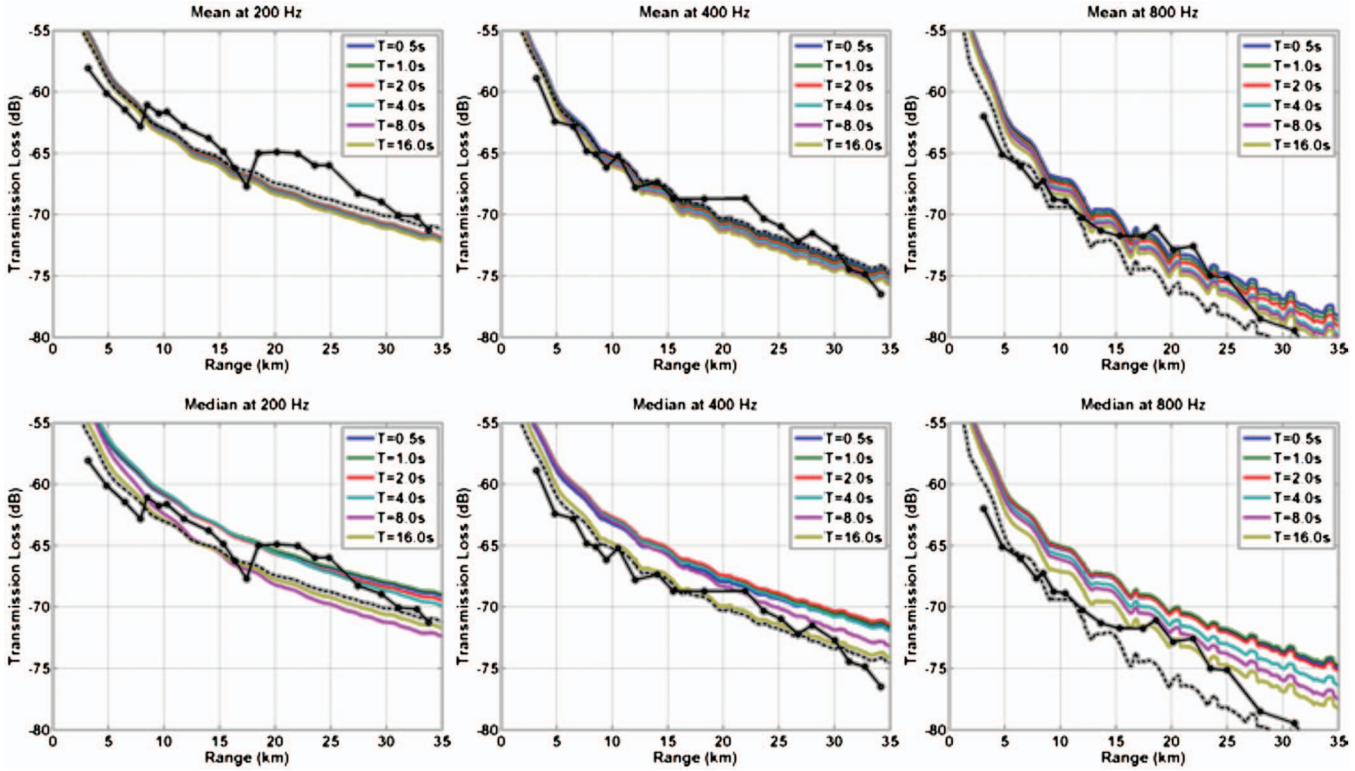


FIG. 8. Comparisons of modeled transmission loss at three frequencies using the incoherent-mean noise inverted bottom loss (top) and median bottom loss (bottom) against measurements (black solid curves) and modeled transmission loss with predicted bottom loss from geoacoustic inversion parameters (black dashed curves). In general, the inferred transmission loss is directly proportional to the integration time.

critical angles, the median and mean bottom loss curves surround the ground truth at the lower frequencies. However, the incoherent-mean curves are closer to the black curves for either integration time at the higher frequency.

As expected, the longer integration time narrows the spread of noise-inverted bottom loss curves at these three frequencies. A point of curiosity with the spread is how wide it becomes near the dominant critical angle for each frequency. Any relationship with the critical angle may be coincidental, but there is also a possibility that the large variability in bottom loss from below to above the critical angle across the beam's resolution cell causes this spread when the ocean currents distort the array. Array shape distortions cause sidelobe levels to increase, despite the shading that is applied, and the main lobe to widen and move around. During some instances, there may be a large sidelobe at steeper angles, which would pull the bottom loss curve to incorrectly register a higher or lower loss. These results indicate that long integration times and intensity averaging the predicted bottom loss curves could overcome some of the fluctuations caused by array distortions and uncertainties.

Of most interest is how well these bottom loss curves predict the transmission loss. Ground truth transmission loss measurements are available at these frequencies in Ref. 10. The transmission loss values were extracted from this paper; curves at the relevant frequencies are plotted as the black solid curves in Fig. 8. The measured sound speed profile in this paper was used to conduct the predicted transmission loss curves. The median and mean bottom loss curves in Fig. 7 were converted into a table to be imported into a Gaussian ray bundle (GRAB) shallow water sound propagation model by Keenan and Weinberg.[12] The incoherent transmission loss was computed limiting the number of eigenrays to 100 surface and bottom bounces with an eigenray tolerance of 0.01, which is more than needed, but these values were used to ensure accuracy in the solutions. The source depth was set to 18 m, the receiver at 27 m, and the bottom at 119 m. Water column attenuation was also included, even though this effect is minimal at these frequencies and for ranges shorter than 35 km.

The dashed black curves in Fig. 8 are the calculated transmission loss for the computed bottom loss with the Knobles *et al.*[10] geoacoustic inversion bottom parameters. The top three plots in Fig. 8 contain the transmission loss using the noise-inverted incoherent-mean bottom loss for six integration times from 0.5 to 16 s. The bottom plots are the respective curves with the median bottom loss curves in Fig. 7. Both black curves in the top and bottom plots are identical for the respective frequencies. In general, the transmission loss is directly proportional to the integration time.

The closeness of the modeled transmission loss curves to the measurements further confirms the previous statement that the bottom loss below the critical angle is more important than that at steeper angles in Fig. 7. Even though none of the modeled curves in Fig. 8 follow every detail of the measurements, they are very close in magnitude, and in their range and frequency trends. Although the bottom loss below the critical angles is very close to zero, a separate calculation with zero bottom loss at all grazing angles showed significant discrepancy with the measurements, particularly at the

higher frequencies. Therefore, this simple noise inversion method was able to predict the transmission loss with similar accuracy to the more computationally demanding and difficult geoacoustic inversion in this environmental condition.

In all cases, it appears that the mean and median transmission loss curves approach the geoacoustic-inverted curves with increasing integration time. However, the curves for the incoherent-mean bottom loss values are closer to the measurements and do not require a longer integration time as the median curves do. For a VLA tilted and distorted by ocean currents, its robustness for predicting the bottom loss using ambient noise is improved—even at frequencies far below the array's design frequency—by intensity averaging and longer integration time.

V. SUMMARY

This paper quantifies the robustness and constraints of ambient noise inverted bottom loss via simulations using processed data collected during ASIAEX 2001. Errors due to water absorption, array distortions, and noise were simulated to quantify the constraints. Ambient noise inversion was found to be particularly robust to array tilt. However, array element locations must be known to be within one-fifteenth of an acoustic wavelength. Electronic and flow noise must be more than 10 dB below the omnidirectional ambient noise level across the array. Nearby loud ship interference is particularly damaging to the accuracy of the bottom loss estimation.

Element-level data from a vertical array were beamformed to infer the bottom loss in the East China Sea. Bottom parameters from a geoacoustic inversion were used to compute the ground truth bottom loss. The ambient noise inverted bottom loss was found to vary significantly across time snapshots, indicating a need for relatively long integration times. In addition, it was observed that the inverted intensity-averaged bottom loss requires a shorter integration time and matches the ground truth better than the median curves at grazing angles below the critical angle. Discrepancies between the bottom loss from the geoacoustic inversion and the noise inversion were found at steeper grazing angles. However, steep grazing angles' contributions to long-range propagation were shown to be negligible. Noise inverted intensity-averaged bottom loss was shown to match the measured transmission loss very well down to very low frequencies. Even in the presence of serious array shape uncertainties, it is still possible to infer the bottom loss at frequencies as low as the sound speed over the array length.

ACKNOWLEDGMENTS

The author is very grateful for the support of the Undersea Warfare Business Area (UWBA) Independent Research & Development (IRAD) Board of JHU/APL, and to Dan Rouseff and Dajun Tang for furnishing the ASIAEX 2001 data. A special thanks is also extended to the reviewers for their constructive feedback to help significantly improve this paper.

[1]The incoherent mean is the average intensity ratio and the coherent mean is the average pressure ratio.

[1]M. J. Buckingham and S. A. S. Jones, "A new shallow-ocean technique for determining the critical angle of the seabed from the vertical directionality of the ambient noise in the water column," J. Acoust. Soc. Am. **81**(4), 938–946 (1987).
[2]N. M. Carbone, G. B. Deane, and M. J. Buckingham, "Estimating the compressional and shear wave speeds of a shallow water seabed from the vertical coherence of ambient noise in the water column," J. Acoust. Soc. Am. **103**(2), 801–813 (1998).
[3]C. H. Harrison and D. G. Simons, "Geoacoustic inversion of ambient noise: A simple method," J. Acoust. Soc. Am. **112**(4), 1377–1389 (2002).
[4]C. H. Harrison, "Formulas for ambient noise level and coherence," J. Acoust. Soc. Am. **99**(4), 2055–2066 (1996).
[5]D. J. Kewley, D. G. Browning, and W. M. Carey, "Low-frequency wind-generated ambient noise source levels," J. Acoust. Soc. Am. **88**(4), 1894–1902 (1990).
[6]R. M. Kennedy and T. K. Szlyk, "Modeling high-frequency vertical directional spectra," J. Acoust. Soc. Am. **89**(2), 673–681 (1991).
[7]P. A. Ferat and J. I. Arvelo, "Mid to high-frequency ambient noise anisotropy and notch-filling mechanisms," *High-Frequency Ocean Acoustics Conference Proceedings*, La Jolla, CA, March 1–5, 2004, edited by M. Porter, M. Siderius, and W. Kuperman (AIP, Melville, New York, 2004), pp. 497–507.
[8]S. R. Ramp, C.-S. Chiu, F. L. Bahr, Y. Qi, P. Dahl, J. Miller, J. Lynch, R. Zhang, and J. Zhou, "The shelf-edge frontal structure in the central East China Sea and its impact on low-frequency acoustic propagation," IEEE J. Ocean. Eng. **29**(4), 1011–1031 (2004).
[9]D. Rouseff and D. Tang, "Internal wave effects on the ambient noise notch in the East China Sea: Model/data comparison," J. Acoust. Soc. Am. **120**(3), 1284–1294 (2006).
[10]D. P. Knobles, T. W. Yudichak, R. A. Koch, P. G. Cable, J. H. Miller, and G. R. Potty, "Inferences on seabed acoustics in the East China Sea from distributed acoustic measurements," IEEE J. Ocean. Eng. **31**(1), 129–144 (2006).
[11]H. Schmidt and F. B. Jensen, "An efficient numerical solution technique for wave propagation in horizontally stratified ocean environments," SM-173, SACLANT ASW Research Center, 1984.
[12]R. E. Keenan and H. Weinberg, "Gaussian Ray Bundle (GRAB) model shallow water acoustic workshop implementation," J. Comput. Acoust. **9**(1), 133–148 (2001).

Sounds and vibrations in the frozen Beaufort Sea during gravel island construction

Charles R. Greene, Jr.,[a] Susanna B. Blackwell, and Miles Wm. McLennan
Greeneridge Sciences, Inc., 1411 Firestone Road, Goleta, California 93117

(Received 28 June 2007; accepted 12 November 2007)

Underwater and airborne sounds and ice-borne vibrations were recorded from sea-ice near an artificial gravel island during its initial construction in the Beaufort Sea near Prudhoe Bay, Alaska. Such measurements are needed for characterizing the properties of island construction sounds to assess their possible impacts on wildlife. Recordings were made in February–May 2000 when BP Exploration (Alaska) began constructing Northstar Island about 5 km offshore, at 12 m depth. Activities recorded included ice augering, pumping sea water to flood the ice and build an ice road, a bulldozer plowing snow, a Ditchwitch cutting ice, trucks hauling gravel over an ice road to the island site, a backhoe trenching the sea bottom for a pipeline, and both vibratory and impact sheet pile driving. For all but one sound source (underwater measurements of pumping) the strongest one-third octave band was under 300 Hz. Vibratory and impact pile driving created the strongest sounds. Received levels of sound and vibration, as measured in the strongest one-third octave band for different construction activities, reached median background levels <7.5 km away for underwater sounds, <3 km away for airborne sounds, and <10 km away for in-ice vibrations.
 [DOI: 10.1121/1.2821970]

PACS number(s): 43.30.Xm, 43.50.Rq, 43.28.Hr [KGF] Pages: 687–695

I. INTRODUCTION

During winter 2000, BP Exploration (Alaska) began construction of an artificial gravel island 5 km seaward of the barrier islands northwest of Prudhoe Bay, Alaska (see Fig. 1). Water depth was 12 m, and the area was covered by landfast ice ~1.2–1.6 m thick. The island was built on the remnants of an eroded gravel island (Seal Island) built in 1982 for exploratory drilling. The new island was to become the platform for 30 wells to produce oil from the Northstar prospect. The work in early 2000 included building two ice roads, one for the gravel trucks and the second for the two pipelines to be buried in the sea bottom, as well as building the island itself. Within a considerable apron of gravel, sheet piles were driven into the gravel to contain a 120-m-square area where the 30 wells, living quarters, crude oil stabilization plants, and gas injection facilities were to be built subsequently.

To document construction noise, sound and vibration measurements were made during the period 1 February–17 May 2000 in the water, air and ice surrounding the island. The ringed seal (*Phoca hispida*) is the main species of seal present in the Prudhoe Bay area during the winter and spring. Noise measurements in 2000 (this study) and subsequently (Blackwell *et al.*, 2004a) were obtained to support studies of the seals before, during, and after construction to determine if there were any adverse effects. Those studies showed that ringed seals were not displaced (Blackwell *et al.*, 2004b; Moulton *et al.*, 2005; Williams *et al.*, 2006). In this paper, we first present basic measurements for each sound source operating during the period of initial heavy construction in early 2000. We then summarize those data relative to one another and to previous related measurements.

II. METHODS

A. Equipment

The sensors included a hydrophone, a three-axis geophone, and a microphone, all calibrated. • The *hydrophone* was an International Transducer Corporation model 6050C, an instrument with a low-noise preamplifier next to the sensor and a 30 m cable. It was calibrated at ITC when new (1985) and again at ITC in June 1999. • The omnidirectional *microphone* was an ACO model 7013 condenser microphone with a 4012 preamplifier and a spherical (omnidirectional) windscreen. It was calibrated at the factory when new (November 1999) and with a secondary calibrator ten days before the field work. • The *geophone* was a GeoSpace model 20 DM three-axis geophone with critical damping on the two horizontal axes and the vertical axis. Calibration was at the factory when new in 1999. Resonant frequency was 10 Hz and the coil resistance was 280 Ω. The frequency response was flat from 30 to 500 Hz and 3 dB lower at 10 Hz. The geophone was used to sense particle velocity in three orthogonal directions: vertical, longitudinal (H1), which was oriented towards the sound source, and transverse (H2). In this paper we only report data from the vertical channel, which often provided the highest levels.

The frequency responses of all sensors rolled off at low frequencies (<20 Hz). To compensate for the low frequency "roll-off," all recorded signals were transformed by digital filtering into signals flat over frequencies from 4 to 500 Hz for the geophone and from 4 to 10,000 Hz for the hydrophone and microphone.

[a]Author to whom correspondence should be addressed. Electronic-mail: cgreene@greeneridge.com

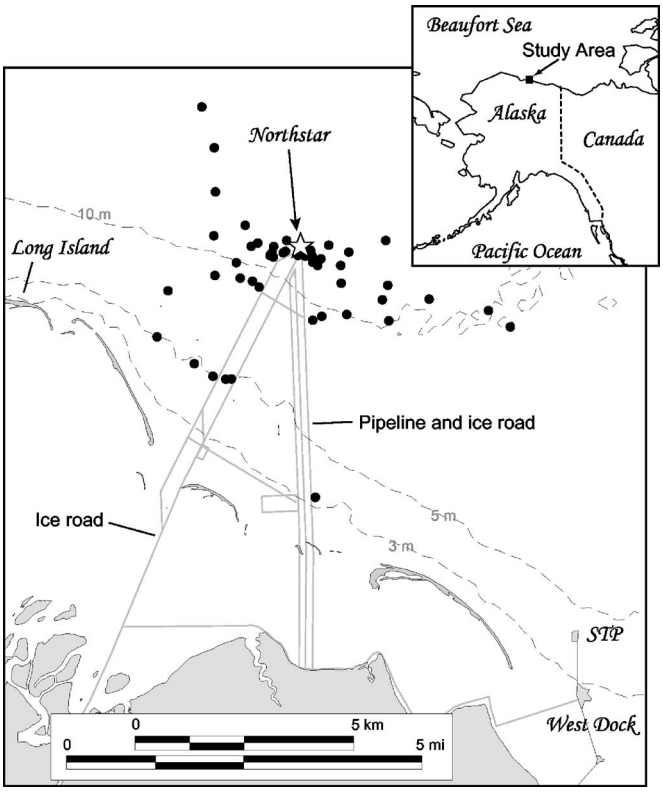


FIG. 1. Location of the study area around Northstar Island (star), Beaufort Sea, Alaska. Acoustic recordings of various sounds during construction of the island were made on 1 and 2 February, 18, 20, and 21 March, and 17 May 2000 at the locations shown by filled circles.

The hydrophone signal and the three geophone signals were amplified with adjustable gain postamplifiers. All five sensor signals were recorded on separate channels of a SONY model PC208Ax instrumentation-quality digital audio tape recorder. The recorder ran at double speed, providing a frequency response nearly flat from 4 to 10,000 Hz on each channel. Quantization was 16 bits, providing a dynamic range of about 90 dB between an overloaded signal and the instrumentation noise. A memo channel on the tape recorder allowed voice announcements, and date and time were recorded automatically.

B. Field procedures

Sounds and vibrations were recorded along transects that extended from the sound source of interest (e.g., ice road or piece of heavy equipment) across the landfast ice. Recordings were obtained along these transects at a range of distances, from ~40 to ~5300 m (Fig. 1). A large, articulated, 10 bag (10 tire) rolligon-type vehicle (model RD-105 by Crowley All Terrain Corporation: CATCO) was used to transport the instrumentation on the sea ice. A plywood structure on the trailer served as an instrumentation shelter. There were electric heaters in the shelter, powered by a generator on the "rolligon." During recording sessions, the RD-105 engine and the generator were shut down to avoid contaminating the recordings with local noise and vibration.

A second, smaller rolligon-type vehicle (Light All-Terrain Vehicle: LATV by CATCO) was used for augering 30 cm holes in the ice through which the hydrophone was lowered. The LATV moved past the recording site by at least 400 m and its engine idled during recording periods. No sound or vibration from the LATV was detectable at the recording site during recordings or at any other time while it idled at least 400 m away. The LATV served as a safety vehicle in the event that the instrumented RD-105 was unable to restart after a recording session.

In general, each recording station was about twice as far from the sound source as was the previous station, e.g., 100, 200, 500, 1000 and 2000 m, but pressure ridges usually prevented the transects from being laid out in a straight line (see Fig. 1). The field crew attempted to obtain recordings at distances out to one station beyond the closest station where the sound was not audible to the recordist. However, this was sometimes not possible because of logistical constraints. Also, the dominant components of the sounds from some sources (e.g., vibratory pile driving) were at infrasonic frequencies (25 Hz) barely audible to humans, but detected by the instruments and characterized in the subsequent analysis.

After the field crew on the LATV drilled the hole for a recording station, the instrumented RD-105 pulled alongside and the hydrophone was deployed to a position nominally 1 m above the bottom. A spike on the geophone was driven firmly into the ice. Snow was packed around for additional support, and a 1.4 kg weight was placed atop the geophone. The microphone was mounted on the corner of the Rolligon 1.5 m above the ice. The RD-105 was positioned so there was a clear path from the microphone and geophone toward the ice roads, island, or other probable sources of sound. Recording periods were usually 5–10 min in duration. Distances from specific sound sources were determined by a Bushnell model 20-0400 laser rangefinder at distances <300 m and by a Garmin handheld GPS at greater distances. Table I gives summary information for the field recordings.

TABLE I. Summary information for field recordings made in February, March, and May 2000 close to Northstar Island, Beaufort Sea.

Date	Local times	Range (m)	Activity recorded
1 Feb	11:00–16:15	100–2100	Ice road construction
2 Feb	9:45–17:45	97–1200	Ice road construction
18 Mar	14:15–17:00	145–3000	Pipeline trenching, gravel trucks, vibratory sheet pile driving
20 Mar	9:15–11:30	40–2200	Gravel trucks traveling to and from the island
20 Mar	13:15–15:45	97–3100	Ditchwitch cutting ice. Vibratory sheet pile driving
21 Mar	9:15–14:00	91–5300	Vibratory sheet pile driving
17 May	14:30–16:45	730	Impact sheet pile driving

C. Signal analysis

The recorded digitized signals were transferred as time series to a computer hard drive. They were then equalized and calibrated with a flat frequency response over the data bandwidth, 10–10,000 Hz for the hydrophone and microphone, and 10–500 Hz for the geophone. Analysis was performed with routines in MATLAB (The MathWorks, 3 Apple Hill Drive, Natick, MA 01760-2098) and custom programs.

For each recording we computed narrowband spectral densities by Fourier analysis, from which one-third octave band levels were derived. Broadband levels were calculated for the 10–10,000 Hz band for underwater and airborne sounds and the 10–500 Hz band for particle velocities. Analyses were performed on 8.5-s-long segments of time, except for impact pile driving sounds (see below). Variability in some of the operating source levels, such as the back and forth motion of a bulldozer clearing snow, caused variability in the results. Most sound samples were also played over a speaker to allow the analyst to hear the recorded sounds and thus to help identify the probable source of sounds.

Impact pile driving sounds contained sufficient low-frequency reverberation that they were for all practical purposes continuous and could not be analyzed using routines normally used for impulsive sounds (see, for example, Blackwell *et al.*, 2004b). Thus, pulse durations were taken from the start of one pulse (still clearly distinguishable in the sound pressure time series) to right before the start of the next pulse. Sound pressure level (SPL) and instantaneous peak level analyses spanned the duration. This means that background contributions were included, although those levels were generally weak.

In the March data, there were many stations at which it was not possible to distinguish between the various sources operating simultaneously. Most notably, signals from vibratory sheet pile driving often interfered with those from other sources. Additional interference came from wind-generated noise, which created turbulence at the surface that was received on the hydrophone and geophone but most conspicuously on the microphone, despite the use of a windscreen. Being broadband, wind noise could not be filtered out.

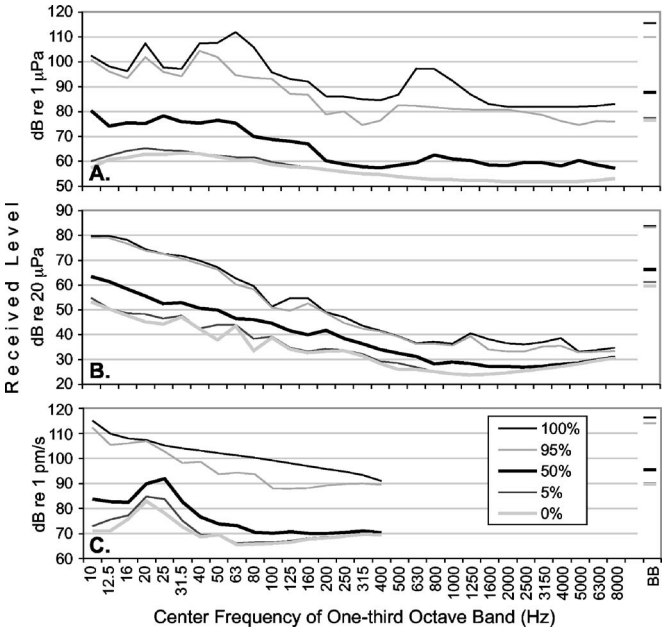


FIG. 2. Ambient sound statistics for (A) underwater sounds; (B) airborne sounds; and (C) iceborne vibrations. Broadband (BB) percentiles are also shown (at right) for each sensor.

1. Sound propagation modeling

One of two propagation models was fitted by the least squares method to broadband or tone data in order to characterize propagation loss underwater, in air, and in the ice. These models were based on logarithmic spreading loss:

$$\mathrm{RL(Received\ Level)} = A + B\log(R) \tag{1}$$

or, for some underwater sound cases,

$$\mathrm{RL} = A + B\cdot\log(R) + C\cdot R. \tag{2}$$

In these equations, R is the distance from sound source to receiver in m, and the units for RL are dB re: 1 μPa underwater, dB re: 20 μPa in air, and dB re: 1 pm/s (picometer/second) in the ice. The constant term (A) is close to the hypothetical level 1 m from the source, extrapolated back to 1 m range based on far-field measurements. The estimated A

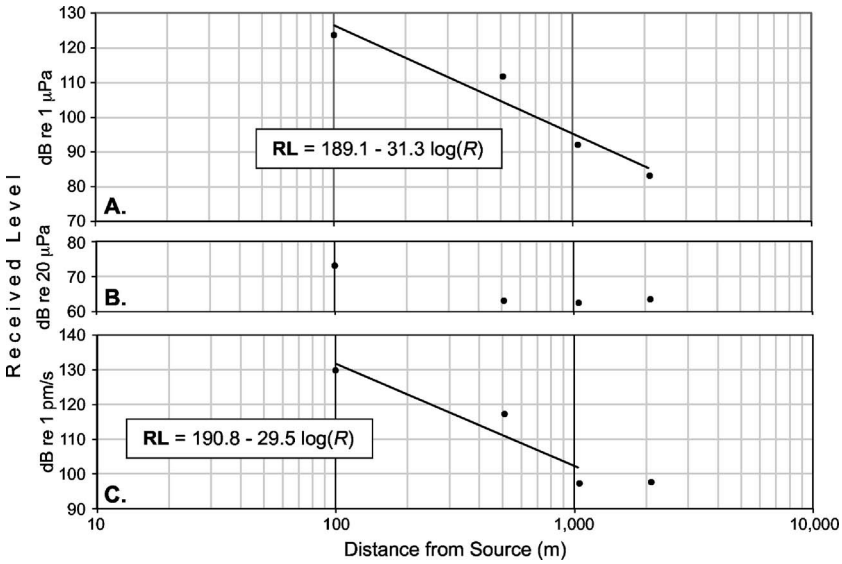


FIG. 3. Broadband received levels vs. distance for general ice-road construction activities. (A) underwater sounds; (B) airborne sounds; and (C) iceborne vibrations.

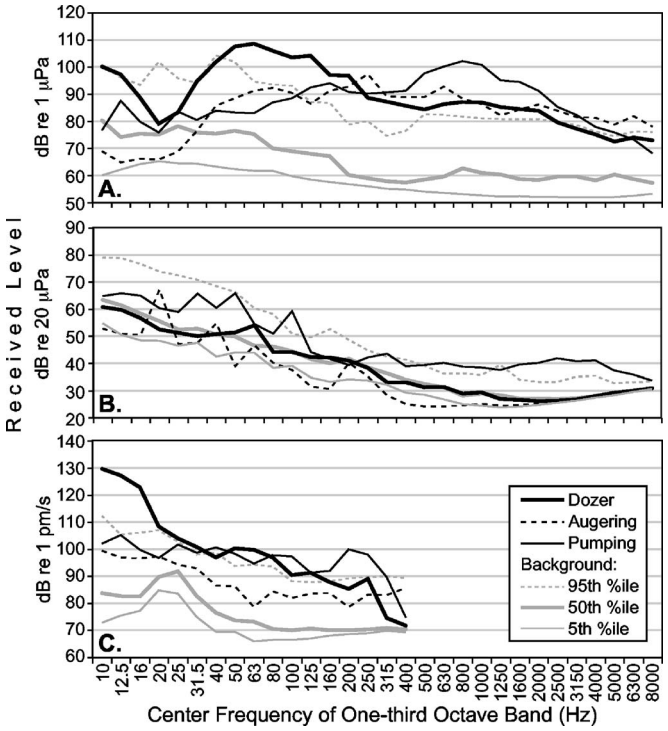


FIG. 4. One-third octave band levels for three activities during ice-road construction: a bulldozer pushing snow on the sea ice, augering a hole through the ice, and pumping sea water. Distance from sound sources is 100 m. (A) Underwater sounds; (B) airborne sounds; (C) iceborne vibrations. The corresponding 5th, 50th, and 95th percentile background sound levels are plotted for comparison.

value is useful mainly as a basis for comparison with other sources operating in the same region. The spreading loss term (B), which is negative, varies with the dominant frequencies in the sound source, water depth, bottom topography, and bottom composition. C (in dB/m) is the coefficient describing losses due to scattering and absorption; like B it is negative. The water depth, composition and topography of the seabed, and sea surface roughness will affect such losses.

When applying these models we used empirical data out to the distance beyond which values no longer decreased, i.e., background values were reached. In one case (see Fig. 5 later) the measurement closest to the sound source was also omitted because received levels there were lower than measurements at farther distances.

One-third octave band levels are presented for each type of machinery at a standardized distance of 100 m. For bulldozers, augering, pumping, and vibratory pile driving, a recording was obtained 100 m from the source. For trucks on the ice road, the Ditchwitch, and trenching with a backhoe, recordings were obtained 90–200 m from the sound source and one-third octave band levels were adjusted to a distance of 100 m using the logarithmic regression obtained with the broadband levels for each sound source.

A recording near each sound source was used to identify the one-third octave band with the highest received levels. The slope of the logarithmic propagation model obtained above for each sound source was then used to compute the distance at which the received level for the strongest one-third octave band was equal to the median level of background sound in the same one-third octave band (see below).

2. Background sounds

Background sound recordings were obtained whenever machinery seemed to be shut down, idling on the ice, or when recording equipment was distant from the ice roads and the island (typically 1–4 km from the predominant source). Additional ambient recordings were obtained near the island site when no activities were occurring (during lunch breaks and shift changes) and once during a storm when all personnel and equipment were evacuated. For each sensor, background sound recordings were pooled and the minimum, 5th-percentile, median, 95th-percentile and maximum levels were computed for one-third octave bands and for broadband values.

D. Reference units

The standard reference unit of sound pressure underwater is one micropascal (1 μPa). The standard reference unit of sound pressure in air is 20 micropascals (20 μPa). In this

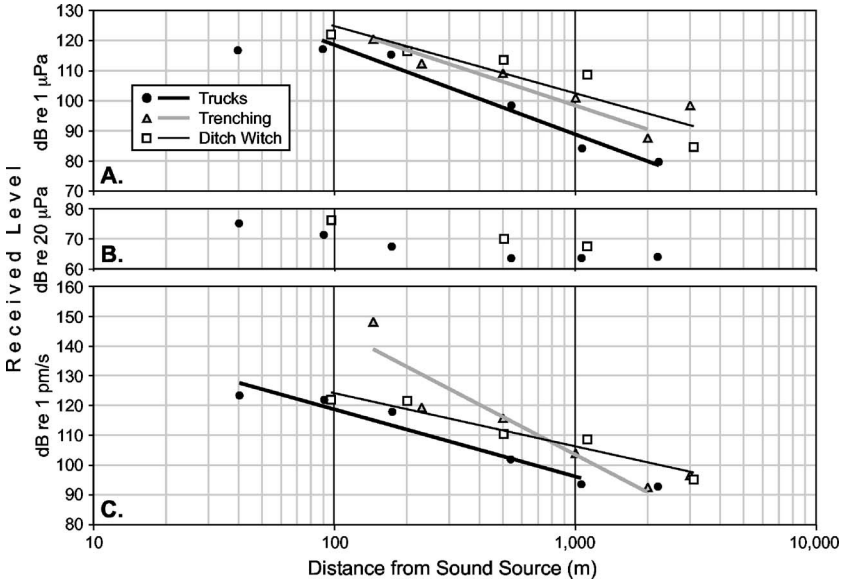


FIG. 5. Broadband received levels vs. distance for sounds from gravel trucks (filled circles), a backhoe digging a submarine trench for the pipelines (filled triangles), and a Ditchwitch sawing ice (open squares). (A) underwater sounds; (B) airborne sounds; and (C) iceborne vibrations.

TABLE II. Propagation loss equations for broadband levels of construction sounds presented in Fig. 4. No usable data were obtained with the microphone during backhoe trenching.

Sound source	Sensor	Equation
Trucks on ice road	Hydrophone	RL=179.1−30.1 log(R)
	Microphone	RL=109.9−21.5 log(R)
	Geophone	RL=164.8−22.8 log(R)
Backhoe digging trench	Hydrophone	RL=177.7−26.4 log(R)
	Geophone	RL=230.1−42.2 log(R)
Ditchwitch sawing ice	Hydrophone	RL=169.6−22.4 log(R)
	Microphone	RL=102.2−13.1 log(R)
	Geophone	RL=159.9−17.9 log(R)

chapter, we have used a reference particle velocity of one picometer per second (1 pm/s), or 1×10^{-12} meters/second. A velocity of 1 inch/second, which is 0 dB re 1 in./s, is 198 dB re 1 pm/s.

III. RESULTS

A. Background noise

Figure 2 presents a statistical distribution of one-third octave band levels vs. center frequency for the three types of sensors recorded during the February–May recording periods. Broadband levels of background noise (also shown in Fig. 2) extended from 77 to 116 dB re 1 μPa underwater and 59 to 84 dB re 20 μPa in air, in both cases for 10–10,000 Hz bandwidth, and from 90 to 116 dB re 1 pm/s for particle velocities (vertical channel as shown in Fig. 2, 10–500 Hz bandwidth). Predictably, the lowest levels coincided with the lowest wind speeds, minimum industrial activity and greatest distances from such activity. The highest levels indicate the noise levels during high winds but may include relatively low-level components of man-made sounds unknown to the authors. At the most distant recording stations (typically 1–4 km from the predominant source), man-made sounds were often not evident upon listening to the sounds, but narrowband spectrum analysis sometimes revealed tonal components of such sounds. Especially common in this regard were sources at low frequencies. For example, Fig. 2(C) shows elevation of the minimum, 5th percentile, and median geophone measurements in the one-third octave bands centered at 20 and 25 Hz; this peak is probably associated with vibratory sheet pile driving (see later).

B. Ice-road construction

To support the heavy trucks hauling gravel over the landfast ice to the island site and the pipeline installation machinery operating on the ice, ice roads up to 3 m thick were required. There were three elements to ice-road construction: bulldozing snow off the frozen surface, augering 30 cm holes through the ice, and pumping seawater up through the holes to flood the ice. Figure 3 shows broadband levels during ice-road construction as a function of distance from the source for all three sensors. Where applicable, Eq. (1) was fitted to the data. Figure 4 shows received one-third octave band levels for each of the three components of ice-road construction and each sensor, at a distance of 100 m from the source. The respective background noise range for each sensor type is also shown.

C. Ditchwitch, gravel trucks, and backhoe operations

A large Ditchwitch R100 machine was used to cut through the sea ice at the island site so gravel could be dumped. It also cut the slot in the ice along the pipeline route. Kenworth Maxihaul dump trucks (23 m^3 or 30 yd^3 capacity) transported the gravel from a river bed to the island site. A large Hitachi EX-450 backhoe straddled the pipeline slot and dug a trench in the sea bottom to a depth of 6 m. Figure 5 presents a summary of broadband levels for these three sound sources. Where applicable, Eq. (1) was fitted to received broadband levels, and the equations for the regressions shown are presented in Table II. Figure 6 shows re-

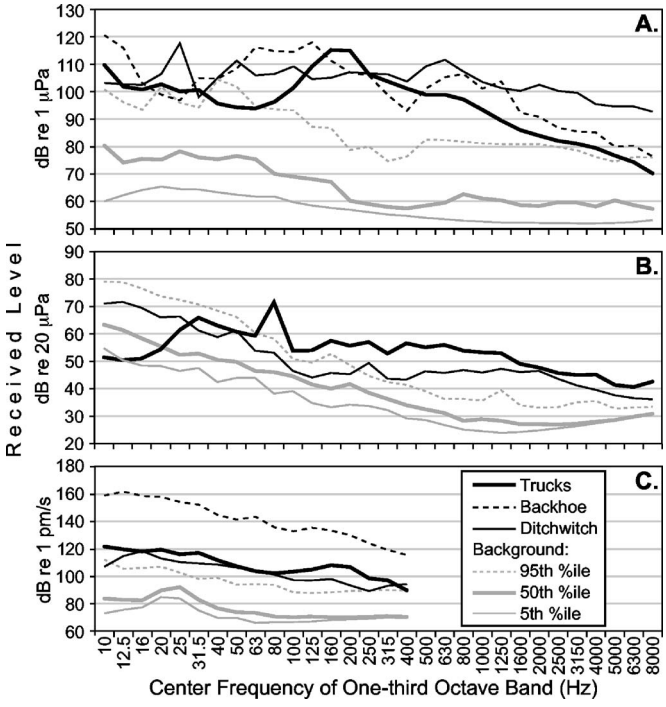


FIG. 6. One-third octave band levels for three sound sources: gravel trucks on the ice road, a backhoe digging a submarine trench for the pipelines, and the Ditchwitch sawing ice, all at distance 100 m. (A) Underwater sounds; (B) airborne sounds; and (C) iceborne vibrations. The corresponding 5th, 50th, and 95th percentile background sound levels are plotted for comparison.

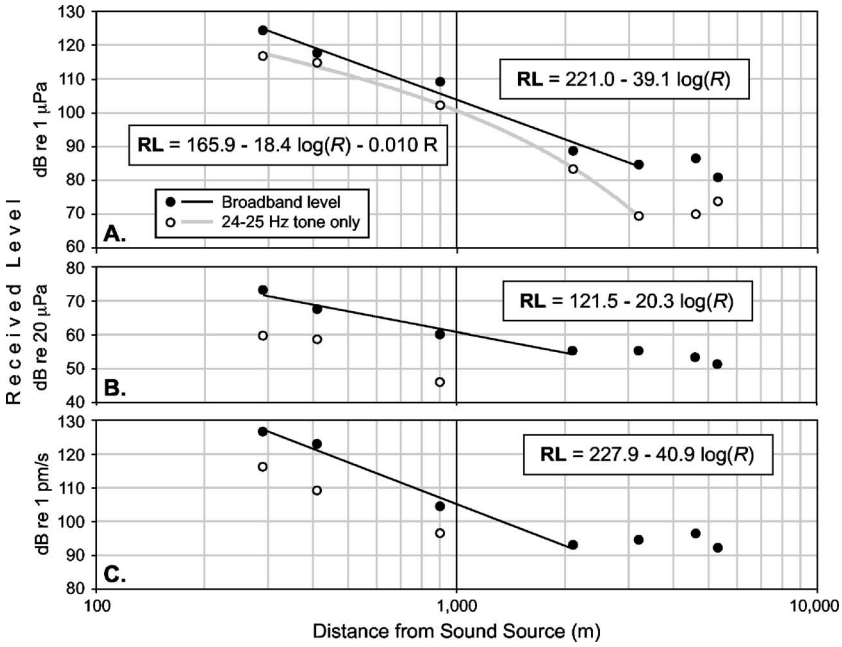


FIG. 7. Broadband received levels vs. distance during vibratory pile driving (filled circles), and received levels for the 24–25 Hz tone (open circles). Note that this is the SPL for the tone, not a sound pressure spectral density level. (A) Underwater sounds; (B) airborne sounds; and (C) iceborne vibrations.

ceived one-third octave band levels for each of the three sound sources and each sensor, at a distance of 100 m from the source.

D. Pile driving of sheet piles

Sheet piles were driven into the island to bound the operating area. The bases of these piles were surrounded by gravel, about 25 m from the waterline at the edge of the island, and thus were not in contact with the water. Vibratory pile driving (APE model 200) was used at first, and when the sheet piles reached a certain depth the vibratory pile driver had to be replaced by an impact pile driver (DELMAG D62-22). The vibrator could not drive the sheet piles through the frozen material of the old Seal Island. During most of the measurements with both types of pile drivers other activities were taking place, such as gravel hauling, island shaping, and pipeline construction. In addition, a crane was scooping gravel from beyond the island edge to form a subsea moat.

1. Vibratory sheet pile driving

Figure 7 presents received broadband (10–10,000 Hz) values during vibratory pile driving for all three sensors at seven different distances, and their respective logarithmic regressions. Most of the sound energy during vibratory pile driving, particularly underwater and in the ice, was in a tone at 24–25 Hz, the values of which are also shown in Fig. 7 for all three media. Equation (2) was fitted to the underwater tone data.

2. Impact sheet pile driving

Satisfactory recordings of impact pile driving sounds were only obtained at one station, 730 m from the sheet pile being driven into the island. Figure 8 presents the time series of a complete impact pile driving sequence (A) as recorded underwater, as well as details on an expanded time scale of both the beginning (B) and end (C) of the sequence. The sound from impact pile driving pulses appeared to reach the hydrophone by two dominant modes of propagation: (1) through the ground into the sea and propagating through the water, therefore undergoing the high-pass filtering effect of the shallow water (i.e., only higher frequencies reached the hydrophone), and (2) by refraction through the bottom layers over a longer path supporting the low frequencies. Sound from these two arrival paths can be seen in Fig. 8(B). The end of the sequence is shown in Fig. 8(C), with the last high-frequency pulse followed by the trailing low-frequency sound arriving last. In the underwater spectrum (not shown), there is a peak at about 6 Hz corresponding to the low frequency, bottom-traveling energy and a broad peak centered near 40 Hz and extending from about 20 to 55 Hz. Overall, most of the energy was below 55 Hz. Table III presents mean peak and rms values underwater and in the ice for 22 pulses analyzed at distance 730 m.

E. Comparison of sound sources

Table IV presents a summary of the levels of sounds and vibrations measured during winter 2000 around the Northstar prospect. For each sound source, the distance is presented at which the level in the strongest one-third octave band equals the median background level in the corresponding one-third octave band. The distances were calculated using the slope of the logarithmic regression obtained from broadband levels of each respective sound source.

IV. DISCUSSION

The sound and vibration levels presented in this paper are those measured for the sources operating near Northstar Island during the period of initial heavy construction in winter 2000. The data presented show that received levels of sound and vibration, as measured in the strongest one-third octave band for different construction activities, reached median background levels <7.5 km away for underwater

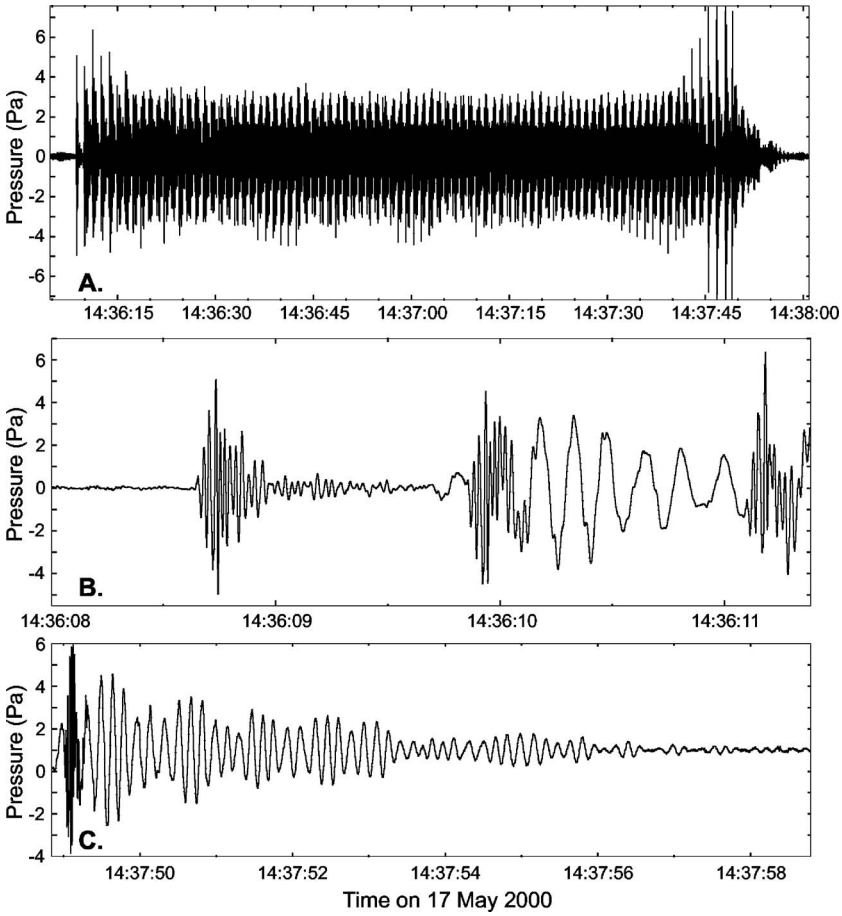


FIG. 8. Sound pressure time series of impact pile driving sounds received underwater at distance 730 m. (A) Complete sequence; (B) Expanded time signature of the start of the sequence shown in (A); (C) Expanded time signature of the end of the sequence shown in (A).

sounds, <3 km away for airborne sounds, and <10 km for iceborne vibrations. The measured background levels no doubt contained some industrial components at times, so distances to actual median ambient levels would be somewhat larger. Given the wide variability in natural ambient levels (e.g., Milne and Ganton, 1964; Lewis and Denner, 1987; see also Fig. 2), distances to actual ambient would be considerably greater and less at times of low and high ambient noise, respectively.

Results for the different sound sources are discussed in more detail below. Our measurements are compared to five other studies done in the same area and involving a variety of sound measurements. The first three are the most relevant because the sound sources recorded were the same or similar to those in this study:

(1) Underwater measurements of construction sounds were made in February and April 1982, when Seal Island was built the first time (Greene, 1983). The site was identical to the Northstar Island site, but in 1982 there were two parallel ice roads from Long Island just 5.7 km (3.1 n.mi.) long, as compared with the longer roads in the triangular configuration of 2000 (Fig. 1). In 1982 there were fewer activities than in 2000, as trenching

TABLE III. Peak and mean square pressure levels, and particle velocity levels, for 22 pulses from impact sheet pile driving received at distance 730 m from Northstar Island, 17 May 2000. Mean interpulse interval is also given.

	Peak pressure (dB re 1 μPa) or peak velocity (dB re 1 pm/s)	Mean square pressure (dB re 1 μPa) or mean square velocity (dB re 1 pm/s)	Interpulse interval (s)
Hydrophone			
Mean	131.6	123.7	1.22
Minimum	129.0	123.0	1.16
Maximum	136.1	124.7	1.33
Geophone			
Mean	140.3	132.3	1.22
Minimum	139.4	131.7	1.11
Maximum	141.1	133.3	1.30

TABLE IV. Summary of levels of sounds and vibrations from seven principal sound sources, for three parameters: (1) Broadband levels at 100 m. Measured broadband levels are shown if a recording was obtained at 100 m; if not, the broadband level at 100 m was calculated from the regressions obtained for each sound source (shown in Figs. 3, 5, 7, and Table II). (2) The center frequency of the strongest one-third octave band for each sound source, as determined from the closest recording (usually 100 m or less). For airborne pumping sounds two one-third octave bands (centered at 12.5 and 50 Hz) tied for the highest levels. (3) The distance from the source at which the level in the strongest one-third octave band was equal to the median level of background sound (see Fig. 2) in the same one-third octave band. This distance was calculated using the slope of the regressions obtained for each respective type of sound source. The "ice road construction" regression (Fig. 3) was used for the bulldozer, augering and pumping sounds except for airborne data, for which spherical spreading (20·log(R)) was assumed.

	Hydrophone (10–10,000 Hz)			Microphone (10–10,000 Hz)			Geophone (10–500 Hz)		
Sound Source	Broadband @ 100 m (dB re 1 μPa)	Center of strongest 1/3 OB (Hz)	Distance to 0 dB S/N in 1/3 OB (m)	Broadband @ 100 m (dB re 20 μPa)	Center of strongest 1/3 OB (Hz)	Distance to 0 dB S/N in 1/3 OB (m)	Broadband @ 100 m (dB re 1 pm/s)	Center of strongest 1/3 OB (Hz)	Distance to 0 dB S/N in 1/3 OB (m)
Bulldozer	114.2	63	1163	64.7	10	73	129.8	10	3613
Augering	103.3	250	1702	67.9	20	389	104.3	10	338
Pumping	108.1	800	1832	72.0	12.5	168	111.1	12.5	582
					50	631			
Ditchwitch	122.0	20	7292	76.3	12.5	612	121.9	16	9963
Trucks	123.2	160	3256	74.8	80	828	126.2	10	3310
Backhoe	124.8	10	3275	NA	NA	NA	145.7	12.5	2500
Vibr.sheet	142.9	25	2930	81.0	50	2,822	146.1	25	1207

with a backhoe did not take place nor did any type of pile driving. Analysis procedures in 1982 were not as comprehensive as they are today and neither broadband nor one-third octave band levels were computed, rendering comparisons difficult.

(2) The ANIMIDA project (Arctic Nearshore Impact Monitoring in the Development Area) included acoustic measurements, independent of ours, at the Northstar construction site during April 2000 (Shepard *et al.*, 2001). They reported narrowband spectrum levels and one-third octave band levels for underwater and airborne sounds, and iceborne vibrations, but not broadband levels. Their near-bottom hydrophone generally showed higher levels than did their mid-depth hydrophone and the data collected by their near-bottom hydrophone should be comparable to our data.

(3) During spring 2000 disposal well conductor pipes (51 cm diameter) were driven into the gravel on Northstar Island using an impact pile driver. Blackwell *et al.* (2004b) report on underwater and airborne sounds.

(4) Blackwell *et al.* (2004a) measured drilling and production sounds (underwater, in air, and in the ice) from Northstar during the winters 2001 and 2002.

(5) During winter 1979 BBN measured the underwater sounds of exploratory drilling on natural and artificial islands north of Prudhoe Bay, where an exploratory drill rig was operating (Malme and Mlawski, 1979).

Ice road construction was an activity that was difficult to separate into its individual components, as one or more bulldozers and several rolligons were normally working concurrently. For that reason broadband levels as a function of distance (Fig. 2) are reported for the ice road construction activity as a whole. Of the construction activities reported in this study, ice road construction produced the least amount of sound, in all three media. The distance to median background for the strongest one-third octave bands for bulldozers, augering, and pumping was <2 km for underwater sounds, <1 km for in-air sounds, and <4 km for iceborne vibrations (Table IV). During ice-road construction in 1982 (Greene, 1983) levels were comparable to 2000. A systematic comparison of narrowband samples obtained in 2000 during ice road construction (not shown) with those presented in Greene (1983) showed that, for comparable distances, values obtained in 2000 were both higher and lower than in 1983, i.e., 1983 values were within the range of variability of the values obtained in 2000.

In 1982 gravel trucks were also hauling gravel along ice roads, but their sounds were not distinguished from other components of "early island construction sounds" (Greene, 1983), making comparisons difficult. However, a comparison of spectral density levels from various samples in 2000 with those presented in 1983 does not lead us to suspect that general island construction sounds were much different in the two years.

Vibratory sheet pile driving produced the strongest extended duration sounds, with broadband underwater levels of 143 dB re 1 μPa at 100 m. Figure 7 shows that most of the sound energy was in a tone close to 24 Hz. Distances to background underwater (~3 km) were somewhat smaller than expected because of the larger than expected spreading loss term (39.1 dB/decade). During vibratory driving of sheet piles, Shepard *et al.* (2001) reported a 23 Hz tone which had underwater received levels of 114–116 dB re 1 μPa at distance 150 m and 90–93 dB at 2 km. The 24–25 Hz tone that we recorded in February 2000 would have had received levels of ~124 dB and ~85 dB at 150 m and 2 km, respectively (Fig. 7(A)). Our estimates were therefore 8–10 dB higher at 150 m and 5–8 dB lower at 2 km than the measurements by Shepard *et al.* (2001). The ANIMIDA study reported that the tone was detectable out to at least 2 km, and probably further, consistent with our own observations.

At a distance of 730 m in May, impact pile driving of sheet piles produced received peak and rms levels of 132 and

124 dB re 1 μPa underwater, respectively. Impact pile driving of disposal well conductor pipes occurred a month after these measurements were made (Blackwell *et al.*, 2004b). Acoustic recordings were made from the sea ice, which was beginning to break up. Received peak and rms levels underwater were 141 and 133 dB re 1 μPa at a distance of 650 m, nine decibels greater in each metric. However, the pipes driven in June were considerably larger than the sheet piles in May, a more powerful pile driver was used, and the conductor pipes were of a different configuration (and were driven deeper) than the sheet piles. These differences likely account for the differences in received sound levels.

Underwater sounds from drilling during the winter were measured by Malme and Mlawski (1979) north of Prudhoe Bay and Blackwell *et al.* (2004a) at Northstar. Malme and Mlawski (1979) concluded that broadband components from exploratory drilling were generally not detectable beyond about 1.8 km (1 n.mi.). Tonal components were detectable to about 10.6 km (6 n.mi.) under low noise conditions and considerably less under high noise conditions. In the Blackwell *et al.* (2004a) study underwater background sound levels were reached about 10 km from the island when ambient sound levels were very low (wind speed <1 m/s). This distance seems large compared to this study, but it is very much dependent on ambient sound levels. For example, if broadband ambient sound levels underwater in the Blackwell *et al.* (2004a) study were equal to median levels as recorded in 2000 (Fig. 2(A)), then the distance to background for drilling sounds would have been halved (4–5 km).

The present study indicates that broadband sounds from various construction activities and sources diminished into the range of the median background noise levels at distances ranging from 1 to 5 km at most, and usually less. Distances to median background for the strongest one-third octave band for each sound source were larger. For all three sensors combined these distances were <1 km in 35% of cases, <2 km in 55% of cases, and <4 km in 90% of cases. Not surprisingly the exception was the Ditchwitch, with distances to 0 dB S/N >7 km underwater and in the ice. Distances to background levels will vary greatly from time to time, given the wide variability in ambient levels.

ACKNOWLEDGMENTS

Jonah Leavitt, Jack Lawson, Mike Williams, Craig Perham, and Jesse Coltrane of LGL Alaska helped with the field recordings. CATCO Rolligon drivers went out of their way to support the field work. Bob Hall, Tom Barnes, Luke Franklin, John Phillips, and Earl Beverly of the BP Northstar project were helpful to the acoustics team in the field. Mike Williams provided a valuable review. W. John Richardson, LGL Ltd., served as Project Director and provided invaluable guidance and reviews. Dave Trudgen of Oasis Environmental and Val Moulton of LGL provided helpful comments on an early draft. Ted Elliott of LGL produced Fig. 1. Dr. Ray Jakubczak and Dr. Bill Streever of BP Exploration (Alaska) Inc. supported the work, and Dr. Streever provided useful comments on a draft of the manuscript. The anonymous reviewers for JASA were helpful. We thank them all.

Blackwell, S. B., Greene, C. R., Jr., and Richardson, W. J. (**2004a**). "Drilling and operational sounds from an oil production island in the ice-covered Beaufort Sea," J. Acoust. Soc. Am. **116**(5), 3199–3211.

Blackwell, S. B., Lawson, J. W., and Williams, M. T. (**2004b**). "Tolerance by ringed seals (*Phoca hispida*) to impact pipe-driving and construction sounds at an oil production island," J. Acoust. Soc. Am. **115**(5), 2346–2357.

Greene, C. R. (**1983**). "Characteristics of underwater noise during construction of Seal Island, Alaska, 1982, pp. 118–150," In: Biological studies and monitoring at Seal Island, Beaufort Sea, Alaska 1982, edited by B. J. Gallaway. Report from LGL Ecol. Res. Assoc. Inc., Bryan, TX, for Shell Oil Co., Houston, TX. 150 p.

Lewis, J. K., and Denner, W. W. (**1987**). "Arctic ambient noise in the Beaufort Sea: Seasonal space and time scales," J. Acoust. Soc. Am. **82**(3), 988–997.

Malme, C. I., and Mlawski, R. (**1979**). "Measurements of underwater acoustic noise in the Prudhoe Bay area," Report submitted to Exxon Production Research Co. Bolt, Beranek and Newman Inc. Tech. Memo. No. 513. 74 p.

Milne, A. R., and Ganton, J. H. (**1964**). "Ambient noise under Arctic-Sea ice," J. Acoust. Soc. Am. **36**(5), 855–863.

Moulton, V. D., Richardson, W. J., Elliott, R. E., McDonald, T. L., Nations, C., and Williams, M. T. (**2005**). "Effects of an offshore oil development on local abundance and distribution of ringed seals (*Phoca hispida*) of the Alaskan Beaufort Sea," Marine Mammal Sci. **21**(2), 217–242.

Shepard, G. W., Krumhansl, P. A., Knack, M. L., and Malme, C. I. (**2001**). "ANIMIDA Phase I: Ambient and industrial noise measurements near the Northstar and Liberty Sites during April 2000," Report submitted to U.S. Dept. of the Interior, Minerals Management Service, Alaska OCS Office, Anchorage, by BBN Technologies. Report No. OCS/MMS 2001-0047, 73 p. Obtainable at http://www.mms.gov/alaska/reports/2001rpts/akpubs01.HTM. Viewed 28 June 2007.

Williams, M. T., Nations, C. S., Smith, T. G., Moulton, V. D., and Perham, C. J. (**2006**). "Ringed seal (*Phoca hispida*) use of subnivean structures in the Alaskan Beaufort Sea during development of an oil production facility," Aquat. Mamm. **32**(3), 311–324.

Finite element model for waves guided along solid systems of arbitrary section coupled to infinite solid media

Michel Castaings[a)]
Laboratoire de Mécanique Physique, Université Bordeaux 1, UMR CNRS 5469, 33400 Talence, France

Michael Lowe
Department of Mechanical Engineering, Imperial College, London SW7 2AZ, United Kingdom

(Received 20 June 2007; accepted 13 November 2007)

The Semi-Analytical Finite Element (SAFE) method is becoming established as a convenient method to calculate the properties of waves which may propagate in a waveguide which has arbitrary cross-sectional shape but which is invariant in the propagation direction. A number of researchers have reported work relating to lossless elastic waves, and recently the solutions for nonpropagating waves in elastic guides and for complex waves in viscoelastic guides have been presented. This paper presents a further development, addressing the problem of attenuating waves in which the attenuation is caused by leakage from the waveguide into a surrounding material. This has broad relevance to many practical problems in which a waveguide is immersed in a fluid or embedded in a solid. The paper presents the principles of a procedure and then validates and illustrates its use on some examples. The procedure makes use of absorbing regions of material at the exterior bounds of the discretized domain.
 [DOI: 10.1121/1.2821973]

PACS number(s): 43.35.Cg, 43.20.Mv [YHB] Pages: 696–708

I. INTRODUCTION

Guided wave modes are attractive, and increasingly successful, for the nondestructive testing (NDT) of a large variety of structures, because of their potential to propagate long distances while interrogating the whole thickness of the tested specimen.[1–4] It is well known that the successful development of such inspection techniques first requires careful model studies to be performed in order to properly understand the nature of the waves. In particular, dispersion curves and through-thickness mode shapes are necessary for evaluating the time of arrival at a monitoring position of any given mode, as well as its attenuation or its sensitivity to specific target defects to be tested.

Analytical dispersion equations[5,6] or matrix methods[7–10] can be established for predicting such data for waveguides of certain regular geometries. Specifically, such models can deal with flat plate or cylindrical structures, including multiple layers, a variety of isotropic and anisotropic elastic and viscoelastic materials, and the waveguides may be in vacuum, immersed in a fluid, or embedded in a solid. However these models cannot be used to study the waves in guides having irregular cross-sectional geometries, such as railway lines or T-shaped beams. For such cases the semi-analytical finite element (SAFE) method is becoming popular (also variously called spectral element, strip element, or waveguide finite element methods). This uses a finite element representation of the cross section of the waveguide, thereby enabling arbitrary definitions of shapes, together with a harmonic description along the propagation direction.[11–29]

There have been some significant recent developments of the SAFE method. Damljanovic and Weaver[30] have reported the rigorous solution of all of the roots for waves (propagating and nonpropagating) in an elastic waveguide, an important outcome being that necessary results are available for modal composition calculations, for example for the solution of scattering problems. Predoi *et al.*[31] have reported the implementation of periodic boundary conditions in the SAFE method, allowing cases of infinitely wide guides with periodic changes in geometry or material properties along the width to be modeled. Bartoli *et al.*[32] have studied waves in guides whose material has damping properties, so that the solutions are necessarily complex, the damping of the guided waves being represented by the imaginary part of the wave number. Wilcox *et al.*[33] have shown how the SAFE method may be applied easily using standard commercial finite element (FE) programs, thus avoiding the necessity to write specific FE code. Jezzine and Lhemery[34] have developed a SAFE model which includes the excitation and reception by a realistic representation of a transducer on the end of the waveguide, so enabling the prediction of the behavior in practical waveguide setups.

However, the work reported so far relates almost entirely to waveguides with stress-free exterior boundaries, with only two exceptions known to the authors. One is a successful attempt by Liu and Achenbach, who used artificial viscoelastic elements for modeling either an unbounded solid or a half-space submitted to dynamic excitation.[35] However, neither of these two models concerns the propagation along a waveguide with radiation into an infinite surrounding me-

[a)]Author to whom correspondence should be addressed. Electronic mail: m.castaings@1mp.u-bordeaux1.fr

dium, such as we do in this paper. The other case is much more related to our work, but it is a brief example result in the work by Jezzine and Lhemery,[34] also discussed in the Ph.D. thesis by Jezzine,[36] although their paper does not investigate the procedures. Thus an important class of problem has been largely neglected. This is the case of wave modes propagating along a guide of arbitrary cross section, and radiating into a surrounding fluid or solid medium of infinite extent, the so-called "leaky" waves. Such cases are practically important, for instance for the NDT of immersed or embedded waveguide structure,[37] or for the development of fluid sensors,[38–40] some of which benefit from irregularity of the cross section.[40] Also, the possibility of exploiting waves that are preferentially guided by structural features such as stiffeners or welds has recently been identified.[41] Such waves would in general be leaky, radiating some of their energy into the adjacent structural material.

This paper presents the implementation of the established FE absorbing regions methodology[42,43] into the SAFE method, such that the radiating leaky waves are absorbed at the exterior of the discretized domain. Energy is then lost from the guided waves, and this attenuation becomes part of the calculation of the wave properties, being expressed by the imaginary part of the wave number solutions. Knowledge of the attenuation is very important to any practical applications, so the paper critically aims toward the goal of the reliable use of absorbing regions to achieve accurate prediction of both the real and the imaginary parts of the solutions.

The paper starts with a brief reminder of the SAFE formalism, and introduces the absorbing regions concept for the leaky waves solutions. Two example cases of simple waveguides are then solved in order to illustrate the implementation and provide validation against known solutions. The first validation case is the propagation of guided modes along a flat plate attached to a semi-infinite half-space of an elastomer material. The second is a steel bar embedded in an infinite concrete medium. This is a much more challenging problem, which reveals important considerations to be addressed when constructing models. Finally two example cases which illustrate the potential relevance to NDT are solved. These are the propagation of guided wave modes along a welded joint between two very large steel plates, and along a steel stiffener bonded on a viscoelastic material plate.

II. MODEL

A. SAFE method

The mathematical model considered in this study is based on the three-dimensional elasticity approach, so that no simplifications are made to the elastic tensor, or to the displacement field. Harmonic guided waves propagating along the x_3 axis are considered. Consequently, the displacements vector $\boldsymbol{u}_i$ in the waveguide can be written:

$$\boldsymbol{u}_r(x_1,x_2,x_3,t) = \boldsymbol{U}_r(x_1,x_2)e^{I(kx_3-\omega t)}, \quad I=\sqrt{-1}, \tag{1}$$

in which k is the wave number along the axial direction (x_3) of the waveguide, $\omega=2\pi f$ is the angular frequency, f being the frequency, t is the time variable, and the subscript $r=1,2,3$. For general anisotropic materials, the equation of dynamic equilibrium can be written in the following form of an eigenvalue problem:

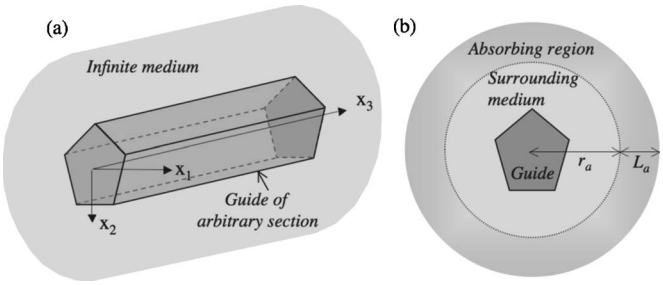


FIG. 1. Schematics of a (a) three-dimensional waveguide embedded in infinite medium and (b) two-dimensional model with absorbing region used for FE calculation of waveguide properties of the system shown in (a).

$$C_{sprq}\frac{\partial^2 U_r}{\partial x_p\,\partial x_q} + I(C_{s3rp}+C_{spr3})\frac{\partial(kU_r)}{\partial x_p} - kC_{s3r3}(kU_r) + \rho\omega^2\partial_{sr}U_r, \tag{2}$$

with summation over the indices $r=1,2,3$ and $p,\ q=1,2$.[31] The coefficients C_{sprq} are the stiffness moduli, which will take real or complex values for elastic or viscoelastic materials, respectively. In the following, these moduli will be expressed using the standard contracted notation C_{ij} $(i,j=1,\ldots,6)$.

The nature of the solution is thus to find eigenvalues of wave number (k) for chosen values of angular frequency (ω). In the context of the construction of dispersion curves, each solution at a chosen frequency will reveal the wave numbers of all of the possible modes at that frequency; the full dispersion curve spectrum is then found by repeating the eigenvalue solutions over the desired range of frequencies.

In the commercial FEM code used for the implementation of the developments of this study,[44] the formalism for eigenvalues problems has the following expression:

$$\nabla(c\,\nabla\,U+\alpha U-\gamma)-\beta\,\nabla\,U-aU+\lambda d_a U=0. \tag{3}$$

Equation (3) is a general form of an eigenvalue equation and the various coefficients c, α, γ, β, a, λ, and d_a do not have any particular physical meaning except that they come from functions of the parameters of the investigated problem. As shown in Ref. 31, c, α, and β depend on the material stiffness properties, a is a function of the mass density and circular frequency, d_a depends on the stiffness properties, the mass density, and the circular frequency, and γ and λ are null for our purpose.

Considering a solid waveguide with a cross section of arbitrary but constant geometry, embedded in an infinite solid medium [Fig. 1(a)], and considering that all derivative terms with respect to variable x_3 in the classical equation of dynamic equilibrium[45] are replaced by (Ik) factors, as shown in Eq. (2), the problem can be reduced to a two-dimensional problem, which requires meshing the cross section of the guide only. The purpose of the study consists in finding eigenvalues k of Eq. (2) that represent mechanical resonances between the lateral sides of the inner waveguide, and that radiate energy at infinity in the surrounding medium while propagating along the x_3 direction, at a given frequency.

However, in FE models, each element of the system has to be limited by borders, so that the infinite surrounding medium must be of finite extent. Thus there is a problem with the present SAFE approaches because, however large the surrounding material is chosen to be, it is still finite, so the set of eigenvalues k found when solving Eq. (2) represents resonances of the whole bounded system, including the inner-guide and the surrounding medium, and consequently the leaky solutions are not found.

B. Absorbing regions

If the waveguide is embedded in an infinite solid medium then it becomes necessary to include some description of the infinite exterior domain in the FE model. Many approaches have been used and indeed are available in established FE programs, including so-called nonreflecting boundary conditions and infinite elements. However, the authors increasingly favor the alternative use of a finite region of material in which the waves are simply absorbed.[42,43] This can be achieved using perfectly matched layers or materials with physical damping properties. The reason for the preference of this approach is its robustness of performance for general use with a variety of angles of incidence and kinds of waves.

An absorbing region is modeled all around the surrounding medium, as shown in Fig. 1(b). As described in reports of previous work with absorbing regions,[42,43] the absorbing region is rendered more and more viscoelastic as the distance away from the region of interest increases. Specifically, the absorbing region has the same mass density and elastic properties as the medium of interest which it surrounds, but its viscoelasticity, i.e., the imaginary parts of its elastic moduli, gradually increases according to the following law:

$$C_{ija} = C'_{ij}\left[1 + I\alpha\left(\frac{r - r_a}{L_a}\right)^3\right] + IC''_{ij}, \quad I = \sqrt{-1}, \tag{4}$$

where C'_{ij} represents the stiffness of the medium surrounding the waveguide, C''_{ij} its viscoelasticity, r_a is the inner radius of the absorbing region from the center of the region of interest, L_a its radial length, r the radial position in this absorbing region ($r > r_a$), and C_{ija} are the resulting viscoelastic moduli of the absorbing region, α is a coefficient that defines the proportion, relative to the stiffness, of the viscoelasticity at the outer limit of the absorbing region. This coefficient must be optimized such that reflections from the outer boundary of the absorbing region are negligible, while avoiding too brutal changes in the acoustic impedance between the surrounding medium and the absorbing region which would themselves cause backscatter. Optimization studies have been performed in earlier work relating to conventional FE modeling of bulk and guided wave problems, and have been reported by the authors.[42,43] Furthermore, more recent work has investigated the performance of these layers analytically, including the influence of the angle of incidence of the wave at the boundary.[46] The guidelines from that work were used here to set the parameters of the absorbing regions. Such tests have shown that this coefficient should be chosen to be higher at low frequencies than at high frequencies in order to compensate for the relatively small attenuation of bulk waves at low frequencies. Typically, for an absorbing region having a length comprised between two and three times the longest wavelength of any radiated wave in the whole frequency range, α, can take values between 4 and 1 for small and high frequencies, respectively. In fact, both the length of the absorbing region and the value of α have to be optimized according to the type of radiated waves, which can be either bulk waves or guided waves (the latter will be illustrated later in the examples).

With the absorbing regions in place around the area of interest, radiating waves are no longer reflected from the exterior of the model, and so the solution of Eq. (2) can now include guided waves which attenuate due to the radiation. In such cases the attenuation is described by the imaginary part of the wave number. An important task when processing the eigenvalue results is to identify and separate these guided wave solutions from unwanted eigenvalues such as resonances of the whole body. An approach to achieve this will be discussed in Sec. III.

III. PRACTICAL SOLUTION AND VALIDATION

The practical implementation of the technique to the solution of realistic waveguide problems is best discussed by presenting it in the context of solving example problems. Here we choose two problems whose geometry is sufficiently simple that solutions can be found by alternative means, for validation, but which also possess sufficient complexity to expose the key issues of the procedures.

A. Steel plate attached to semi-infinite half-space of elastomer material

The first example was chosen to be particularly simple, with the principal aims of showing how the solution is obtained and demonstrating that the results agree with those from an alternative established method. Alternative methods such as matrix techniques[7–10] are generally limited to simple geometric shapes. Here we have used a commercial matrix-method program,[9] which can model continuous flat plates or cylinders. The example consists of a continuous flat aluminum plate which is attached to a semi-infinite half-space of an elastomer material. Such a structure can support guided waves which propagate along the plate while leaking energy into the elastomer. Thus the solution involves the calculation of the guided wave velocity dispersion curves as well as the attenuation due to the leakage. The elastomer material has been chosen to have an acoustic impedance which is very much lower than that of the plate, so as to define a low-leakage simple first problem.

1. Geometry, material properties, and FE input data

To represent infinitely wide crested waves in a continuous flat plate it is only necessary to model a narrow strip of the structure, provided that it is given appropriate boundary conditions at its edges. Figure 2 shows the spatial domain of the FE model. The thickness of the plate is 4 mm. Attached to this is a 10-mm-thick layer of the elastomer material, this thickness chosen to be large enough to be able to examine

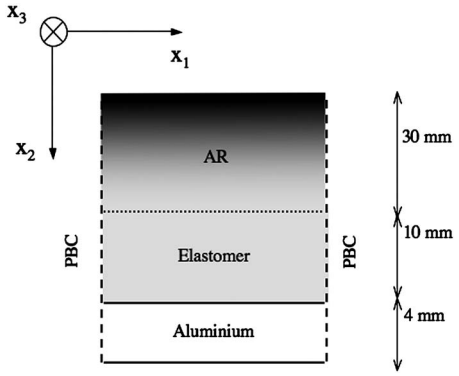


FIG. 2. Schematic of the FE model used for the aluminum plate attached to a semi-infinite half-space of an elastomer material.

the leaking waves. Then attached to that is a 30-mm-thick absorbing region having the same elastic properties as the elastomer, but increasing damping properties in order to absorb the leaking waves, thus simulating their radiation to infinity. A narrow strip of the plate of just 1 mm width is modeled. The material properties are given in Table I.

Continuity of displacements and stresses is imposed between the plate and the elastomer and between the elastomer and the absorbing region. Stress-free conditions are imposed at the outer surface of the plate and of the elastomer. Periodic boundary conditions[31] are imposed at the lateral boundaries of the domain. The whole geometry is meshed by 864 square elements of second order, with side lengths of 0.25 mm. The number of degrees of freedom is 23 382, and Eq. (2) is solved for 18 frequencies from 50 to 400 kHz. For each frequency, several complex solutions of the wave number k are obtained. In order to select the solutions of interest from the full set of solutions obtained at each frequency, the axial component of the energy flow is calculated at each nodal position of the mesh and used to plot an image of the energy flow over the cross section. This quantity is expressed by the following well-known formulas:

$$P_{x_3} = -\frac{1}{2}\,\mathrm{Re}(\boldsymbol{v}^* \cdot \bar{\bar{\boldsymbol{\sigma}}}) \cdot x_3$$

$$= \frac{1}{2}\,\mathrm{Re}(I\omega \boldsymbol{u}^* \cdot \bar{\bar{\boldsymbol{\sigma}}}) \cdot x_3$$

$$= -\,\mathrm{Re}\left[\left(\frac{I\omega}{2}\right)(u_1^*\sigma_{31} + u_2^*\sigma_{32} + u_3^*\sigma_{33})\right], \qquad (5)$$

where

$$\sigma_{31} = C_{55}\left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1}\right) = C_{55}\left(Iku_1 + \frac{\partial u_3}{\partial x_1}\right),$$

$$\sigma_{32} = C_{44}\left(\frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3}\right) = C_{44}\left(\frac{\partial u_3}{\partial x_2} + Iku_2\right),$$

$$\sigma_{33} = C_{13}\frac{\partial u_1}{\partial x_1} + C_{23}\frac{\partial u_2}{\partial x_2} + C_{33}\frac{\partial u_3}{\partial x_3}$$

$$= C_{13}\frac{\partial u_1}{\partial x_1} + C_{23}\frac{\partial u_2}{\partial x_2} + C_{33}Iku_3$$

are components of the stress tensor $\bar{\bar{\boldsymbol{\sigma}}}$, the asterisk stands for complex conjugate, and $\boldsymbol{v}$ is the particle velocity. All of these quantities are easily extracted from the eigenvectors of the FE solutions, although of course the eigenvectors have arbitrary scale, so it is important to use the energy flow values in a comparative way, by comparing values at one location to another, rather than using any absolute values. Solutions k for which the energy-flow component in the images is clearly dominant in the aluminum plate are selected.

2. Results

Figure 3 presents the dispersion curves for the guided wave modes in the plate. The dots are from the SAFE solutions obtained with the model and the postprocessing described in Sec. II, and the lines are predictions made with a commercial matrix-method program.[9] Figure 3(a) shows the phase velocity dispersion curves, calculated from the real part of the wave number, and Fig. 3(b) shows the attenuation, corresponding to the imaginary part of the wave number. Excellent agreement can be seen for all calculated points, thus demonstrating the validity of the approach taken and the accuracy, which is achievable.

As a final comment, a thumb rule may be proposed for defining the length of the absorbing region in the model. As stated in Refs. 42 and 43, these regions should be as long as twice the maximum wavelength in the frequency range of interest, and this length was set in the direction of propagation of guided wave modes. In the cases considered in this paper, waves to be absorbed are those radiated in media surrounding guides, and since the cross section only of the entire domain is modeled, then the absorbing regions will be defined in this section and their role should be to absorb any radiation modeled in this section. Therefore, their length should be twice the longest wavelength in the plane of the mesh, λ_{x_2}, that is to say, in the x_2 direction for the case of the steel plate attached to the semi-infinite elastomer.

TABLE I. Mechanical properties of materials used in models.

	ρ (g/cm^3)	C_{11} (GPa)	C_{66} (GPa)	V_L (m/s)	V_T (m/s)	α_L and α_T (dB/λ)
Aluminum	2.78	112	27	6347.3	3116.4	⋯
Elastomer	1.1	$1\times(1+i0.1)$	$0.3\times(1+i0.1)$	953.5	522.2	2.73
Steel	7.932	281.8	84.3	5960.5	3260	⋯
Concrete	2.3	41	16	4222.1	2637.5	⋯
Adhesive	1.1	$7\times(1+i0.03)$	$1.1\times(1+i0.03)$	2523	1044	0.819
Perspex	1.45	$8\times(1+i0.05)$	$1.5\times(1+i0.05)$	2348.8	1017	1.36

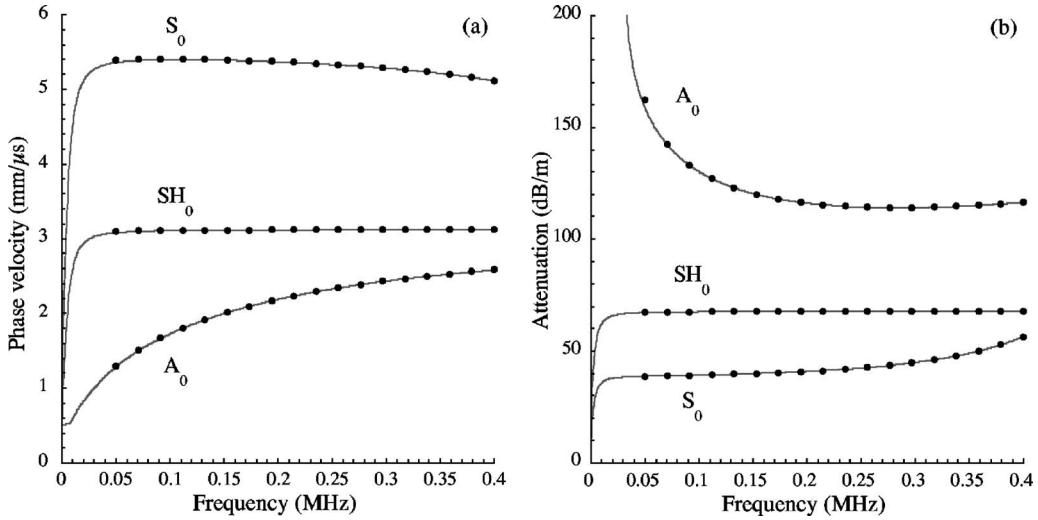


FIG. 3. Dispersion curves for modes guided in the aluminum plate and radiating into the elastomer; (a) phase velocities and (b) attenuations. Comparison of SAFE calculations (•••) with conventional matrix-method solutions (—).

This value may be approximated by following a systematic process which consists in (1) roughly estimating the smallest phase velocity, $\min(V_m)$, of all modes m that may propagate along the guide, in the frequency range of interest (this may be done by a quick run for the free guide), (2) using the well-known Snell–Descartes law[45] for calculating the maximum angle of radiation in the elastomer, obtained for the compression wave radiated with high velocity V^L_{rad}:

$$\max(\theta_{\text{rad}}) = \sin^{-1}\left(\frac{V^L_{\text{rad}}}{\min(V_m)}\right), \tag{6}$$

and (3) calculating the projection onto the cross section of the corresponding longest wavelength radiated at the highest angle:

$$\max(\lambda_{x_2}) = \frac{\max(\lambda_{\text{rad}})}{\cos[\max(\theta_{\text{rad}})]} = \frac{V^L_{\text{rad}}/f}{\cos[\max(\theta_{\text{rad}})]}. \tag{7}$$

The length of the absorbing region should then be set equal to twice this maximum projected wavelength for all of the calculations, to achieve reliable absorption. Figure 4 shows two situations corresponding to (a) a not optimized absorbing region and (b) an optimized one. Although this rule of using absorbing regions of two projected wavelengths thick has led to quite good results in many trials, it is sometimes possible to reduce this dimension, especially if the surrounding medium is made of a viscoelastic material, thus causing some absorption of the radiated waves before they reach the absorbing region. This was the case for the actual example where the absorbing region above the elastomer was set equal to one only for the maximum projected wavelength.

B. Steel bar embedded in infinite concrete

The second example also enables the method to be validated, because the results can be compared with those from alternative dependable solutions. However, this example is significantly more complex, involving a cylindrical geometry and a surrounding material whose properties present much less contrast to those of the waveguide than in the preceding case. This will serve to illustrate important considerations for the preparation of models. The example consists of finding the complex roots k of Eq. (2), representing modal solutions for a steel bar embedded in an infinite medium made of concrete. Solutions for this problem have already been found using the global matrix method.[8,9,37]

1. Geometry, material properties, and FE input data

The geometry of the system is presented in Fig. 5. The steel bar is 20 mm in diameter. The surrounding ring of concrete is attached at its inner radius to the steel and has a thickness of 20 mm. The concrete is intended to be infinite in extent, so the chosen value of the thickness of the concrete ring has no physical significance to the problem being studied. The mechanical properties of the steel and concrete are

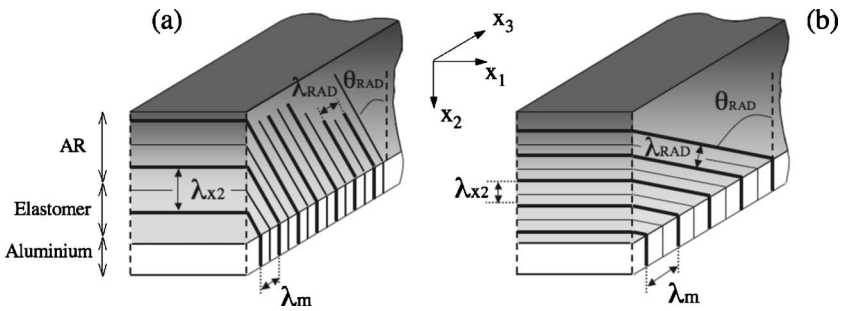


FIG. 4. Defining length of absorbing regions from projected wavelength of radiating waves; Cases of (a) not well defined and (b) well defined absorbing regions.

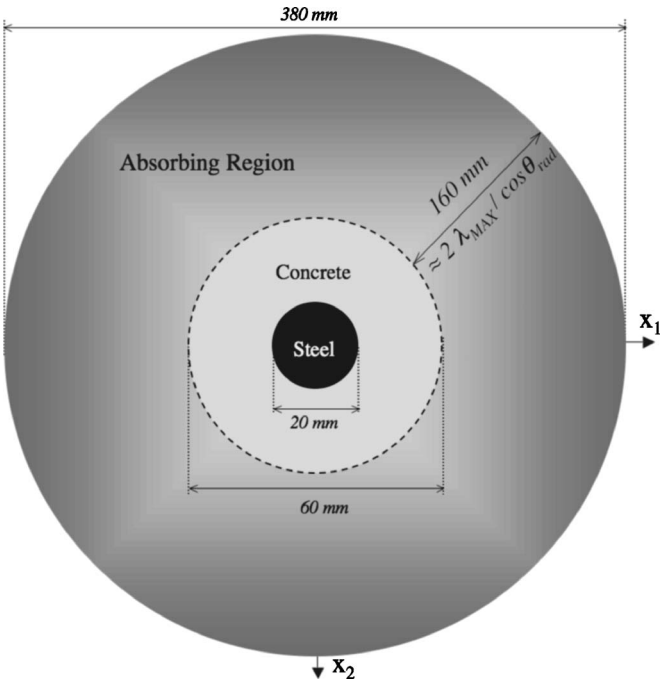


FIG. 5. Schematic of the FE model used for the steel bar embedded in infinite concrete.

given in Table I. The absorbing region is modeled by an exterior ring attached to the outside of the concrete, with a thickness of 160 mm. According to previous studies,[42,43] the thickness of this region was chosen to be equal to twice the maximum wavelength of the bulk waves expected to be radiated into the concrete when guided modes propagate along the steel bar, i.e., the wavelength of longitudinal waves at the lowest frequency of the investigated frequency range. However, radiated waves in the concrete are unlikely to propagate in the direction normal to the steel bar–concrete interface, but should form angles that will depend on the ratio between their wave number and that of the wave mode guided along the steel bar, according to the Snell–Descartes law.[45] The required length of the absorbing region should then be found by projecting the wavelength of the bulk waves onto the radial direction, and taking twice this value. This can be done using

$$\lambda_{x_1} = \frac{\lambda_{\mathrm{rad}}}{\cos(\theta_{\mathrm{rad}})}. \tag{8}$$

However, there is a problem here. If one considers that the angles of radiation θ_{rad} may be comprised between 0° and almost 90° (at exactly 90°, one considers that there is no radiation), then Eq. (8) can yield values tending to infinity. In our example, let us consider setting a pragmatic upper limit of 89.9°. The maximum possible wavelength of the leaking bulk waves is about 84 mm, at 50 kHz, then its projection onto the radial direction λ_{x_1}, is calculated to be 48 000 mm. This would imply the need of a 100-m-long absorbing region, and of a tremendous number of mesh elements for the cross section of the system. It was thus decided, arbitrarily, to set the length of the absorbing region equal to 160 mm, and to accept that some results involving large radiation angles will probably be erroneous due to inefficiency of the absorbing regions. This is acceptable as long as the erroneous results can be identified. This will of course be straightforward in this example because there is an alternative solution for comparison, but careful attention to this point will have to be paid in the general use of the method.

Continuity of displacements and stresses is imposed at each internal border, i.e., between steel and concrete and between concrete and the absorbing region. Stress-free conditions are imposed at the outer limit of the system. The whole geometry is meshed by 8930 triangular elements of first order (each element has three nodes), with sidelengths comprised between 2 and 4.5 mm. These elements are automatically generated by the software used,[44] and have sizes slightly increasing with the distance away from the central axis. Equation (2) is solved for 15 frequencies over the range from 50 to 200 kHz. For each frequency, several solutions k are obtained, those unwanted representing resonances of the whole system, and those sought corresponding to resonances of the steel bar, which represent modes guided along the bar and radiating into the concrete.

2. Results

As in the previous example, the axial energy-flow component is calculated at each nodal position for each eigensolution, and those for which this is clearly more dominant across the region of the steel bar than anywhere else are selected. In addition, solutions for which the ratio $|\mathrm{Im}(k)/\mathrm{Re}(k)|$ is greater than 15% are eliminated because these correspond to waves having tremendous attenuation, and are not directly of interest for NDT applications.

As an example, two results are presented in Fig. 6, one showing an unwanted solution that corresponds to a resonance of the absorbing region [Fig. 6(a)], and the other showing a wanted solution that corresponds to a resonance of the steel bar coupled to the surrounding concrete [Fig. 6(b)]. Figure 7 presents some of the dispersion curves of wave modes propagating along the steel bar and radiating energy into the infinite concrete. The lines are predictions made with a commercial matrix-method program,[9] while dots represent the SAFE solutions obtained with the model described in Sec. II and the postprocessing described in case (a). Only the solutions for circumferential order zero (axially symmetric, solid lines) and order one (one harmonic order around the circumference, dashed lines) are shown. The names given to the various curves are chosen for the close correspondence of the wave solutions to modes of these names in a free cylindrical bar, the difference here being the leakage of energy into the surrounding concrete. Note that the discontinuities of the curves $L(0,1)$ and $L(0,1)'$ correspond to their interceptions with the bulk velocities of the materials, as can be seen in Fig. 7(a). The attenuation in Fig. 7(b) is defined by $20\log_{10}(e^{1000k''})$, expressed in dB/m, with k'' being the imaginary part of the wave number solution, in rad/mm.

For most of the SAFE calculation points there is very good agreement with the independently calculated dispersion curves. This can be seen in the figure, showing that the SAFE calculation has found both the propagation and the

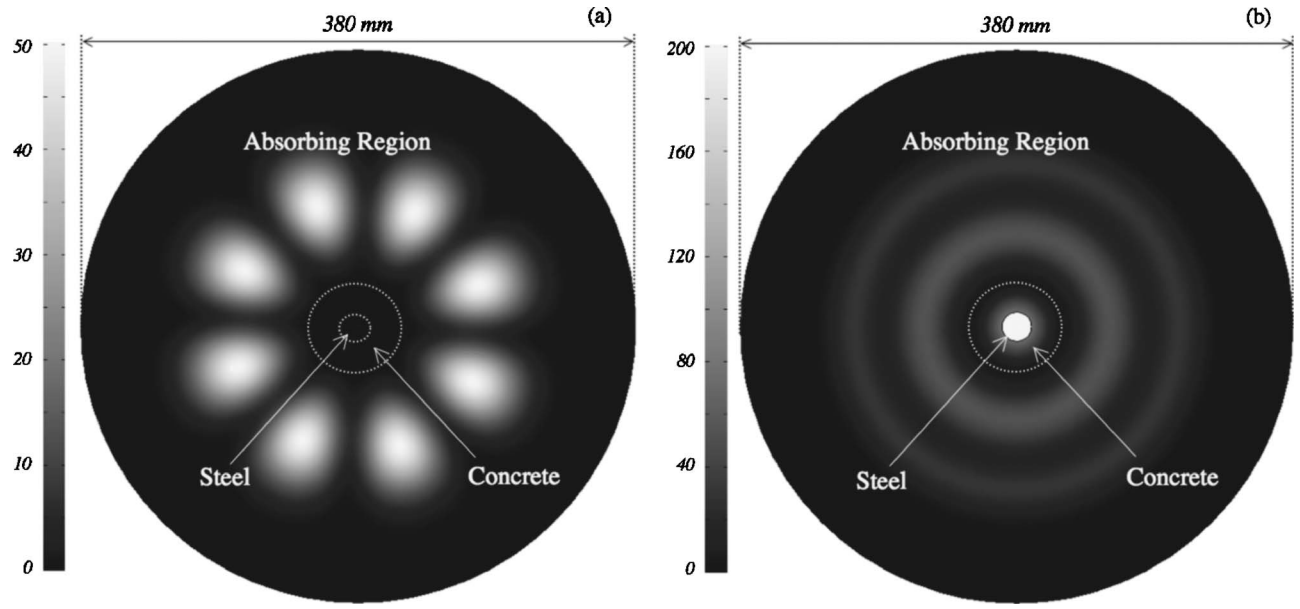


FIG. 6. Cross-section distribution of axial energy-flow (in W/mm^2) at 50 kHz for example modal results: (a) Mode resonating in the absorbing region and (b) mode resonating in the steel bar and radiating into the concrete (white=high energy-flow, black=low energy-flow).

attenuation characteristics of the leaky waves. However, SAFE points have not been found for some of the regions of the curves, and in some places points are plotted but are not in very good agreement, especially for the attenuation in Fig. 7(b). Examination of Fig. 7(a) reveals the conditions when the comparison is not good. Here the velocities of bulk longitudinal and shear waves in the concrete have been added to the plot, as horizontal lines; these are significant because leaky waves whose velocities are just above these values have large angles of radiation θ_{rad}, and we have noted earlier that the absorbing regions do not perform well in such cases. Furthermore, the size of the absorbing region for this model was limited in order to preserve a manageable model size. Thus we can see that conditions where the SAFE solutions lose their accuracy are, as expected, when the radiation angle is large and when the wavelength is long (low frequency).

We can conclude from this and the previous example that the SAFE calculation procedure is validated, provided that the absorbing region is large enough to prevent reflections from the outer boundary. It follows that practical use of the method, as it stands, is going to require restrictions to the range of applicability, to exclude solutions with large angles of radiation. In the longer term, it is hoped that this may be overcome by further development of the absorbing regions methodology.

IV. EXAMPLES OF RELEVANCE TO NDT INTERESTS

Here, we present two example applications which aim to illustrate the potential use of the technique to solve complex waveguide problems of relevance to NDT applications. These two examples are both too complex to solve by the conventional models, yet typical of the interests of researchers developing advanced guided wave NDT methods. These are presented here simply as examples; there is no intention in this paper to draw conclusions relating to the specific NDT opportunities.

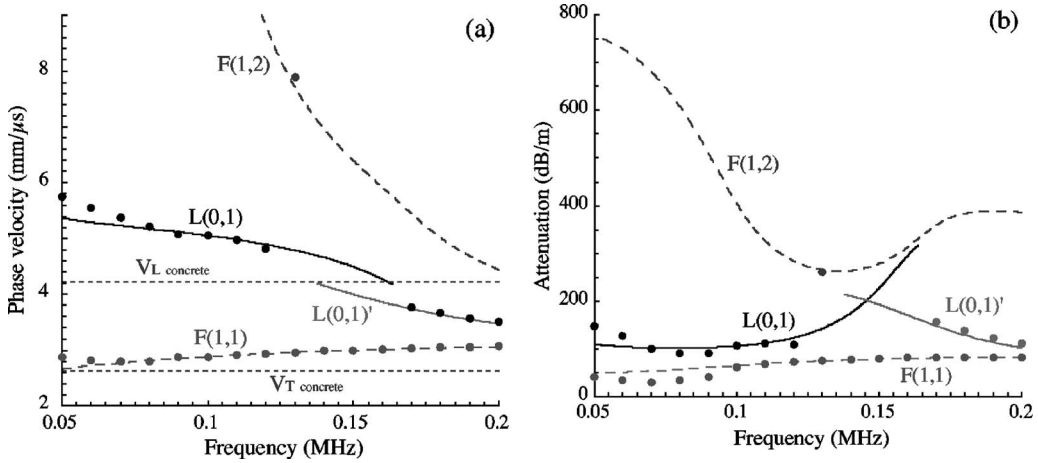


FIG. 7. Comparison of SAFE calculations (•••) with conventional matrix-method solutions for order zero modes (—) and order one modes (---).

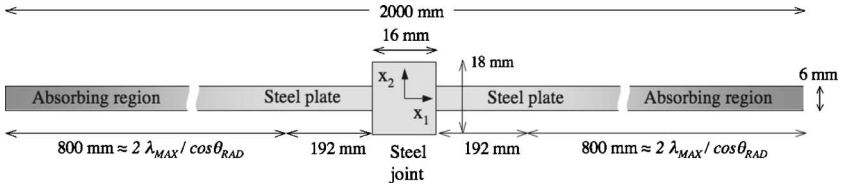


FIG. 8. Schematic of two-dimensional model of welded steel plates.

A. Welded joint between two steel plates

This study concerns waves which propagate along a butt weld between steel plates of the same thickness, inspired by the observation by Sargent[41] that waves traveling along the weld attenuate much less than waves propagating in an open region of the plate. Thus the weld appears to act as a waveguide, albeit with some attenuation due to leakage into the adjacent plates. A model system, consisting of a thickened rectangular region to represent the weld, was used as a simple approach to examine the physics of the phenomena. In this work we present a leaky wave SAFE analysis of the problem. Juluri *et al.*[47] have also performed conventional time-stepping FE simulations of wave propagation in this geometry, similarly demonstrating the guiding effect.

1. Geometry, material properties, and FE input data

The geometry of the system is presented by Fig. 8. The plates are made of steel and they both have a thickness equal to 6 mm. The idealized weld is also made of steel and it has a rectangular cross section of 18 mm thick and 16 mm wide. Waves which propagate along the weld would in general be expected to leak some energy out into the adjacent plates; thus the SAFE model of this problem must have outer boundaries which absorb these leaking waves without reflection. The absorbing regions, one at each extreme of the model, are shown in Fig. 8. A series of numerical tests covering a range of sizes of the absorbing regions, in which the acceptance criterion was convergence of the complex solution k, confirmed that the width of each of these regions had to be equal to about twice the projection along the x_1 axis of the maximum wavelength of the waves that could be radiated along the plates. Since these plates have a finite thickness, these radiated waves must necessarily be guided modes, i.e. Lamb-like or *SH*-like modes, not bulk modes.

The maximum possible wavelength in the adjacent plate is that of the S_0 mode at the lowest frequency of the investigated frequency range, i.e., ≈180 mm at 30 kHz, and its projection along the x_1 axis will depend on the angle at which it is radiated. Since these data are unknown, it is again difficult to set the length of the absorbing regions, but by choosing a limit of valid solutions to angles of radiation comprised between 0° and 65° allows the projected wavelength to be comprised between about 180 and 400 mm, according to Eq. (8). The length of the absorbing regions was therefore set equal to 800 mm, and the width of the cross section of the system (plate–weld–plate) was then equal to 2 m, including the absorbing regions. For S_0 modes radiated at angles greater than 65°, the 800-mm-long absorbing regions are likely to be inefficient, and thus we would then expect erroneous eigenvalue solutions. For the A_0 or SH_0 modes that may be radiated in the plates, which have maximum wavelengths equal to 42 and 108 mm, respectively, at 30 kHz, the 800 mm length of each absorbing and region is greater than twice the projection along x_1 of these maximum wavelengths, up to radiation angle limits of 84° and 74°, respectively.

Continuity of displacements and stresses are imposed at the internal borders between the steel plates and the steel joint. Stress-free conditions are imposed at the outer limit of the system. The whole geometry is meshed by 1800 rectangular elements of second order, with side lengths of 1.5 mm in direction x_1 and 2.2 mm in direction x_2, for each lateral plate, and 120 rectangular elements of second order, with side lengths of 1.6 mm in direction x_1 and 1.5 mm in direction x_2, for the weld. The number of degrees of freedom is 100 350, and Eq. (2) is solved for 28 frequencies in the range from 30 to 300 kHz. For each frequency, several solutions k are obtained, those unwanted representing resonances of the whole system or of the plates, and those sought corresponding to resonances of the steel joint and representing modes guided along the weld and radiating to infinity in the lateral steel plates.

2. Results

In order to select the solutions of interest among all the solutions obtained at each frequency, the axial component of the energy flow is calculated along a cross-section line run-

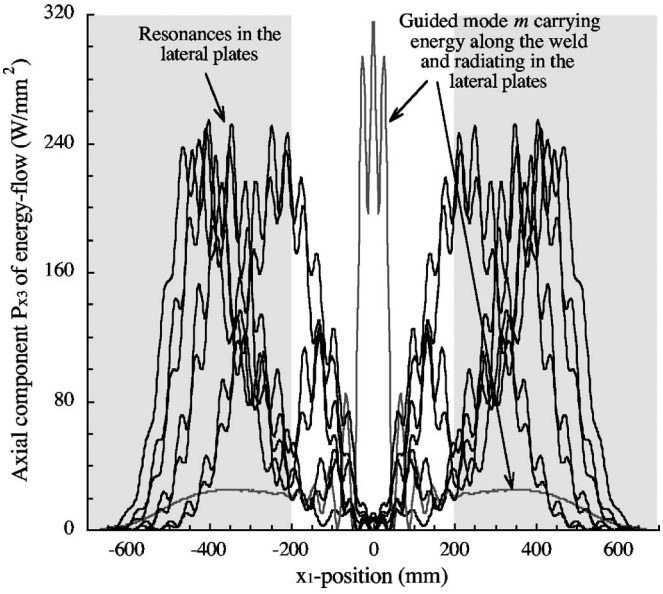


FIG. 9. Cross-section distribution of axial energy flow for several eigenvalue solutions obtained for the welded steel plates, at frequency 100 kHz. The solid line shows a weld-guided leaky wave, while the dashed lines show unwanted plate resonance solutions. The grey zones indicate the absorbing regions.

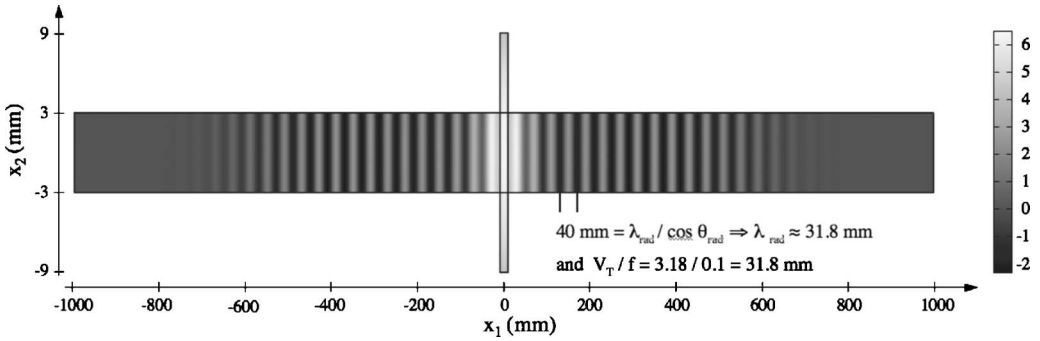


FIG. 10. Axial displacement snapshot for mode m of Fig. 9, guided along the weld joint and radiating in the steel plates, at frequency 100 kHz.

ning from $x_1=-700$ mm to $x_1=700$ mm, at $x_2=0$ mm, and high attenuation solutions neglected, in exactly the same way as was explained earlier in Sec. III B.

Figure 9 presents, for example, the distributions of the axial energy-flow component for several solutions obtained at 100 kHz. It is clear that one mode, labeled "m," carries energy mainly along the weld and leaks part of this energy in the lateral plates. This mode has relatively high values of energy flow in the weld, with decay of the energy flow toward the outsides of the absorbing regions. For clarity, this mode is shown by a solid line, while the other solutions are plotted as dashed lines.

Figure 10 shows a snapshot of the axial displacement of mode m, at 100 kHz, obtained by plotting the real part of the axial displacement from the eigenvector results. The eigenvalue for this solution is $k_m=0.1198-I8.145\times10^{-4}$ rad/mm, from which the corresponding phase velocity is: $V_m=2\pi f/0.1198=5.24$ mm/μs. According to the Snell–Descaries law,[45] only modes of the lateral plates having phase velocities smaller than V_m could be radiated in these plates. Thus the S_0 mode, with its phase velocity of 5.36 mm/μs, cannot radiate, while in principle the other two fundamental modes, A_0 and SH_0 could. However, since the mode "m" is symmetric with respect to the midplane of the plates and weld, the A_0 mode, which is antisymmetric, cannot be launched. Therefore, the SH_0 mode is the only mode which the m mode can leak into the plates, and its radiation must be at an angle equal to $\theta_{\text{rad}}=\sin^{-1}(3.18/5.24)\approx37°$, with respect to the direction normal to the plates–weld interfaces. Its wavelength at the frequency 100 kHz would then be equal to $\lambda_T=3.18/0.1=31.8$ mm. The wavelength λ_{x1} is easily found from the image in Fig. 10 to be 40 mm, and it is easy to check that this corresponds to the projection of the wavelength of SH_0 along the x_1 direction, using Eq. (8): 40 mm$\approx\lambda_T/\cos 37°$, thus proving that the radiated mode is effectively the SH_0 mode.

Figure 11 presents the dispersion curves for the mode m, found by repeating the SAFE solutions over a range of frequencies, and plotted as dots. In Fig. 11(a), the wave numbers are compared to those of the S_0, A_0, and SH_0 Lamb modes that would propagate along a free steel plate of the same thickness as the weld, i.e., 18 mm, and which have been predicted using a standard method.[8–10] The SAFE solutions indicate that the mode in the weld is similar to the S_0 Lamb mode in the 18-mm-thick plate. This is confirmed by the similarities between the through-thickness displacement shapes of these two modes, as shown for example in Fig. 10 by the axial displacement which is similar to the in-plane displacement produced by a S_0 Lamb mode in a free plate.

Figure 11(b) displays the attenuation of the mode m due to leakage of energy into the adjacent plates. This attenuation curve shows an interesting shape, specifically two peaks at 110 and 170 kHz, a dip at 140 kHz, and zero values above 190 kHz. In order to investigate further the nature of the

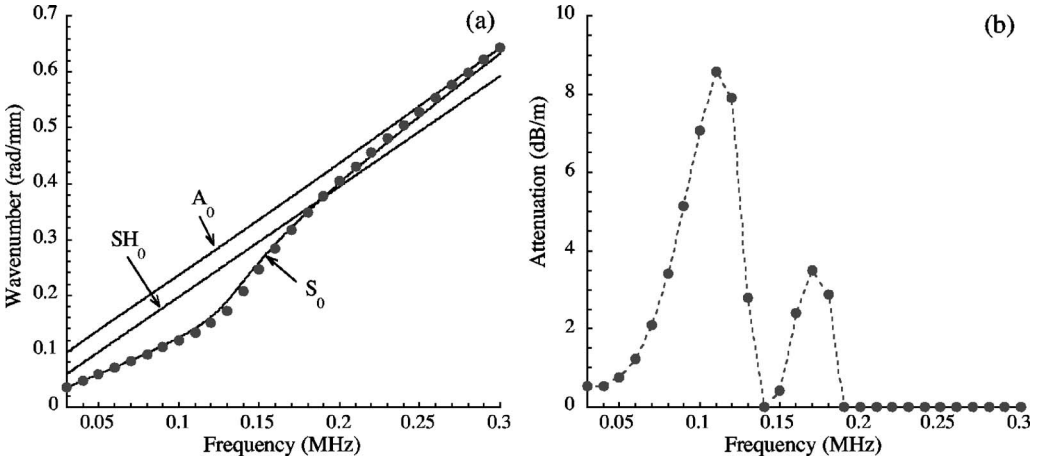


FIG. 11. Comparison between solutions predicted by SAFE method for the S_0-like mode (mode m in Fig. 9) guided along the weld joint and leaking into the plates (•••) with conventional solutions for S_0, A_0, and SH_0 modes propagating along a free plate of the same thickness as the weld (—); (a) wave number and (b) attenuation.

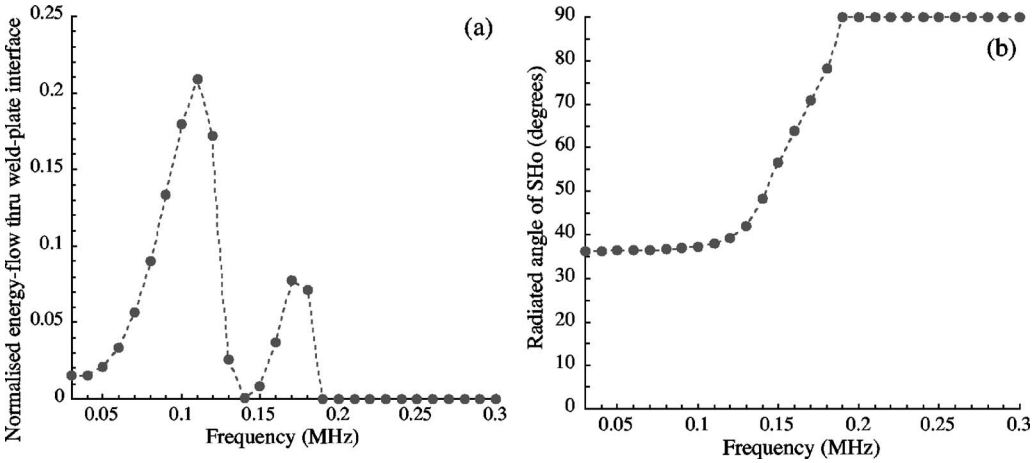


FIG. 12. (a) x_1 component of energy-flow transmitted through one plate-weld interface, normalized by the whole axial energy flow of the mode, calculated for the S_0-like mode (mode m in Fig. 9); (b) radiating angle of the SH_0 mode in the adjacent plates.

results, the x_1 component of the energy flow has been monitored along one of the two plate–weld interfaces, from $x_2=-3$ mm to $x_2=3$ mm at $x_1=8$ mm, and normalized by the axial component of the energy flow of the mode, obtained by integrating the energy flow vector over the whole cross section of the weld. This calculation has been repeated over the range of frequencies. As shown in Fig. 12(a), the variation with frequency of this normalized energy-flow x_1 component is very similar to that of the attenuation, thus showing, as should be expected, that the attenuation of the mode guided along the weld is governed by the displacements, stresses, and therefore energy flow normal to the plate–weld interfaces. The zero attenuation above 190 kHz is explained by the Snell–Descartes' law, which does not allow the SH_0 mode to be radiated above this frequency because its velocity is greater than that of the S_0-like mode propagating along the weld. Figure 12(b) displays the variation with frequency of the angle of radiation of this SH_0 mode. It clearly appears that this mode is no longer radiated for frequencies greater than 190 kHz.

From the predicted wave numbers shown in Fig. 11(a), the angles of radiation θ_{rad} shown in Fig. 12(b), and the phase velocity of the radiated SH_0 mode, the projection along the x_1 axis, i.e. λ_{x1}, of the wavelength λ_{rad} of the radiated SH_0 mode has been calculated, at each frequency of the FE calculations, using Eq. (8). These results are not shown here but the values obtained for λ_{x1} do not exceed 100 mm [the highest angle of leakage in the calculated results is the point at 78° in Fig. 12(b); this corresponds to $\lambda_{x1}=82$ mm], thus confirming that the 800-mm-long absorbing regions have been chosen to be long enough to be sure that they are effective in absorbing the leaking waves.

B. Stiffened plate

This last study concerns the case of a very large Perspex plate with a steel stiffener bonded onto one face. The purpose of this study is to find complex eigenvalue solutions that represent guided wave modes propagating along the stiffened region and radiating energy into the plate. A comparison of results for two different states of bonding is used to illustrate the potential to use such feature-guided waves to detect defects.

1. Geometry, material properties, and FE input data

The geometry of the system is presented in Fig. 13. The Perspex plate has a thickness equal to 4 mm and its density and viscoelastic properties are given in Table I. The stiffener has a T shape and it is made of steel, the mechanical properties of which are also given in Table I, as are the properties of the 1-mm-thick layer of adhesive between the plate and the stiffener. In order to suppress unwanted reflections, an absorbing region is modeled at each side of the Perspex plate, in a similar manner to the weld case study (Sec. IV A). The width of the absorbing regions was chosen in the same manner as in the other examples. In this example, the maximum possible wavelength of leaky waves is that of the S_0 mode at the lowest frequency of the investigated frequency

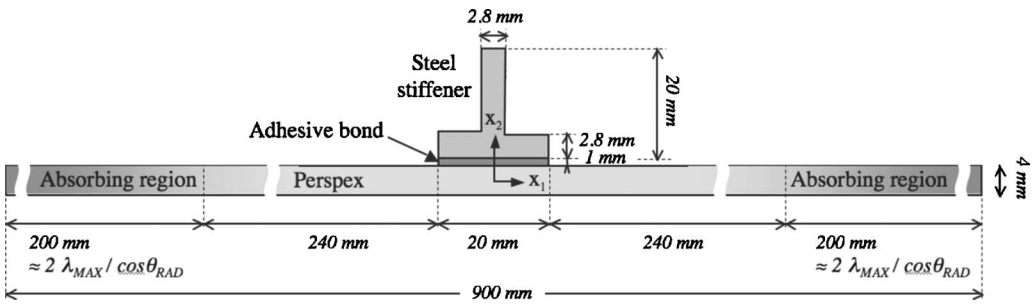


FIG. 13. Schematic of 2D model of Perspex plate with steel stiffener.

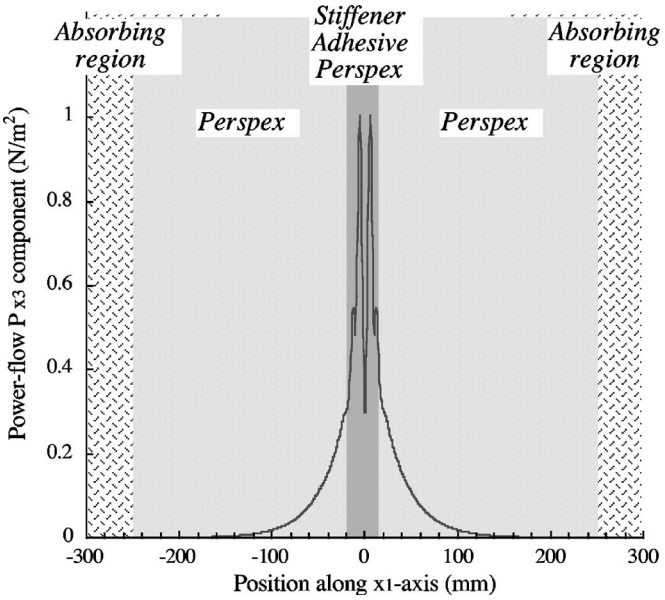


FIG. 14. Cross-section distribution of axial energy flow for one selected eigenvalue solution corresponding to a wave mode propagating energy along the stiffener-bond-Perspex region and radiating energy in the Perspex plate, at 100 kHz.

range, i.e., ≈36 mm at 50 kHz. The length of the absorbing regions was thus set equal to 200 mm, thus allowing good absorption of the three possible radiating modes S_0, A_0, and SH_0 at almost any possible angles between 0° and 70°, in the frequency range of interest, which was chosen from 50 to 200 kHz.

Continuity of displacements and stresses are imposed at the internal borders between the steel stiffener, the adhesive layer, and the Perspex plate. Stress-free conditions are defined at all outer limits of the system. The whole geometry is meshed by 2140 rectangular elements of second order having different sizes depending on the zone which is meshed: For example, 1 mm by 2.2 mm elements were used in the lateral sides of the Perspex plate (out of the region situated underneath the stiffener) and 0.67 mm by 0.25 mm elements were used in the adhesive layer. The various element sizes have been chosen to ensure good spatial representation of the field in the cross section of the system, and also after a series of numerical trials indicates a good convergence of the FE solutions. The number of degrees of freedom is 56 934, and Eq. (2) is solved for seven frequencies in the range from 50 to 200 kHz. For each frequency, several solutions k are obtained, including the wanted waveguide modes and the unwanted resonances of the whole system.

2. Results

The solutions of interest were found in the same way as in the previous examples. In this case the axial component of the energy flow was calculated along a cross-section line running from $x_1=-300$ mm to $x_1=300$ mm, at $x_2=0$ mm. Figure 14 presents the distribution of the axial energy-flow component for one selected solution, at 100 kHz. In this example, it is clear that the absorbing regions do not play any role since the mode radiation is totally absorbed in the Perspex plate due to its viscoelasticity. However, the use of the absorbing regions was required for other frequencies in this model. The eigenvalue for this mth solution is $k_m=0.412-i7.1\times10^{-5}$ rad/mm, from which the corresponding phase velocity is 1.52 mm/μs. The possible radiating waves in the Perspex plate can thus be calculated to be the A_0 or SH_0 modes, at radiation angles of about 32° and 42°, respectively.

Figure 15(a) presents the displacement shape of this mode, at the frequency 100 kHz. This mode shape indicates flexural behavior of the stiffener, and Fig. 15(b) shows that it also exists for the single stiffener, i.e., if the stiffener is uncoupled from the Perspex. The oscillatory pattern of the plate shown in Fig. 15(a) indicates the radiation of guided modes. Figure 16 shows the variations of the phase velocity and of the attenuation of this mode versus the frequency, for two qualities of the bond. The first one, considered as a good quality bond, is modeled using the adhesive property data (ρ, C_{11}, and C_{66}) given in Table I, while the second one, considered as a bad quality bond, is modeled by dividing the values of C_{11} and C_{66} by a factor of 10^5, the density being kept constant, thus simulating an extremely weak bond. It is clear

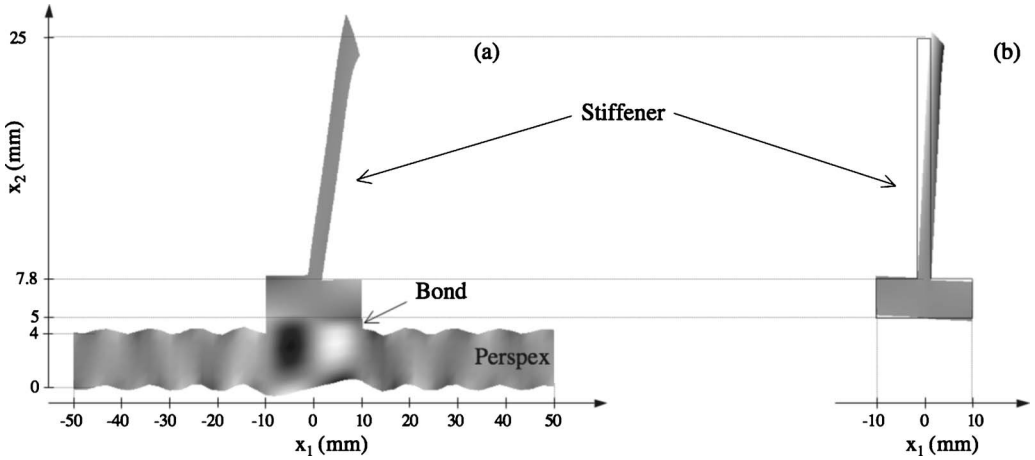


FIG. 15. Displacement shape of flexural mode of Fig. 14 guided along (a) the stiffener-bond-Perspex region and radiating in the Perspex plate and (b) the stiffener alone, at frequency 100 kHz; gray scale represents axial displacement component while the in-plane cross-section motion is shown by the deformed domains.

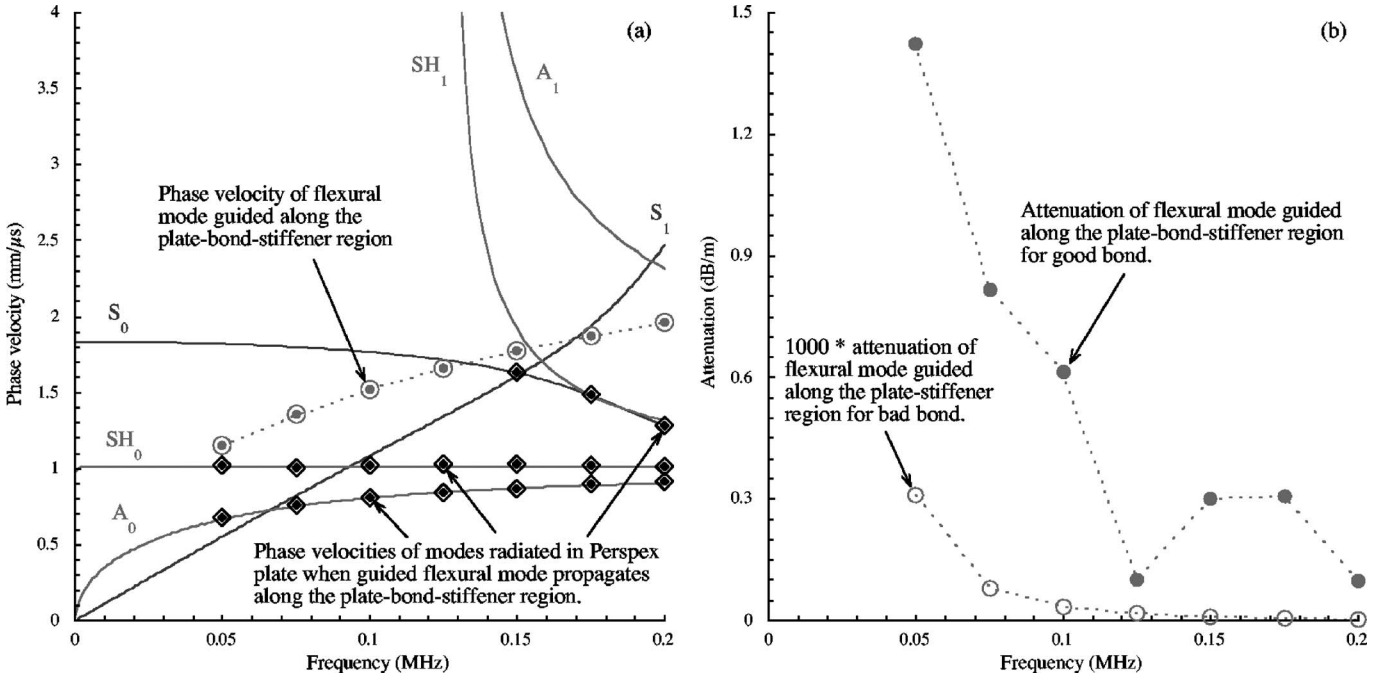


FIG. 16. (a) Phase velocity and (b) attenuation vs frequency for flexural mode of Fig. 14 guided along the stiffener-bond-Perspex region, modeling a good (●) and a bad (○) quality for the bond; In (a), dispersion curves of guided modes along free Perspex plate are also plotted (—) and compared to data (♦ for good bond and ◇ for bad bond) obtained by postprocessing displacements monitored in the Perspex plate in the SAFE model.

that the phase velocity is not sensitive to the change in bond quality. However, the attenuation shows strong sensitivity: It is almost equal to zero for the very weak bond (no leakage of the flexural mode in the Perspex plate) while it takes significant values (up to 2 dB/m) for the good bond quality.

Following the manner of the analysis of the results in case study (*C*), the x_1 components of the wave numbers of the leaky waves were extracted from the displacement solutions, and then resolved, using the knowledge of the wave numbers along the propagation direction, x_3, to yield the wave numbers in the leakage directions. In this case the field of leaky waves was composed of all three modes S_0, A_0, and SH_0, so their contributions had to be separated by a spatial Fourier transform. The leaky wave velocities thus found are plotted in Fig. 16(a), confirming perfect agreement with the Lamb wave solutions for the Perspex plate. It is interesting to observe that results for leakage of the S_0 mode in the Perspex plate do not appear below 0.15 MHz, and this is because its phase velocity is greater than that of the stiffener-guided mode.

As in the previous studies, the projections of the wavelengths of the leaky waves onto the x_1 direction were used to check that the size of the absorbing region was large enough. None of these wavelengths exceeded 100 mm, thus validating the choice of the length of the absorbing regions.

V. CONCLUSIONS

The SAFE method has been implemented for the study of elastic waveguide problems in which the guided waves may leak into an infinite surrounding solid material. This has been achieved by using established absorbing region modeling techniques in order to absorb the leaking waves and thus simulate an infinite extent of the surrounding material. The resulting SAFE eigensolutions in wave number are complex, the attenuation characteristics of the leaking guided waves being represented by the imaginary part of the wave number. The technique has been implemented and validated using two problems of simple geometry with known alternative solutions, and then applied to two example problems of relevance to NDT. The technique has been shown to work well, but there remains a problem if there are leaking waves of large leakage angle, when the absorbing region may not perform well enough. Therefore careful choice of the absorbing region is essential.

[1]P. Cawley, "The rapid non-destructive inspection of large composite structures," Composites **25**, 351–357 (1994).
[2]D. Alleyne and P. Cawley, "Long range propagation of Lamb waves in chemical plant pipework," Mater. Eval. **55**, 504–508 (1997).
[3]K. Diamanti, J. M. Hodgkinson, and C. Soutis, "Detection of low-velocity impact damage in composite plates using Lamb waves," Struct. Health Monit. **3**, 33–41 (2004).
[4]K. S. Tan, N. Guo, B. S. Wong, and C. G. Tui, "Experimental evaluation of delaminations in composite plates by the use of Lamb waves," Compos. Sci. Technol. **53**, 77–84 (1995).
[5]W. Hassan and P. B. Nagy, "On the low-frequency oscillation of a fluid layer between two elastic plates," J. Acoust. Soc. Am. **102**, 3343–3348 (1997).
[6]J. G. Harris, "Rayleigh wave propagation in curved waveguides," Wave Motion **36**, 425–441 (2002).
[7]A. L. Shuvalov, O. Poncelet, and M. Deschamps, "General formalism for plane guided waves in transversely inhomogeneous anisotropic plates," Wave Motion **40**, 413–426 (2004).
[8]M. J. S. Lowe, "Matrix techniques for modelling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[9]B. Pavlakovic, M. Lowe, D. Alleyne, and P. Cawley, *DISPERSE: A General Purpose Program for Creating Dispersion Curves*, Review of Progress in Quantitative NDE, Vol. **16**, edited by D. Thompson and D. Chimenti (Plenum, New York, 1987), pp. 185–192.
[10]B. Hosten and M. Castaings, "Surface impedance matrices to model the propagation in multilayered media," Ultrasonics **41**, 501–507 (2003).

[11]P. J. Shorter, "Wave propagation and damping in linear viscoelastic laminates," J. Acoust. Soc. Am. **115**, 1917–1925 (2004).
[12]L. Gavric, "Finite element computation of dispersion properties of thin walled waveguides," J. Sound Vib. **173**, 113–124 (1994).
[13]L. Gavric, "Computation of propagative waves in free rail using a finite element technique," J. Sound Vib. **185**, 531–543 (1995).
[14]U. Orrenius and S. Finnveden, "Calculation of wave propagation in rib stiffened plate structures," J. Sound Vib. **198**, 203–224 (1996).
[15]T. Mazuch, "Wave dispersion modelling anisotropic shells and rods by the finite element method," J. Sound Vib. **198**, 429–438 (1996).
[16]S. Finnveden, "Spectral finite element analysis of the vibration of straight fluid-filled pipes with flanges," J. Sound Vib. **199**, 125–154 (1997).
[17]S. Dong and R. Nelson, "On natural vibrations and waves in laminated orthotropic plates," J. Appl. Mech. **39**, 739–745 (1972).
[18]R. Nelson and S. Dong, "High frequency vibrations and waves in laminated orthotropic plates," J. Sound Vib. **30**, 33–44 (1973).
[19]S. Dong and K. Huang, "Edge vibrations in laminated composite plates," J. Appl. Mech. **52**, 433–438 (1985).
[20]S. Datta, A. Shah, R. Bratton, and T. Chakraborty, "Wave propagation in laminated composite plates," J. Acoust. Soc. Am. **83**, 2020–2026 (1988).
[21]Z. Xi, G. Liu, K. Lam, and H. Shang, "Dispersion and characteristic surfaces of waves in laminated composite circular cylindrical shells," J. Acoust. Soc. Am. **108**, 2179–2186 (2000).
[22]J. M. Galan and R. Abascal, "Numerical simulation of Lamb wave scattering in semi-infinite plates," Int. J. Numer. Methods Eng. **53**, 1145–1173 (2002).
[23]J. M. Galan and R. Abascal, "Lamb mode conversion at edges. A hybrid boundary element-finite-element solution," J. Acoust. Soc. Am. **117**, 1777–1784 (2005).
[24]T. Hayashi, W.-J. Song, and J. L. Rose, "Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example," Ultrasonics **41**, 175–183 (2003).
[25]B. R. Mace, D. Duhamel, M. J. Brennan, and L. Hinke, "Finite element prediction of wave motion in structural waveguides," J. Acoust. Soc. Am. **117**, 2835–2843 (2005).
[26]C. H. Yang, H. Huh, and H. T. Hahn, "Investigation of effective material properties in composites with internal defect or reinforcement particles," Int. J. Solids Struct. **42**, 6141–6165 (2005).
[27]A. Chakraborty and S. Gopalakrishnan, "A spectrally formulated finite element for wave propagation analysis in layered composite media," Int. J. Solids Struct. **41**, 5155–5183 (2004).
[28]J. M. Mencik and M. N. Ichchou, "Multi-mode propagation and diffusion in structures through finite elements," Eur. J. Mech. A/Solids **24**, 877–898 (2005).
[29]D. Roy Mahapatra and S. Gopalakrishnan, "A spectral finite element model for analysis of axial–flexural–shear coupled wave propagation in laminated composite beams," Compos. Struct. **59**, 67–88 (2003).
[30]V. Damljanovic and R. L. Weaver, "Propagating and evanescent elastic waves in cylindrical waveguides of arbitrary cross section," J. Acoust. Soc. Am. **115**, 1572–1581 (2004).
[31]M. V. Predoi, M. Castaings, B. Hosten, and C. Bacon, "Wave propagation along transversely periodic structures," J. Acoust. Soc. Am. **121**, 1935–1944 (2007).
[32]I. Bartoli, A. Marzani, F. Lanza di Scalea, and E. Viola, "Modeling wave propagation in damped waveguides of arbitrary cross-section," J. Sound Vib. **295**, 685–707 (2006).
[33]P. Wilcox, M. Evans, O. Diligent, M. J. S. Lowe, and P. Cawley, "Dispersion and excitability of guided acoustic waves in isotropic beams with arbitrary cross-section," *Review of Progress in Quantum Nondestructive Evaluation*, edited by D. O. Thompson and D.E. Chimenti [AIP Conf. Proc. **21**, 203–210 (2002)].
[34]K. Jezzine and A. Lhemery, "Diffraction effects on ultrasonic guided waves radiated or received by transducers mounted on the section of the guide," *Review of Progress in Quantum Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti [AIP Conf. Proc. **25**, 134–141 (2006)].
[35]G. R. Liu and J. D. Achenbach, "A strip element method for stress analysis of anisotropic linearly elastic solids," J. Appl. Mech. **61**, 270–277 (1994).
[36]K. Jezzine, "Approche modale pour la simulation globale de contrôles non-destructifs par ondes elastiques guides," (Modal approach for the full simulation of nondestructive tests by elastic guided waves), Ph.D. thesis, University Bordeaux 1, No. 3241, CEA-R-6147, 2006, available at http://www-ist.cea.fr/publicea/exl-doc/200600007322.pdf, most recently viewed 10 November 2007.
[37]M. D. Beard, M. J. S. Lowe, and P. Cawley, "Ultrasonic guided waves for the inspection of tendons and bolts," ASCE J. Mater. Civ. Eng. **15**, 212–218 (2003).
[38]T. Vogt, M. J. S. Lowe, and P. Cawley, "Measurement of the material properties of viscous liquids using ultrasonic guided waves," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 737–747 (2004).
[39]F. B. Cegla, P. Cawley, and M. J. S. Lowe, "Fluid bulk velocity and attenuation measurements in non-Newtonian liquids using a dipstick sensor," Meas. Sci. Technol. **17**, 264–274 (2006).
[40]J. O. Kim and H. H. Bau, "On line, real-time densimeter—Theory and optimization," J. Acoust. Soc. Am. **85**, 432–439 (1989).
[41]J. P. Sargent, "Corrosion detection in welds and heat affected zones using ultrasonic Lamb waves," Insight **48**, 160–167 (2006).
[42]M. Castaings, C. Bacon, B. Hosten, and M. V. Predoi, "Finite element predictions for the dynamic response of thermo-viscoelastic material structures," J. Acoust. Soc. Am. **115**, 1125–1133 (2004).
[43]M. Drozdz, M. J. S. Lowe, P. Cawley, L. Moreau, and M. Castaings, "Efficient numerical modeling of absorbing regions for boundaries of guided waves problems," *Review of Progress in Quantum Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti [AIP Conf. Proc. **25**, 126–133 (2006)].
[44]"COMSOL, User's Guide and Introduction," Version 3.2 by–COMSOL AB 2005 http://www.comsol.com/, most recently viewed 20 June 2007.
[45]B. A. Auld, *Acoustic Fields and Waves in Solids* (R. Krieger, Malabar, FL, 1990).
[46]M. Drozdz, M. J. S. Lowe, E. Skelton, and R. V. Craster, "Modeling bulk and guided wave propagation in unbounded elastic media using absorbing layers in commercial FE packages," *Review of Progress in Quantum Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti [AIP Conf. Proc. **26**, 87–94 (2007)].
[47]N. Juluri, M. J. S. Lowe, and P. Cawley, "The guiding of ultrasound by a welded joint in a plate," *Review of Progress in Quantum Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti [AIP Conf. Proc. **26**, 1079–1086 (2007)].

Vibrational modes of free nanoparticles: From atomic to continuum scales

Fernando Ramirez[a)]
Universidad de los Andes, Bogotá, Colombia

Paul R. Heyliger[b)]
Department of Civil Engineering, Colorado State University, Fort Collins, Colorado 80523

Anthony K. Rappé[c)]
Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523

Robert G. Leisure[d)]
Department of Physics, Colorado State University, Fort Collins, Colorado 80523

(Received 10 September 2007; revised 7 November 2007; accepted 19 November 2007)

Vibration analysis of free standing silicon nanoparticles, with sizes ranging from 0.732 to 4.223 nm, are calculated using two different methods: molecular mechanics and classical continuum elasticity. Three different geometries are studied: cubes, spheres, and tetrahedrons. Continuum mechanics methods provide good estimates of the lowest natural frequency of particles having at least 836 ($R>1.5$ nm) and 800 ($R>1.28$ nm) atoms for cube- and tetrahedron-shaped nanostructures, respectively. Equations for vibrational frequencies of smaller particles as a function of size are proposed. The vibrational modes obtained from both methods were practically the same for the sphere- and tetrahedron-shaped particles with a large number of atoms. However, for the cube geometry only the shape of the modes corresponding to the lowest couple of frequencies occur in the same order. In general, vibrational modes shapes obtained using both methods are the same although the order in which they appear may be shifted. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2823065]

PACS number(s): 43.35.Gk, 43.40.Hb [RLW] Pages: 709–717

I. INTRODUCTION

In the past decade, significant advances have been achieved in the area of nanoengineering. The existence of numerous different types of nanostructures has been reported in the literature, including nanoparticles or quantum dots, single- and multiwalled nanotubes, nanowires, and nanoropes. One of the most important characteristics of nanostructures is the size dependence that the properties of these materials and components exhibit. Because of the extremely small size of these structures, the development of efficient methods to evaluate their electronic, mechanical, optical, and acoustical properties is a challenge to researchers.

Nanoparticles show optical and acoustic properties different from those of the bulk crystals of the same material. These differences are mainly attributed to the three-dimensional confinement of electrons and holes in small volumes known as quantum size effect,[1] and to the fact that most of the atoms are near the particle surface for these small sizes. Optical properties of nanoparticles have attracted much interest due to their application as optical processing devices and semiconductor lasers. Additionally, recent investigations predict extensive applications in biology, bioengineering, and medicine as biological tags and as active electrical contacts to neurons.[2] Several studies exist in which the acoustic modes of spherical nanoparticles embedded in a matrix are experimentally measured using Raman scattering. In these studies[1–7] frequency results were analyzed and compared to results obtained using continuum elasticity methods. Most of these comparisons were based on Lamb's theory for the vibration of elastic homogeneous spheres,[8] and good general agreement was reported. Calculations of Raman intensities of nanoscale germanium (Ge) and silicon (Si) quantum dots of varying sizes have been performed using a microscopic valence force field model.[9–11] These studies reported that when the size of the quantum dots increased, their highest frequencies approach the frequencies of the optical-phonon frequencies of bulk Si, and that the lowest frequencies are roughly proportional to the inverse of the quantum dots diameters.

Determination of the natural frequencies of nanostructures is important not only for applications in optical devices and lasers, but it also provides enough information to evaluate their elastic constants. Vibrational frequencies and modes of rectangular parallelepiped and cylinders are involved in resonant ultrasound spectroscopy to determine their elastic constants by matching the measured frequencies to those obtained solving the equations of motion in an iterative procedure.[12,13] Texture and orientation-distribution coefficients in polycrystalline brass were studied using a similar approach.[14]

[a)]Electronic mail: framirez@uniandes.edu.co
[b)]Electronic mail: prh@engr.colostate.edu
[c)]Electronic mail: rappe@lamar.colostate.edu
[d)]Electronic mail: leisure@lamar.colostate.edu

Different analytical methods are available for the determination of the properties of nanostructures. They can be divided into two main categories: atomistic models and elastic continuum model theories. Atomistic approaches include both quantum mechanics and molecular dynamics and mechanics. In principle, these are the most suitable methods to deal with molecular or atomic motions. However, the application of atomistic models is computationally expensive, and the application is limited to nanostructures formed by a small number of molecules or atoms. Conversely, continuum mechanics approaches are computationally much faster, but they do not explicitly consider atomic and electronic interactions that can affect the properties of these novel structures. At very small length scales, nanostructures cannot be considered continuous. The validity of classical continuum approaches must be evaluated and potentially modified.

In the present study, the natural vibration frequencies and modes of free standing silicon (Si) nanoparticles, with sizes ranging from 0.732 to 4.223 nm, were calculated using two different methods: molecular mechanics at the atomic level, and classical elasticity at the continuum level. Three different geometries are studied: cubes, spheres, and tetrahedrons. Once the frequency spectra have been obtained, vibrational modal shapes obtained from both methods are compared, and the presence and number of degenerate modes are analyzed. Equations describing the variation of the natural frequencies as a function of the particle size and number of atoms forming the nanoparticles are also constructed, for general use.

II. THEORY

A. The atomic model

Molecular mechanics is a relatively simple, empirically based "ball-spring" atomistic model for molecular structures. Atoms are connected by bonds that can be stretched or compressed due to intra- or intermolecular forces. The size of the atoms and the stiffness of the bonds are determined empirically by being chosen to reproduce experimental data. The force field/molecular mechanics model evaluates the potential energy, V, for a molecule as the sum of energies from two-; three-; and four-body interactions.[15] These energies are expressed as

$$V = V_r + V_\theta + V_\phi + V_\omega + V_{\mathrm{vdW}} + V_{\mathrm{el}}. \tag{1}$$

The first four energies on the right-hand side of Eq. (1) are related to bond stretching, bond angle bending, dihedral angle torsion, and inversion terms, respectively. The nonbonded energies include van der Waals and electrostatic contributions, represented by the final two terms on the right-hand side. The total potential energy gives a mathematical representation for how each atom would move under the influence of the motions of all other atoms in the system. Vibrational frequencies are determined using molecular mechanics by determining the equilibrium geometry from energy minimization, and then using a harmonic oscillator approximation to reproduce the vibrational frequencies. Expressions for the different terms that contribute to the total potential energy are functions of the atomic positions, interatomic equilibrium distances, and different parameters that are obtained and calibrated by fitting experimental data and quantum mechanics results.

In this study, the approximate pair theory[16] is used to calculate the vibrational spectra of particles at the nanoscale. This method uses parametric models of electronic interactions to permit reasonable descriptions of the electronic changes associated with nanoscale chemistry. This theory makes use of the force field known as the universal force field,[17] which is a set of simple functional forms and parameters used to model the structure, movement, and interactions of molecules. The parameters are defined empirically or by combining atomic parameters according to certain rules. A more detailed description of this process may be found elsewhere.[15–19] In this study, the focus is on Si particles for which the different energy terms have been presented by the authors to evaluate the size-dependent vibrational spectra of Si nanoparticles.[20] They are presented here for completeness.

For the bond stretch term, an extended Rydberg function[21,22] is used as given by

$$V_r = -D_e \mathrm{BO}_{IJ}, \tag{2}$$

where D_e is 85.0 kcal/mol. The bond order BO_{ij} between atoms i and j is given by

$$\mathrm{BO}_{ij} = e^{-4\alpha_{ij}(r-r_{ij})}\left[1 + 4\alpha_{ij}(r-r_{ij})\left[1 + \alpha_{ij}(r-r_{ij})\left[1 + \alpha_{ij}(r-r_{ij})\right]\right]\right]. \tag{3}$$

The bond distance parameter r_{ij} for Si is 2.330 Å, and the exponent α_{ij} is 0.6955 Å^{-1}. A bond-order-dependent cosine expansion is used for the bond angle bending term. Bond order mediation of angular deformation terms has previously been used for the water potential surface[23] and in the general Stillinger–Weber potential.[24] It is given by

$$V_\theta = \tfrac{1}{2} K_\theta \mathrm{BO}_{ij}\mathrm{BO}_{jk}\left(\cos\theta_{ijk} - \cos\theta_{0ijk}\right), \tag{4}$$

where the force constant K_θ is 43.382 kcal/mol and the equilibrium angle θ_{0ijk} is 109.471°. A bond-order-dependent threefold torsional potential is used and is given by

$$V_\phi = K_{ijkl}\mathrm{BO}_{ij}\mathrm{BO}_{jk}\mathrm{BO}_{kl}\left(1 - \cos 3\phi_{ijkl}\right). \tag{5}$$

The silicon torsion parameter, K_{ijkl}, is 1.225 kcal/mol. A Morse–Spline van der Waals potential is used for nonbonded interactions.[25] A Morse potential describes the inner wall region as well as the bottom of the well and is given as

$$V_{\mathrm{vdW}} = D_{ij}\left(e^{-2\beta_{ij}(x-x_{ij})} - 2e^{-\beta_{ij}(x-x_{ij})}\right). \tag{6}$$

The dissociation energy D_{ij} is 0.303 kcal/mol, the equilibrium distance is 4.41 Å, and the exponent β_{ij} is 1.266 Å^{-1}. Long-range interactions are described by an inverse-6 potential as expressed by

$$V_{\mathrm{vdW}} = \frac{C_6}{x^6}. \tag{7}$$

The C_6 coefficient, 3689.64 Å^6 kcal/mol, is selected to match the Morse potential for nonbonded interactions at a crossover point or cutoff distance, which is selected to provide continuous first derivatives.

The molecular mechanics calculations begin with the determination of the minimized or equilibrium molecular structures, which correspond to the lowest possible energy level. The equilibrium coordinates for each atom within the molecule are obtained by energy minimization using the Newton–Raphson minimization technique.[26] In the Newton–Raphson algorithm, both the gradient of the potential energy $\nabla \mathbf{V}$ (where we use boldface to denote a vector), and the second derivative or Hessian matrix, $\mathbf{H_V}$, are evaluated at the initial trial geometry $\mathbf{r}_o$. Since the Hessian matrix defines the curvature of the potential energy surface in each gradient direction, multiplication of this matrix by the gradient results in a vector which translates the system toward the minimum at $\mathbf{r}_{\min}$. The general expression describing this procedure is expressed as

$$\mathbf{r}_{\min} = \mathbf{r}_o - \mathbf{H}_\mathbf{V}^{-1} \nabla \mathbf{V}. \tag{8}$$

The evaluation of the natural vibration frequencies begins with the expression for the total kinetic energy (T) of the n-atom system, given by[26]

$$T = \frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{3} m_\alpha \left(\frac{d\Delta x_i^\alpha}{dt}\right)^2 = \frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{3} m_\alpha \left(\frac{q_i^\alpha}{dt}\right)^2 = \frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{3} \dot{q}_i^\alpha, \tag{9}$$

where m_α is the mass of atom α, Δx_i^α is its displacement in the jth coordinate direction relative to the equilibrium position, q_i^α is its mass-weighted displacement $\sqrt{m_\alpha}\Delta x_i^\alpha$, and t is time. For small vibrations, the potential energy may be expressed as a power series in the displacement q_i^α,

$$V = V_0 + \sum_{\alpha=1}^{n}\sum_{i=1}^{3}\left(\frac{\partial V}{\partial q_i^\alpha}\right)_o q_i^\alpha + \sum_{\alpha,\beta=1}^{n}\sum_{i,j=1}^{3}\left(\frac{\partial^2 V}{\partial q_i^\alpha \partial q_j^\beta}\right)_o q_i^\alpha q_j^\beta + O(x_i)^3. \tag{10}$$

By selecting the zero energy at the equilibrium configuration and considering that the corresponding gradient of the potential energy is a minimum, V_0 and $(\partial V/\partial q_i^\alpha)_o$ may be respectively eliminated. Neglecting the higher order terms under a small displacement assumption, Eq. (10) becomes

$$V = \sum_{\alpha,\beta=1}^{n}\sum_{i,j=1}^{3}\left(\frac{\partial^2 V}{\partial q_i^\alpha \partial q_j^\beta}\right)_o q_i^\alpha q_j^\beta. \tag{11}$$

The Euler–Lagrange equation is given by

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_j^\alpha} + \frac{\partial V}{\partial q_j^\alpha} = 0. \tag{12}$$

Since T is a function of the velocities only, and V is a function of the coordinates only, substitution of them in Eq. (12) yields

$$\ddot{q}_j^\alpha + \sum_{\alpha,\beta=1}^{n}\sum_{i,j=1}^{3}\left(\frac{\partial^2 V}{\partial q_i^\alpha \partial q_j^\beta}\right)_o q_i^\alpha = 0. \tag{13}$$

This is a system of $3n$ simultaneous second-order linear differential equations. The terms $(\partial^2 V/\partial x_i^\alpha \partial x_j^\beta)$ are constants evaluated at the equilibrium configuration. Assuming harmonic motion, Eq. (13) takes the form of an eigenvalue problem that can be solved using conventional methods, i.e.,

$$[F]\{x\} = \omega^2[M]\{x\}, \tag{14}$$

where ω and the vector $\mathbf{x}$ represent the frequencies and corresponding eigenvectors, respectively, and the elements of the stiffness matrix $[F]$, and the diagonal mass matrix $[M]$ are given by

$$F_{ij}^{\alpha\beta} = \left(\frac{\partial^2 V}{\partial q_i^\alpha \partial q_j^\beta}\right)_o, \tag{15}$$

$$M_{ii}^\alpha = m_\alpha. \tag{16}$$

B. The continuum model

The continuum approach to compare natural frequencies begins with the equations of motion for a general anisotropic linearly elastic solid given by

$$\sigma_{ij,j} + f_i = \rho\frac{\partial^2 u_i}{\partial t^2}. \tag{17}$$

Here σ_{ij}, f_i, ρ, u_i, and t denote the components of the stress tensor, the vector components of the body force, the mass density, the components of the displacements, and time, respectively. Following the standard formulation used in variational solid mechanics methods, the equations of motion are multiplied by arbitrary functions that represent the virtual displacements δu_i, integrating over the volume V of the solid, integrating by parts, and applying the divergence theorem yields the weak form of the equations of motion as

$$\int_V \left(\sigma_{ij}\delta\epsilon_{ij} - \delta u_i f_i + \delta u_i \rho\frac{\partial^2 u_i}{\partial t^2}\right)dV - \oint_S t_i \delta u_i dS = 0, \tag{18}$$

where t_i are the vector components of the surface tractions, and ϵ_{ij} are the components of the infinitesimal strain tensor.

A Ritz model[27] is used here to seek approximate solutions to the weak form by approximating the displacement components and their variations using finite linear combinations of the form

$$u_i(x,y,z,t) = \phi_0^{u_i}(x,y,z) + \sum_{j=1}^{n} a_j^{u_i}\phi_j^{u_i}(x,y,z)e^{i\omega t}. \tag{19}$$

Here ω is the natural frequency under the assumption of periodic motion, $\phi_j^{u_i}$ are known functions of position selected as power series in terms of the coordinate variables x, y, and z, n represents the number of terms in the approximation for the displacement components, and $a_j^{u_i}$ are constants determined by requiring that each of the variational statements holds for arbitrary variations of the displacements. Substitution of the strain-displacement and stress-strain relations, and the approximate displacement functions into the weak form of the governing equations, assuming the absence of body force, and collecting the terms corresponding to the coefficients of the variation of the displacements leads also to an eigenvalue problem to be solved for the natural frequencies and vibrational modes of the corresponding solid. Results obtained using this method have been compared with other

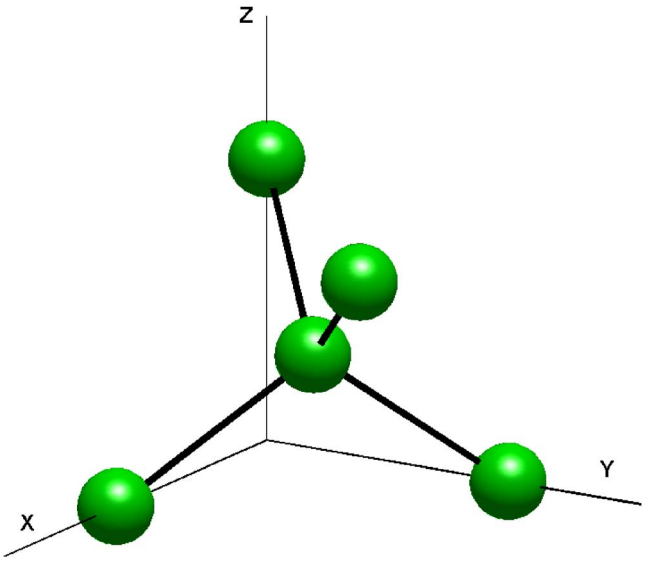


FIG. 1. (Color online) Orientation of the cubic crystal structure.

solutions, with excellent accuracy being obtained.[28,29]

The two methods used to evaluate the vibrational spectra of nanoparticles result in general eigenvalue problems whose solutions provide the natural frequencies and corresponding mode shapes for the particles considered. Although the final equations solved for all the presented methods have the same form, their components are different. The molecular mechanics method results in a diagonal mass matrix whose elements are the masses of each individual atom forming the nanostructure. In the continuum mechanics method, a consistent mass matrix is obtained. The stiffness matrices are also different for each method. For molecular mechanics, the stiffness matrix represents the interactions among particles at the atomic scale. The continuum mechanics stiffness matrix is a function of the continuum elasticity tensor, and its components basically represent the elastic strain energy stored in the solid when it undergoes deformation.

III. RESULTS

In this section, the calculated natural vibrational frequencies of cubic Si nanostructures using molecular mechanics and continuum mechanics are presented. The components of the bulk elastic stiffness tensor and mass density for cubic Si are:[20] $C_{11}=166.0$ GPa, $C_{12}=63.9$ GPa, $C_{44}=79.6$ GPa, and $\rho=2.329$ g/cm^3. These properties are referred to a coordinate system with an orientation relative to the crystal structure shown in Fig. 1.

Three different nanoparticle geometries were considered: spheres, cubes, and tetrahedra. The orientation of the different geometries relative to the Cartesian coordinate system in which they are described is crucial when considering materials having anisotropic behavior. It should match not only the orientation in which the elastic constants were determined, but also the crystal orientation used in the atomistic model. Spherical nanoparticles of radius R were defined with the center located at the origin of the coordinate system and with arbitrary orientation of the (x,y,z) axes. Cubes with side length L were considered with the origin located at one

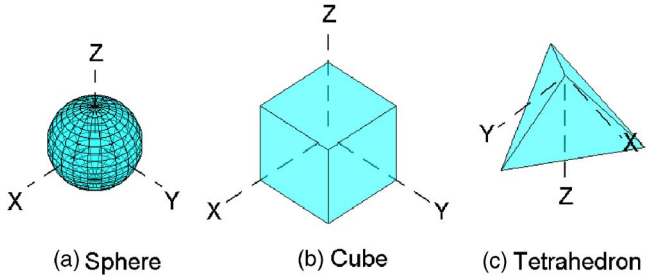


FIG. 2. (Color online) Shapes of nanostructures.

of the corners with the axes parallel to the sides. Tetrahedra with edge length a had the origin of the coordinate system located at one the tips of the particle, and they are oriented such that the z axis coincides with the height and the x axis is parallel to one of the sides of the triangle forming the base of the tetrahedron. These orientations and locations of the Cartesian coordinate systems were used in the continuum mechanics analyses, and they are shown in Fig. 2 for each of the different geometries. Note that the coordinate system for the tetrahedron does not coincide with that of the crystal (Fig. 1). As a consequence, the elastic stiffness tensor was transformed in order to have the components defined in such a way to be consistent with the crystal orientation used in the molecular mechanics computations.

The natural frequencies reported in this study are normalized as

$$\eta=\frac{\omega R}{C_t}, \tag{20}$$

where η represents the normalized frequencies, ω are the frequencies obtained from the analyses, $C_t=\sqrt{C_{44}/\rho}$ is the shear wave speed, and R is the radius of the sphere for the case of spherical nanoparticles. For the other geometries, R is the radius of an equivalent sphere having the same volume as that of the corresponding shape. For the case of cube-shaped nanoparticles, $R=L\sqrt[3]{3/4\pi}$ with L being the side length of the cube. For tetrahedron-shaped particles, $R=a\sqrt[3]{2^{1/2}/16\pi}$ with a being the edge length of the tetrahedron.

To study the size dependence of the natural frequencies and modal shapes, nanostructures having different dimensions and formed by different number of atoms were studied. For spheres, 123, 227, 477, and 933 atoms were considered. These correspond to spheres with radii 0.7328, 0.9200, 1.2846, and 1.5662 nm, respectively. Cube particles are formed by 64, 216, 512, and 1000 atoms, and have side lengths 0.9566, 1.4999, 2.0436, and 2.5884 nm. The edge lengths of the tetrahedrons are 1.1512, 1.9196, 2.6876, 3.4554, and 4.2230 nm resulting in particles formed by 51, 136, 281, 502, and 815 atoms, respectively.

A. Sphere-shaped nanostructures

The lowest 20 normalized natural frequencies for spherical nanoparticles obtained using molecular and continuum mechanics are listed in Table I. Results indicate that there is in fact a dependence on the nanostructure size of the normalized vibrational frequencies at the nanoscale. Not only the value of the normalized frequencies, but also the location of

TABLE I. Normalized frequencies for silicon spheres.

Atoms	123	227	477	933	Continuum
R (Å)	7.3280	9.2000	12.8463	15.6620	
1	1.5441	1.5465	1.9959	2.0415	2.1295
2	1.5848	1.5465	1.9959	2.0415	2.1295
3	1.5848	1.5465	1.9959	2.0415	2.1295
4	1.5848	1.6093	2.3137	2.2895	2.1320
5	1.7646	1.7145	2.3137	2.2895	2.1320
6	1.7646	1.7145	2.3603	2.3910	2.4780
7	2.0129	2.0738	2.3603	2.3910	2.4780
8	2.0129	2.0738	2.5835	2.6449	2.6109
9	2.0988	2.1062	2.7930	2.7467	2.6109
10	2.0988	2.1062	2.7930	2.7467	2.6109
11	2.0988	2.1062	2.7930	2.7467	2.9547
12	2.4550	2.3844	2.9139	2.9716	2.9547
13	2.4550	2.3844	2.9139	2.9716	2.9547
14	2.4550	2.3844	2.9139	2.9716	3.1308
15	2.5006	2.4258	3.1121	3.2225	3.4418
16	2.5006	2.4258	3.1121	3.2225	3.4418
17	2.5006	2.4258	3.1121	3.2225	3.4418
18	2.8712	2.4858	3.2168	3.4871	3.5273
19	2.8712	2.4858	3.2168	3.4871	3.5273
20	2.9472	2.4858	3.2168	3.4871	3.5273

the degenerate modes depend on the nanoparticle size. The normalized frequencies increase as the size of the particles is increased. For the largest spherical particle (R=1.5662 nm) some of the frequencies remain below and others surpass the values obtained from continuum mechanics, with the maximum difference being about ±7.0%. The behavior of the lowest frequency as a function of the number of atoms is illustrated in Fig. 3. Differences with respect to continuum values range from −47.1% for the smallest sphere (R =0.7328 nm) to −4.2% for the largest one (R=1.5662 nm). There is not a smooth variation with number of atoms or radius, and it is not possible to fit an acceptable equation to describe the size dependence of the natural frequencies.[20]

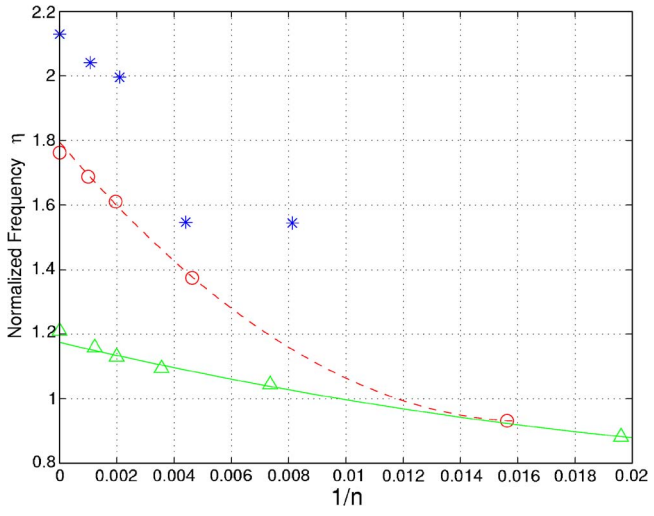


FIG. 3. (Color online) Lowest normalized natural frequency for sphere (*), cube (○), and tetrahedron-shaped (△) silicon nanostructures as function of the number of atoms (n). Dashed and solid lines represent the curve fitting for cubes and tetrahedrons, respectively. Frequencies at $1/n$=0 correspond to continuum mechanics results.

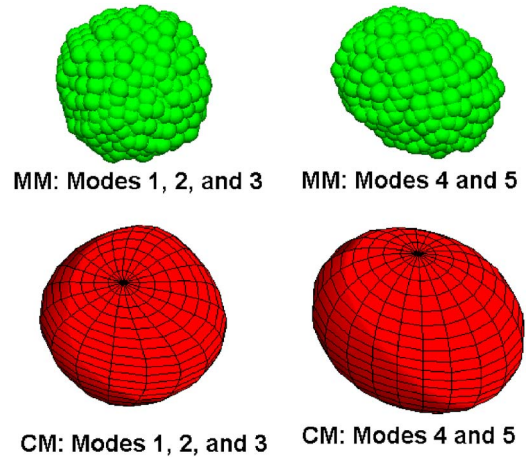


FIG. 4. (Color online) Identical vibrational modes of sphere-shaped nanostructures obtained using molecular mechanics (MM) and continuum mechanics (CM).

This behavior may be caused by the variation of the geometry of the spherical particles with the number of atoms. The geometry of the molecular structures modeled using molecular mechanics is not perfectly spherical. Their geometric shape changes with the number of atoms, and the molecule formed by the highest number atoms is the one with a geometry closest to a sphere. The reported radii were calculated as the average of the maximum and minimum values. These values correspond to the largest and smallest distances between pair of atoms located at the surface of the different sphere-shaped nanoparticles.

Most of the lower mode shapes of silicon spheres formed by 933 atoms are practically identical for both molecular and continuum mechanics approaches. The only difference is provided by the fact that the mode shape corresponding to the 8th molecular frequency matches the mode shape associated with the 14th continuum frequency. The modal shapes corresponding to the first four degenerate groups exhibiting the same deformation for both methods are shown in Figs. 4 and 5. Modal shape animations for all the particles considered in this study are available online.[30]

B. Cube-shaped nanostructures

Table II contains the lowest 20 normalized frequencies for cube-shaped particles calculated using molecular and continuum mechanics. Once again the normalized frequency size dependence is exhibited. The location of the degenerate modes also vary, and the normalized frequencies increase with the number of atoms. Molecular mechanics frequencies of silicon cubes approach the continuum values from below, with differences for the largest cube-shaped nanostructure (L=2.5884 nm) ranging between 4.4% for the lowest frequency and 13.8% for the 20th frequency.

The lowest frequency was 27.5% lower than that obtained using continuum mechanics for the smallest cube (L =0.9566 nm), and 4.1% lower for the largest particle (L =2.5884 nm). The behavior of the lowest frequency is illustrated in Fig. 3 as a function of the number of atoms (n). This time, the variation is smooth, and the second-order polynomial obtained using least-squares fitting is given by

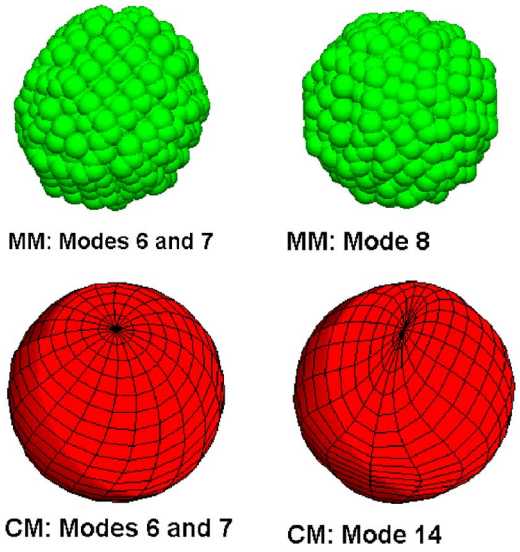


FIG. 5. (Color online) Identical vibrational modes of sphere-shaped nanostructures obtained using MM and CM.

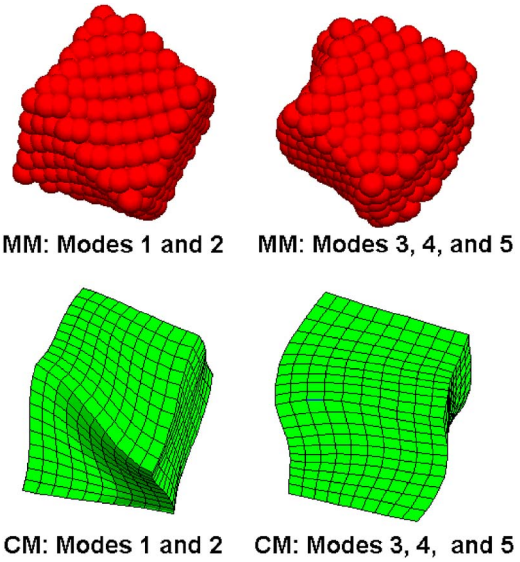


FIG. 6. (Color online) Identical vibrational modes of cube-shaped nanostructures obtained using MM and CM.

$$\eta = 3082.35(1/n)^2 - 103.08(1/n) + 1.793. \quad (21)$$

The polynomial estimate for frequencies evaluated at $1/n=0$ for the lowest normalized frequency of cube-shaped nanoparticles is 1.793, which is only 1.8% higher than the lowest continuum mechanics frequency. Besides, Eq. (30) indicates that continuum mechanics methods may be used to estimate (within 95%) the lowest vibrational frequencies of cube-shaped nanoparticles with at least 836 atoms, and $R > 1.5$ nm.

In spite of the excellent agreement between the cube geometries modeled using the molecular and continuum mechanics methods, the number and location of the degenerate frequency groups are not the same for these two approaches. However, for the largest cube-shaped nanoparticles considered, the mode shapes corresponding to the lowest two frequencies are the same for silicon structures, and some higher modes also have the same deformation as it is for the cases of the 5th (frequencies 12–14) and 6th (frequencies 15–17) modal groups. Figures 6 and 7 illustrate the modal shapes corresponding to these four degenerate frequency groups for silicon cube-shaped nanoparticles obtained from molecular and continuum mechanics methods.

TABLE II. Normalized frequencies for silicon cubes.

Atoms	64	216	512	1000	
L (Å)	9.5656	14.9988	20.4356	25.8838	Continuum
R (Å)	5.9340	9.3045	12.6773	16.0570	
1	0.9318	1.3749	1.6105	1.6875	1.7622
2	1.1012	1.4325	1.6105	1.6875	1.7622
3	1.1012	1.4325	1.6306	1.7547	1.9672
4	1.2333	1.4884	1.6306	1.7547	1.9672
5	1.2333	1.4884	1.6358	1.7948	1.9672
6	1.2636	1.5123	1.6823	1.8087	2.2072
7	1.4073	1.7407	1.9347	2.0819	2.2072
8	1.4073	1.7407	1.9347	2.0819	2.2072
9	1.5487	1.8465	2.0456	2.1785	2.3591
10	1.6177	1.8465	2.1239	2.2632	2.3591
11	1.6177	1.8478	2.1239	2.2632	2.3591
12	1.6702	1.9330	2.1670	2.2702	2.3780
13	1.6702	1.9608	2.1670	2.2736	2.3780
14	1.7329	1.9608	2.1725	2.2736	2.3780
15	1.7537	2.0112	2.2074	2.3922	2.5710
16	1.8512	2.0112	2.2191	2.4179	2.5710
17	1.8512	2.0318	2.2191	2.4179	2.5710
18	1.9179	2.0440	2.3418	2.4465	2.6691
19	1.9179	2.0440	2.3418	2.4465	2.6691
20	2.0030	2.0519	2.3823	2.5590	2.9709

C. Tetrahedron-shaped nanostructures

It is important to begin this section by noting that the geometry of the molecules used in this approach perfectly describes tetrahedra with four faces that are equilateral triangles. Unlike the sphere and cube, there is no atomic discretization error for this geometry. The lowest 20 normalized frequencies obtained using molecular mechanics are shown

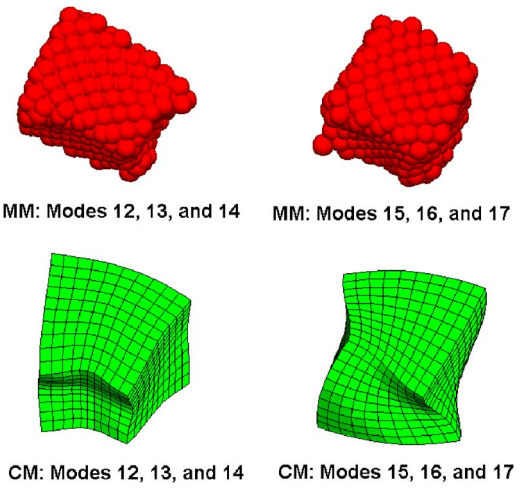


FIG. 7. (Color online) Identical vibrational modes of cube-shaped nanostructures obtained using MM and CM.

TABLE III. Normalized frequencies for silicon tetrahedrons.

Atoms	51	136	281	502	815	
a (Å)	11.5125	19.1960	26.8761	34.5539	42.2301	Continuum
R (Å)	3.5015	5.8384	8.1743	10.5094	12.8441	
1	0.8821	1.0443	1.0940	1.1303	1.1588	1.2100
2	0.8821	1.0443	1.0940	1.1303	1.1588	1.2100
3	0.8821	1.0647	1.1677	1.2338	1.2800	1.3357
4	0.9575	1.0647	1.1677	1.2338	1.2800	1.3357
5	0.9575	1.0647	1.1677	1.2338	1.2800	1.3357
6	1.1083	1.1887	1.2317	1.2644	1.2910	1.3602
7	1.1147	1.1887	1.2317	1.2644	1.2910	1.3602
8	1.1147	1.1887	1.2317	1.2644	1.2910	1.3602
9	1.1147	1.4082	1.5863	1.7023	1.7832	1.9421
10	1.3334	1.5761	1.7127	1.7989	1.8586	1.9731
11	1.3334	1.5761	1.7127	1.7989	1.8586	1.9731
12	1.4090	1.6841	1.8375	1.9088	1.9521	2.0373
13	1.4090	1.6841	1.8375	1.9088	1.9521	2.0373
14	1.4090	1.6841	1.8375	1.9088	1.9521	2.0373
15	1.4152	1.7139	1.8425	1.9382	2.0096	2.1505
16	1.4152	1.7139	1.8425	1.9382	2.0096	2.1505
17	1.4152	1.7139	1.8425	1.9382	2.0096	2.1505
18	1.7316	2.0229	2.1759	2.2802	2.3585	2.5153
19	1.7316	2.0229	2.2072	2.3408	2.4381	2.8278
20	1.7316	2.0229	2.2072	2.3408	2.4381	2.8278

in Table III along with the continuum values. The dependence on the nanostructure size of the normalized frequencies and degenerate modes is confirmed again, with their values increasing with the number of atoms forming the molecular model. Normalized frequencies of silicon obtained using molecular mechanics tend toward the continuum values from below, the frequencies for the largest tetrahedron (a=4.2230 nm) are between 4.2% and 13.7% lower than continuum values.

The lowest frequency for the smallest tetrahedron (a =1.15125 nm) is 27.1% lower than the continuum value, and 4.2% lower for the largest one (a=4.2230 nm). The behavior of the lowest frequency with the number of atoms is shown in Fig. 3. The exhibited smooth variation allows the fitting of a second-order polynomial as

$$\eta = 301.08(1/n)^2 - 20.81(1/n) + 1.175. \tag{22}$$

According to this polynomial, the converged value for silicon tetrahedrons is 1.175, which is only 2.9% lower than the continuum value. Good estimates of the lowest natural frequency within 95% are obtained using continuum mechanics for tetrahedra with more than 800 atoms and R >1.28 nm.

The number and groups of degenerate frequencies match exactly for both molecular and continuum mechanics methods. The mode shapes are also the same as illustrated in Figs. 8 and 9 for silicon tetrahedrons. Because of this exact match in mode and frequency, it is quite possible, though not certain, that discrepancies in modal type for spheres and cubes are caused in part by discretization error.

IV. SUMMARY AND CONCLUSIONS

The main objectives of this work were (1) to evaluate the applicability of the continuum mechanics assumption to determine the natural frequencies and mode shapes of Si particles at the nanoscale, and (2) to obtain expressions for the evaluation of their natural frequencies as a function of size and number of atoms forming the particles. To achieve this goal, the vibrational spectra of Si nanostructures with different shapes and sizes have been studied using two different methods: molecular mechanics and continuum mechanics.

Results obtained using the molecular mechanics approach confirmed the size dependence of particles properties at the nanoscale. Not only the natural frequencies, but also the location of the degenerate frequency groups change with

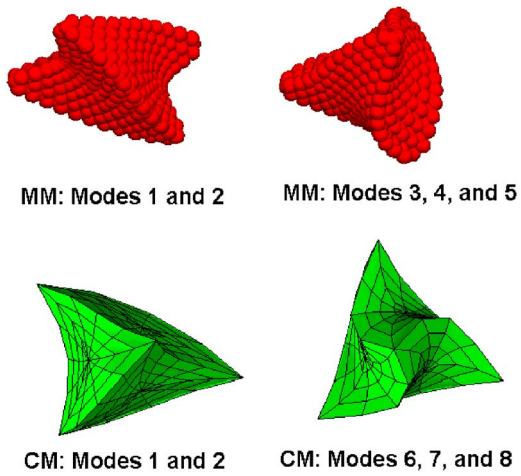


FIG. 8. (Color online) Identical vibrational modes of tetrahedron-shaped nanostructures obtained using MM and CM.

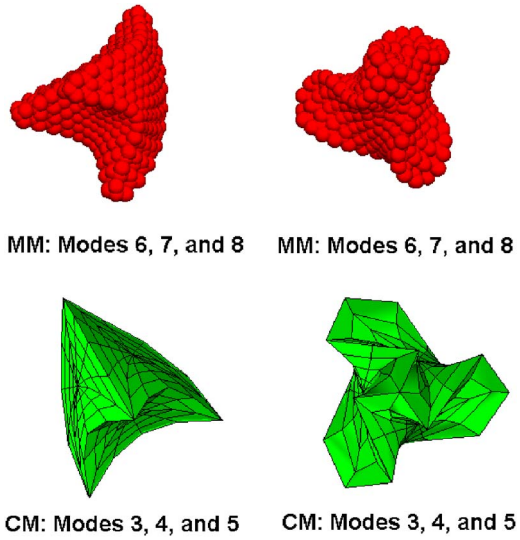


FIG. 9. (Color online) Identical vibrational modes of tetrahedron-shaped nanostructures obtained using MM and CM.

the number of atoms forming the solid. Normalized natural frequencies increase nonlinearly with the number of atoms forming the particles. This behavior for the lowest frequency exhibited a smooth variation for the cases of cube- and tetrahedron-shaped nanostructures allowing the fitting of polynomials to describe this behavior. However, the variation of the natural frequencies for sphere-shaped particles was not smooth, and it was not possible to fit an acceptable expression describing their variation with the size of the nanostructures. This is partially attributed to the fact that the geometry of the molecular structures was not perfectly spherical, and it changes with the number of atoms. The converged values for the lowest normalized frequencies according to the fitted second-order polynomials are 1.793 and 1.1752 for cube- and tetrahedron-shaped nanoparticles, respectively. Compared with the lowest continuum mechanics frequency, these values are 1.8% higher for cubes and 2.9% lower for tetrahedrons. Normalized frequencies of Si nanoparticles may be estimated using the proposed equations for the shapes and particle size range considered here.

The normalized natural frequencies increase linearly with the size of the nanoparticle. This behavior indicates that the scale invariance of the term ωR in continuum elasticity does not hold for the range of particle sizes considered in this study.

Continuum mechanics methods provide good estimates of the lowest natural frequency of particles having at least 836 ($R > 1.5$ nm) and 800 ($R > 1.28$ nm) atoms for cube- and tetrahedron-shaped nanostructures, respectively. For smaller particles a more accurate method such as quantum and molecular mechanics must be used for final analyses or rigorous conclusions. Equations for the evaluation of the lowest natural frequencies of cube- and tetrahedron-shaped nanoparticles as functions of the number of atoms forming the particle were proposed.

The normalized frequencies calculated using the molecular mechanics method approached the continuum values from below. The shape of the vibration modes obtained from molecular and continuum mechanics methods were practically the same for the sphere- and tetrahedron-shaped particles with the largest number of atoms. However, for cubes only the shape of the modes corresponding to the lowest two frequency groups were identical for both methods. In general, the modal shapes obtained using molecular mechanics coincide with those from continuum mechanics, however, they may not occur in the same order, as forces near the surface start to play a more important role.

ACKNOWLEDGMENT

This work was sponsored by the National Science Foundation under Grant No. CMS-030108. This support is gratefully acknowledged.

[1] M. Fujii, T. Nagareda, S. Hayashi, and K. Yamamoto, "Low-frequency Raman scattering from small silver particles embedded in SiO2 thin films," Phys. Rev. B **44**, 6243–6248 (1991).
[2] S. Rufo, M. Dutta, and M. A. Stroscio, "Acoustic modes in free and embedded quantum dots," J. Appl. Phys. **93**, 2900–2905 (2003).
[3] A. Tanaka, S. Onari, and T. Arai, "Low-frequency Raman scattering from CdS microcrystals embedded in a germanium dioxide glass matrix," Phys. Rev. B **47**, 1237–1243 (1993).
[4] N. N. Ovisuk and V. N. Novikov, "Influence of a glass matrix on acoustic phonons confined in microcrystals," Phys. Rev. B **53**, 3113–3118 (1996).
[5] P. Verma, W. Cordts, G. Irmer, and J. Monecke, "Acoustic vibrations of semiconductor nanocrystals in doped glasses," Phys. Rev. B **60**, 5778–5785 (1999).
[6] L. Saviot, B. Champagnon, E. Duval, I. A. Kudriavtsev, and A. I. Ekimov, "Size dependence of acoustic and optical vibrational modes of CdSe nanocrystals in glasses," J. Non-Cryst. Solids **197**, 238–246 (1996).
[7] L. Saviot, D. B. Murray, and M. C. Marco de Lucas, "Vibrations of free and embedded anisotropic elastic spheres: Application to low-frequency Raman scattering of silicon nanoparticles in silica," Phys. Rev. B **69**, 113402 (2004).
[8] H. Lamb, "On the vibrations of an elastic sphere," Proc. London Math. Soc. **13**, 189–212 (1882).
[9] W. Cheng and S. F. Ren, "Calculations on the size effects of Raman intensities of silicon quantum dots," Phys. Rev. B **65**, 205305 (2002).
[10] W. Cheng, S. F. Ren, and P. Y. Yu, "Theoretical investigation of the surface vibrational modes in germanium nanocrystals," Phys. Rev. B **68**, 193309 (2003).
[11] W. Cheng, S. F. Ren, and P. Y. Yu, "Microscopic theory of the low frequency Raman modes in germanium nanocrystals," Phys. Rev. B **71**, 174305 (2005).
[12] A. Stekel, J. L. Sarrao, T. M. Bell, R. G. Leisure, W. M. Visscher, and A. Migliori, "Method for identification of the vibrational modes of a rectangular parallelepiped," J. Acoust. Soc. Am. **92**, 663–668 (1992).
[13] P. Heyliger, A. Jilani, H. Ledbetter, R. G. Leisure, and C. Wang, "Elastic constants of isotropic cylinders using resonant ultrasound," J. Acoust. Soc. Am. **94**, 1482–1487 (1993).
[14] K. Foster, S. L. Fairburn, R. G. Leisure, S. Kim, D. Balzar, G. Alers, and H. Ledbetter, "Acoustic study of texture in polycrystalline brass," J. Acoust. Soc. Am. **105**, 2663–2668 (1999).
[15] A. K. Rappe' and C. J. Casewit, *Molecular Mechanics Across Chemistry* (University Science Books, Sausalito, CA, 1997).
[16] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," J. Am. Chem. Soc. **114**, 10024–10035 (1992).
[17] C. J. Casewit, K. S. Colwell, and A. K. Rappe, "Application of a universal force field to organic molecules," J. Am. Chem. Soc. **114**, 10035–10046 (1992).
[18] C. J. Casewit, K. S. Colwell, and A. K. Rappe, "Application of a universal force field to main group compounds," J. Am. Chem. Soc. **114**, 10046–10053 (1992).
[19] A. K. Rappe, K. S. Colwell, and C. J. Casewit, "Application of a universal force field to metal complexes," Inorg. Chem. **32**, 3438–3450 (1993).
[20] F. Ramirez, P. R. Heyliger, A. K. Rappe, and R. G. Leisure, "Breakdown of frequency–spectra scaling of Si nanoparticles," Phys. Rev. B **76**,

085415 (2007).
[21]J. N. Murrell and K. S. Sorbie, "New analytic form for the potential energy curves of stable diatomic states," J. Chem. Soc., Faraday Trans. 2 **70**, 1552–1556 (1974).
[22]P. Huxley and J. N. Murrell, "Ground-state diatomic potentials," J. Chem. Soc., Faraday Trans. 2 **79**, 323–328 (1983).
[23]K. V. Ermakov, B. S. Butayev, and V. P. Spiridonov, "Model potential functions of nonlinear XY_2 molecules," J. Mol. Struct. **240**, 295–303 (1990).
[24]F. H. Stillinger and T. A. Weber, "Computer simulation of local order in condensed phases of silicon," Phys. Rev. B **31**, 5262–5271 (1985).
[25]R. T. Pack, J. J. Valentini, C. H. Becker, R. J. Buss, and Y. T. Lee, "Multiproperty empirical interatomic potentials for ArXe and KrXe," J. Chem. Phys. **77**, 5475–5485 (1982).
[26]E. B. Wilson, J. C. Decious, and P. C. Cross, *Molecular Vibrations* (Dover, New York, 1980).
[27]J. N. Reddy, *Energy and Variational Methods in Applied Mechanics* (Wiley, New York, 1984).
[28]W. M. Visscher, A. Migliori, T. M. Bell, and R. A. Reinerr, "On the normal modes of free vibration of inhomogeneous and anisotropic elastic objects," J. Acoust. Soc. Am. **90**, 2154–2162 (1991).
[29]P. R. Heyliger and J. Kienholz, "The mechanics of pyramids," Int. J. Solids Struct. **43**, 2693–2709 (2006).
[30]http://wwwprof.uniandes.edu.co/~framirez/QD/FRR_nanovib.html. Author's personal webpage, last viewed 10 September 2007.

Vibration modeling of structural fuzzy with continuous boundary

Lars Friis[a)]
Acoustic Technology, Ørsted DTU, Technical University of Denmark, Building 352, DK-2800 Kgs. Lyngby, and Widex A/S, Ny Vestergaardsvej 25, DK-3500 Vaerloese, Denmark

Mogens Ohlrich[b)]
Acoustic Technology, Ørsted DTU, Technical University of Denmark, Building 352, DK-2800 Kgs. Lyngby, Denmark

(Received 29 June 2007; revised 5 November 2007; accepted 20 November 2007)

From experiments it is well known that the vibration response of a main structure with many attached substructures often shows more damping than structural losses in the components can account for. In practice, these substructures, which are not attached in an entirely rigid manner, behave like a multitude of different sprung masses each strongly resisting any motion of the main structure (master) at their base antiresonance. The "theory of structural fuzzy" is intended for modeling such high damping. In the present article the theory of fuzzy structures is briefly outlined and a method of modeling fuzzy substructures examined. This is done by new derivations and physical interpretations are provided. Further, the method is extended and simplified by introducing a simple deterministic approach to determine the boundary impedance of the structural fuzzy. By using this new approach, the damping effect of the fuzzy with spatial memory is demonstrated by numerical simulations of a main beam structure with fuzzy attachments. It is shown that the introduction of spatial memory reduces the damping effect of the fuzzy and in certain cases the damping effect may even be eliminated completely. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2823498]

PACS number(s): 43.40.At, 43.40.Tm [DF] Pages: 718–728

I. INTRODUCTION

It is commonly known from experiments that vibrations of a complicated system consisting of a main structure and a large number of small attached resonant substructures often appears to be more damped than the main structure's damping properties would imply. Already in 1928 Ormondroyd and Den Hartog[1] revealed that a "dynamic absorber" could produce a considerable reduction in vibration level, however, only in a relatively narrow frequency band around its natural frequency. To be effective such an auxiliary system is usually attached at the forcing point of the main structure. And the auxiliary system is predominantly a reflecting device controlled by its stiffness and mass, although its bandwidth of influence can be slightly increased by a tuning, which involves small adjustments of both stiffness and damping of the device. Therefore, the term "dynamic neutralizer" has been adopted.[2] For a more recent and detailed treatment of this subject the reader is referred to the monograph by Mead.[3] It took almost 60 years before Soize[4] and Chabas *et al.*[5] suggested that attached resonant substructures, in effect, behave like a multitude of dynamic neutralizers with different natural frequencies that introduce a high damping in the main structure over a broader frequency range.

Many complicated engineering systems consist basically of an outer shell- or a box-like *master* structure and a complicated *internal* structure. Examples of such structures varying from small to large sizes are electromechanical hearing aids, machines, aircraft, and ship hulls. The outer master structure is often well defined and its vibration can be predicted using conventional methods of vibrational analysis. In contrast, the dynamic properties of the internals may only be partly known and therefore their dynamics and influence have to be modeled by using an alternative method such as that offered by "fuzzy structure theory."[4–9] This theory is intended for an overall and simple prediction of the vibration of the master structure, and the theory considers the internal parts as a single or several independent "fuzzy substructures," which are known in some statistical sense only.

In some systems the fuzzy substructures are attached to the master through a continuous boundary or junction. This could, for example, be line-coupled machinery in a ship hull or passenger seats and luggage compartments attached to the main structure of an airplane. The continuous connection boundary implies that spatial coupling within the fuzzy has to be considered, and it is only in special cases that this coupling can be neglected.

Often the motion of the continuous junctions is varying significantly with position due to the spatial variation of vibration in the master structure, and spatial coupling forces in the fuzzy have to be accounted for. The present article addresses this problem of including spatial coupling in the modeling of structural fuzzy.

With frequency or vibration wavelength as a parameter, Fig. 1 shows three scenarios, each with three different cases

[a)]Electronic mail: lf@oersted.dtu.dk
[b)]Electronic mail: mo@oersted.dtu.dk

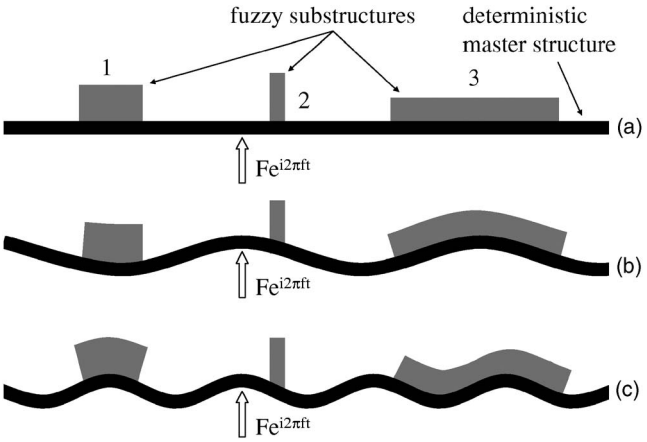


FIG. 1. Three different fuzzy substructures attached to a master structure undergoing harmonic vibration: (a) low frequency rigid body motion, (b) low frequency wave motion, and (c) mid-frequency wave motion.

of fuzzy substructures attached to a master. It is assumed that a time harmonic force of amplitude F and angular frequency $\omega=2\pi f$ excites the master and generates vibration in the whole system. At very low frequencies the master structure vibrates as a rigid body in translational motion [see Fig. 1(a)], and the junction displacement at the boundary of the fuzzy substructures is almost constant. This implies that the spatial coupling within each substructure has no significant effect on the response of the system as a whole. Now, increasing the excitation frequency introduces elastic motion in the master structure and hence at the interface with the fuzzy. When the vibration wavelength of the master becomes comparable with the dimensions of the fuzzy connection area then the spatial coupling begins to take effect. This is the case in Fig. 1(b) where the boundary displacement of substructure 3 is varying whereas those of substructures 1 and 2 are nearly constant. In Fig. 1(c) the frequency has been increased further and the boundary displacements of both substructures 1 and 3 are varying, whereas the boundary displacement of substructure 2 remains close to constant.

Fuzzy structure theory was originally developed by Soize and co-worker and presented in a series of papers[4–7] during a 10-year period starting in 1986. These papers involve probabilistic concepts in order to take the model uncertainties into account. In attempts to explain the main ideas behind the theory, these papers have been subject to numerous simplifications, interpretations, and extensions during the last 20 years.

In particular, Pierce *et al.*[8] and Strasberg and Feit[9] have introduced more simple and deterministic methods in order to predict the average responses of the master. In these methods the main parameter describing the fuzzy is taken to be the distribution of resonating mass per unit frequency. One of the crucial steps in applying fuzzy structure theory is the very estimation of this mass distribution. Both Soize[10,11] and Pierce[12] have addressed this problem throughout the last 10 years.

In 1993 Soize briefly presented a method[6] for including spatial memory in the modeling of structural fuzzy with continuous boundaries. Despite this, elaborating literature has so far mainly been concerned with structural fuzzy *without* spatial coupling effects. However, in order to utilize Soize's innovative theory on structural fuzzy with spatial memory, there is a strong need for a detailed examination of the suggested method in addition to a presentation of necessary supplementary derivations. Moreover, a simple method for implementing the structural fuzzy is still absent in the open literature. The objective of the present article is to extend Soize's theory by using a simplifying approach, which to some extent is based on the methods introduced by Pierce. *et al.*,[8] and Strasberg and Feit.[9] The outline of the present article is as follows. Succeeding a brief outline of the theory of fuzzy structures in Sec. II, the method of including spatial memory is discussed in details and extended in Sec. III; this includes (i) derivation of the boundary impedance of Soize's spatial oscillator, (ii) derivation of the boundary impedance of an infinite number of *identical* spatial oscillators evenly distributed on the fuzzy connection area, and (iii) introduction to Soize's local *equivalent oscillator* and *equivalent coupling factor* and a presentation of new physical interpretations. Further, in Sec. IV we present a new approach for determining the boundary impedance of structural fuzzy with spatial memory. Finally, numerical simulations based on this approach are presented in Sec. V in order to illustrate the damping effects of structural fuzzy, which includes spatial memory.

II. STRUCTURAL FUZZY WITHOUT SPATIAL MEMORY

The purpose of the fuzzy structure theory is to model the overall vibrational response of a master structure, which has an attachment of one or more *resonant* substructures. A fuzzy substructure is considered as being composed of many simple oscillators resonating at different frequencies and being attached to the master at their base. When modeling such a system it is an advantage to separate the fuzzy from its master. Each fuzzy substructure is conveniently represented in terms of its *boundary impedance*.[4,5] Using this approach the modeling of the response of the master with fuzzy attachments can be achieved without exceeding the number of degrees of freedom required for predicting the response of the master structure itself. As mentioned previously, one can neglect the spatial coupling effects in the fuzzy when a fuzzy substructure of multiple resonators is connected to the master over a small length or over a small area of virtually constant motion.

This is illustrated in Fig. 2 showing a fuzzy substructure, which is modeled by N simple oscillators that is attached locally at an area A of the master structure. An expression for the total *boundary* impedance $\underline{z}_{\text{fuzzy}}(f)$ of this substructure can be derived by superposition and by assuming, say, that the nth simple oscillator of the fuzzy is defined by the mass M_n, the undamped resonance frequency $f_{r,n}$ and the loss factor η. By introducing the complex stiffness of the oscillator $\underline{s}_n=s_n(1+i\eta)$, where $s_n=(2\pi f_{r,n})^2 M_n$ is the impedance of the oscillator $\underline{Z}_n=\underline{F}_n/\underline{v}_n$ at the attachment base yields[9]

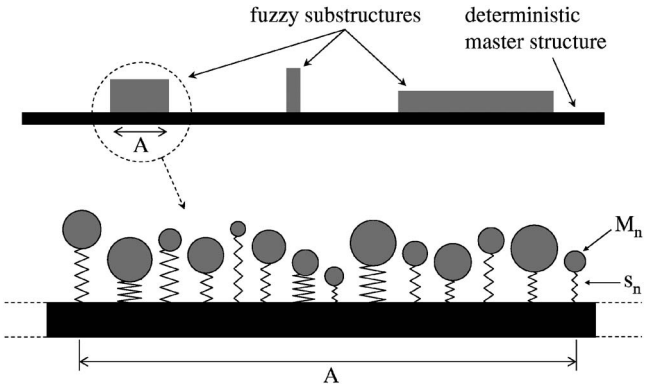


FIG. 2. Master structure exemplifying an attached fuzzy substructure, which is composed of N simple oscillators resonating at different frequencies and without spatial memory.

$$\underline{Z}_n = \frac{\underline{s}_n}{i\omega}\left(1 - \frac{\underline{s}_n}{\underline{s}_n - \omega^2 M_n}\right) = -i2\pi f\left(\frac{f_{r,n}^2}{f^2}\right)(1+i\eta) \times M_n\left(1 - \frac{f_{r,n}^2(1+i\eta)}{f_{r,n}^2(1+i\eta) - f^2}\right). \quad (1)$$

Figure 3 shows the frequency variation of this oscillator impedance in a normalized form for different values of spring damping η. Below and above its resonance frequency the oscillator is, respectively, mass controlled and spring controlled. Further, at resonance of the oscillator, where the impedance is very large and almost purely real, it will strongly oppose any movements of its base. It is this particular feature of the oscillator, which results in the damping effect of the fuzzy substructure.

Generally, the different oscillators of a fuzzy substructure have different masses and natural frequencies and they are attached randomly to the master structure within the considered fuzzy connection area. Also, the total mass of all the oscillators equals the mass of the fuzzy substructure M_{fuzzy} $= \Sigma_{n=1}^{N} M_n$. Below a certain frequency, say $f_{r,\text{lower}}$, the oscillators will all be mass controlled. By increasing the frequency gradually from $f_{r,\text{lower}}$ to an upper limit, say $f_{r,\text{upper}}$, the oscillators will resonate one by one. Now, at each frequency within this "resonant" frequency band $f_{r,\text{lower}} \leqslant f_r \leqslant f_{r,\text{upper}}$ at least one oscillator will be close to its base antiresonance and it will therefore oppose the motion of the master. If the oscillators are attached close to one another within the area A, which has a nearly constant displacement, then the effective boundary impedance of all the oscillators, $\underline{z}_{\text{fuzzy}}(f)$, can be approximated by the sum of each oscillator's impedance $\underline{Z}_n(f)$ divided by the attachment area A:

$$\underline{z}_{\text{fuzzy}}(f) = \frac{1}{A}\sum_{n=1}^{N}\underline{Z}_n(f) = -\frac{i2\pi f}{A}\sum_{n=1}^{N}\left(\frac{f_{r,n}^2}{f^2}\right)(1+i\eta) \times M_n\left(1 - \frac{f_{r,n}^2(1+i\eta)}{f_{r,n}^2(1+i\eta) - f^2}\right). \quad (2)$$

This boundary impedance, however, requires specific knowledge about the properties of each oscillator and it is therefore conveniently replaced by an asymptotic and smoothed version.[8,9] This is obtained by considering infinitely many oscillators resonating within the frequency band of $f_{r,\text{lower}} \leqslant f_r \leqslant f_{r,\text{upper}}$ and having a total mass M_{fuzzy}. This smoothed impedance yields[8,9]

$$\underline{z}_{\text{fuzzy}}(f) = -\frac{i2\pi f}{A}\int_{f_{r,\text{lower}}}^{f_{r,\text{upper}}}\left(\frac{f_r^2}{f^2}\right)(1+i\eta) \times m_{\text{fuzzy}}(f_r)\left(1 - \frac{f_r^2(1+i\eta)}{f_r^2(1+i\eta) - f^2}\right)df_r, \quad (3)$$

where the quantity $m_{\text{fuzzy}}(f_r)df_r$ represents the mass resonating between the frequencies f_r and $f_r + df_r$; this means that the total mass of the fuzzy substructure now is expressed as

$$M_{\text{fuzzy}} = \int_{f_{r,\text{lower}}}^{f_{r,\text{upper}}} m_{\text{fuzzy}}(f_r)df_r. \quad (4)$$

The damping effect of the fuzzy substructure is mainly governed by this frequency dependent resonating mass distribution $m_{\text{fuzzy}}(f_r)$.[8,9] Methods for finding this parameter were suggested by Soize[10,11] and Pierce,[12] and different prototype mass distributions were proposed by Pierce *et al.*[8] and Strasberg and Feit.[9]

As an example of the damping effect of structural fuzzy Fig. 4 shows computed results for the velocity vibration response per unit harmonic force, $\underline{Y} = \underline{v}/\underline{F}$, of a flexurally vibrating master beam, free in space, both without and with an attached substructure represented by 16 different simple oscillators. The resonance frequencies of these oscillators are spaced in geometric progression from 500 to 5000 Hz. Further, the oscillators have identical point masses, weighting in total 10% of the master beam and a spring loss factor of $\eta = 0.05$. It is clearly observed that the attached substructure has a strong effect on the master response; this is seen to be reduced considerable over a broad band of frequencies and by up to 18 dB around 1300 Hz. Further, it is seen that this substructure can be modeled successfully as a smoothed structural fuzzy by using the expression in Eq. (3). An ap-

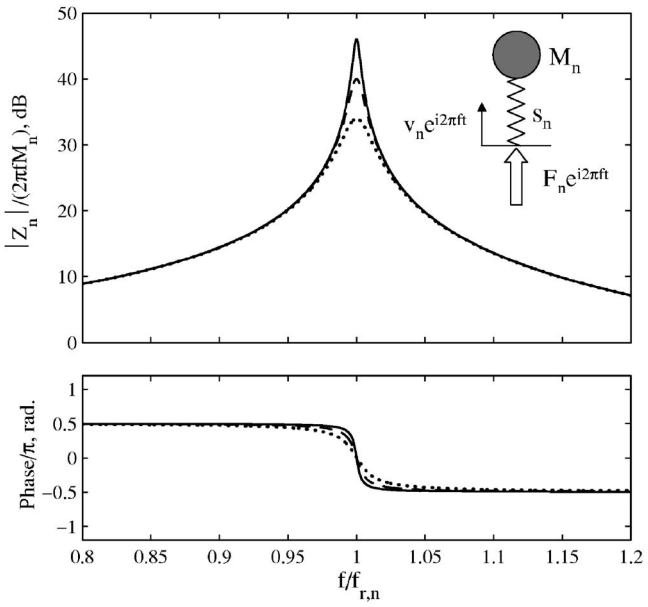


FIG. 3. Frequency variation of normalized impedance, $\underline{Z}_n/(2\pi f M_n)$ for different values of spring damping η: —, 0.005; ---, 0.01; ···, 0.02.

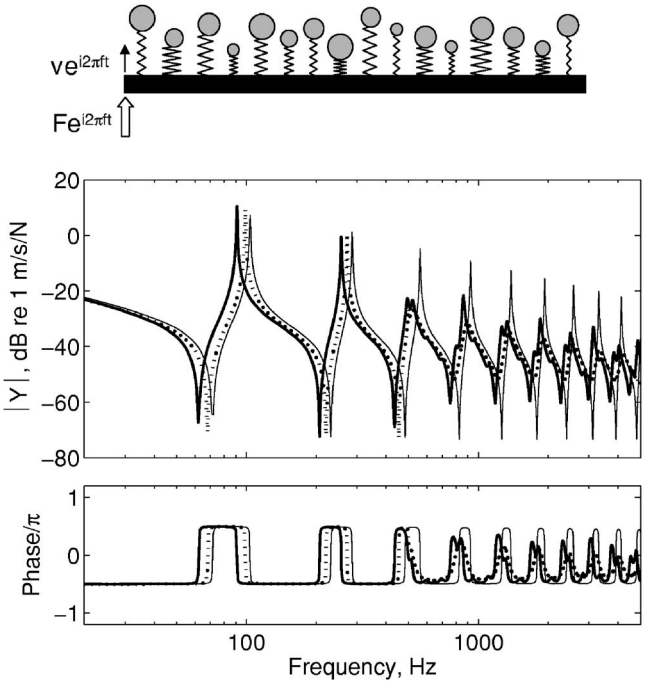


FIG. 4. Vibration velocity response per unit harmonic force, $\underline{Y}=\underline{v}/\underline{F}$, of a master beam free in space. Beam: —, without structural fuzzy; ⋯, with an attached fuzzy substructure represented by 16 simple oscillators, —; with a smoothed layer of structural fuzzy without spatial memory.

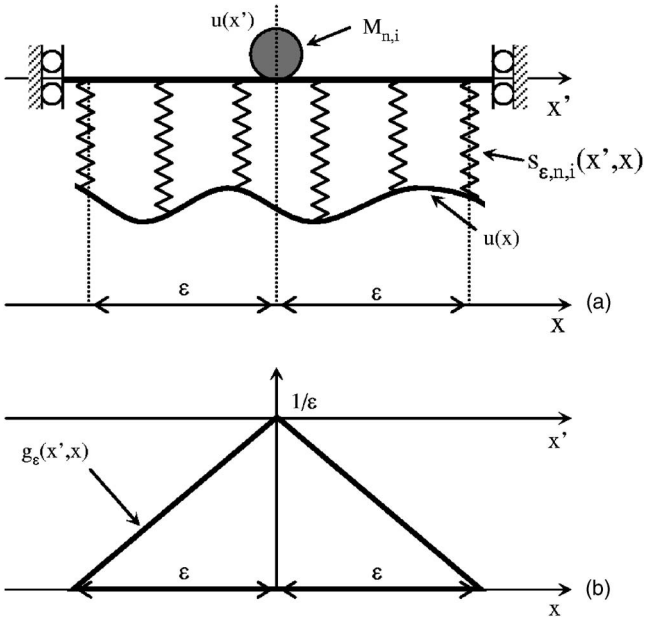


FIG. 5. Fuzzy oscillator with spatial coupling. (a) Oscillator attached to a boundary of motion $\underline{u}(x)$ and (b) stiffness density distribution $g_\varepsilon(x',x)$ of the oscillator spring elements.

proximate condition for using this expression is suggested in Ref. 9 as $\eta > 2\Delta f_{r,n}/f_{r,n}$ where $\Delta f_{r,n}$ is the spacing between adjacent resonance frequencies. In other words this very strict condition requires that $\Delta f_{r,n} < \Delta f_{3\ \mathrm{dB}}/2$ where $\Delta f_{3\ \mathrm{dB}}$ denotes the 3 dB bandwidth of the oscillator at resonance. However, in the present example it applies that the spacing between resonances is $\Delta f_{r,n} \approx 65 \cdot \Delta f_{r,n,3\ \mathrm{dB}}/2$ around 2000 Hz, that is, 65 wider than the suggested requirement. It is therefore evident that acceptable results can be obtained with a much relaxed condition.

III. SOIZE'S STRUCTURAL FUZZY WITH SPATIAL MEMORY

A. Soize's spatial oscillator

Consider a fuzzy substructure connected to the master through a continuous boundary. A fuzzy substructure is generally attached to the master within an area, but for the sake of simplicity we shall here consider a fuzzy attached to the master through a one-dimensional boundary of length L_{fuzzy}. Soize incorporates a spatial memory in the structural fuzzy by introducing a "spatial oscillator" as sketched in Fig. 5(a). A structural fuzzy with spatial memory is composed of N *different sets* ($n \in [1,N]$) of such spatial oscillators. Each of these N sets consists of infinitely many *identical* spatial oscillators spread on the fuzzy connection area—or length. Let us first consider only one spatial oscillator, say the ith ($i \in [1,\infty]$) of the nth set of spatial oscillators of a fuzzy substructure, see Fig. 5(a). This oscillator is defined by the resonance frequency $f_{r,n}$, the lossfactor η and the point mass $M_{n,i}$ located at position x'. Further, the point mass is assumed supported by spring elements of stiffness density $\underline{s}_{\varepsilon,n,i}(x',x)$ that are attached to the master structure at different positions $x \in [x'-\varepsilon, x'+\varepsilon]$. Moreover, the spatial variation of the vibration displacement of the master is shown as $u(x)$ in Fig. 5(a). The actual width 2ε of the distributed springs Soize denotes "the spatial memory" and the stiffness density he defines as[6]

$$\underline{s}_{\varepsilon,n,i}(x',x) = \underline{s}_{n,i} g_\varepsilon(x',x) = (M_{n,i}\omega_{r,n}^2)(1+i\eta) g_\varepsilon(x',x). \quad (5)$$

Here, $\underline{s}_{n,i}$ is the total complex stiffness of the ith oscillator belonging to the nth set, and the quantity $g_\varepsilon(x',x)$ is an even and positive-valued function of area 1. As a one-dimensional spatial memory Soize suggests a simple triangular distribution function $g_\varepsilon(x',x)$ as shown in Fig. 5(b). This is determined as

$$g_\varepsilon(x',x) = \frac{\varepsilon - |x'-x|}{\varepsilon^2} 1_{[x'-\varepsilon,x'+\varepsilon]} \quad (6)$$

where $1_{[x'-\varepsilon,x'+\varepsilon]}$ is a function, which is equal to 1 for $x \in [x'-\varepsilon, x'+\varepsilon]$ and which is 0 elsewhere. As the area under the curve $g_\varepsilon(x',x)$ is 1, the oscillator in Fig. 5(a) has the same natural frequency $f_{r,n}$ as the simple oscillator with mass M_n and stiffness s_n that was considered in Sec. II. From the expression in Eq. (6) it is seen that the distribution $g_\varepsilon(x',x)$ only is dependent on the difference $x'-x$ and therefore it can be written as

$$g_\varepsilon(x'-x) = \frac{\varepsilon - |x'-x|}{\varepsilon^2} 1_{[x'-\varepsilon,x'+\varepsilon]} \quad (7)$$

where it applies that $g_\varepsilon(x'-x) = g_\varepsilon(x-x')$ and further that $\underline{s}_{\varepsilon,n,i}(x'-x) = \underline{s}_{\varepsilon,n,i}(x-x')$.

B. The *n*th set of spatial oscillators

Next, consider the infinitely many identical oscillators belonging to the nth set. Let us assume that the oscillators

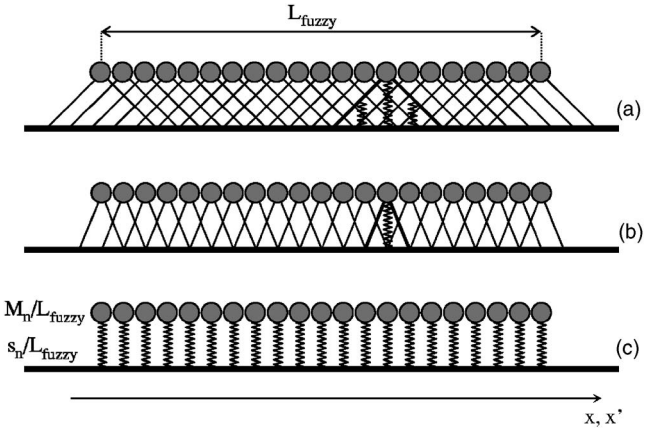


FIG. 6. Structural fuzzy attached to the master; fuzzy with: (a) high spatial memory, (b) small spatial memory, and (c) no spatial memory.

are distributed evenly over the fuzzy boundary, so that each location x' is associated with a point mass $M_{n,i}$. This is illustrated in Fig. 6 where each one of the oscillators representing the one sketched in Fig. 5(a) is depicted as a point mass and the associated triangular stiffness distribution. The point masses can vibrate independently whereas the spring elements overlap spatially at the connection boundary.

These infinitely many *identical* oscillators constitute the nth contribution to the total boundary impedance of the homogeneous fuzzy substructure. The total mass and total stiffness of all the oscillators of the nth set are given as M_n and $\underline{s}_n$, respectively, so that $M_{n,i}=M_n/L_{\text{fuzzy}}$ and $\underline{s}_{n,i}=\underline{s}_n/L_{\text{fuzzy}}$. Figure 6(a) illustrates the case of spatial oscillators with a large width—or a high spatial memory—beacuse the spring elements from the individual oscillators overlap significantly. On the other hand, in Fig. 6(b) the spring elements overlap less because ε is somewhat smaller. Finally, in Fig. 6(c) the spatial memory approaches zero as $\varepsilon \to 0$, and the spatial stiffness density approaches the stiffness of simple discrete springs.

C. Derivation of the boundary impedance of the *n*th set of oscillators

The vibration of the master results in a force at the interface between master and attached fuzzy; this action on the fuzzy we denote the contact force. Now, from the vibration velocity $\underline{v}(x)$ along the fuzzy connection boundary one can express the total contact force $\underline{F}'_{\varepsilon,n}(x_0)$ per unit length at x_0 due to the nth set of oscillators as[6]

$$\underline{F}'_{\varepsilon,n}(x_0) = \int_{L_{\text{fuzzy}}} \underline{z}_{\varepsilon,n}(x_0 - x)\underline{v}(x)dx, \tag{8}$$

where $\underline{z}_{\varepsilon,n}(x_0-x)$ is the boundary impedance associated with the nth set of oscillators. This impedance depends only on the difference (x_0-x), so that $\underline{z}_{\varepsilon,n}(x_0-x)=\underline{z}_{\varepsilon,n}(x-x_0)$ in analogy to the stiffness distribution function in Eq. (7). Although an expression for $\underline{z}_{\varepsilon,n}(x_0-x)$ is shown in Ref. 3, it has not been derived in the open literature. The authors believe that such a derivation is essential in order to appreciate and understand the characteristics of a fuzzy with spatial memory. It is also anticipated that such a derivation will ease the usability of Soize's theory considerably. In view of this, a step by step derivation of $\underline{F}'_{\varepsilon,n}(x_0)$ will be presented in what follows.

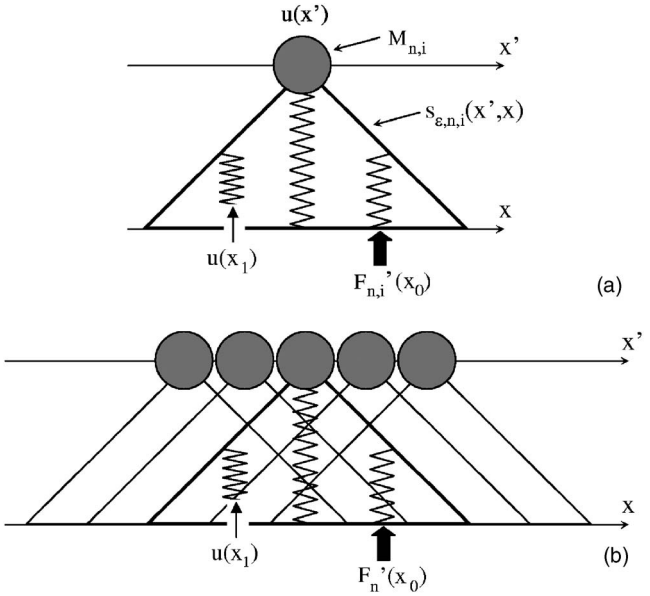


FIG. 7. Derivation of the nth contribution to the fuzzy boundary impedance. (a) Boundary impedance $\underline{z}_{\varepsilon,n,i}(x_0-x_1)$ of the ith oscillator of the nth set of spatial oscillators. (b) Boundary impedance $\underline{z}_{\varepsilon,n}(x_0-x_1)$ of the nth set of spatial oscillators.

Once again, only one spatial oscillator is considered, say, the ith of the nth set of oscillators. This oscillator is sketched in Fig. 7(a). For this particular oscillator we first seek an expression for the boundary impedance $\underline{z}_{\varepsilon,n,i}(x_0-x_1)$ with a spatial contribution over the differential length dx is defined as

$$\underline{F}'_{n,i}(x_0) = \underline{z}_{\varepsilon,n,i}(x_0 - x_1)\underline{v}(x_1)dx|_{\underline{v}(x\notin[x_1,x_1+dx])=0}, \tag{9}$$

where $\underline{F}'_{n,i}(x_0)$ is the force per unit length that excites the connection boundary at x_0 and $\underline{v}(x_1)=i\omega\underline{u}(x_1)$ is the base velocity of the spring element at x_1. Now, the spring element at x_1 is given a displacement $\underline{u}(x_1)$ at its base, whereas all other springs elements are locked such that $\underline{u}(x\neq x_1)=0$, see Fig. 7(a). As the mass $M_{n,i}$ undergoes a displacement $\underline{u}(x')$ the induced spring force $\underline{F}_{S,n,i}(x_1)$ in a differential neighborhood dx around x_1 becomes

$$\underline{F}_{S,n,i}(x_1) = \underline{s}_{\varepsilon,n,i}(x' - x_1)(\underline{u}(x_1) - \underline{u}(x'))dx. \tag{10}$$

Due to the motion of the mass, a reaction force $\underline{F}_{M,n,i}(x')$ influences the spring element and this force is given as

$$\underline{F}_{M,n,i}(x') = -\omega^2 M_{n,i}\underline{u}(x'). \tag{11}$$

Additionally, the motion of the mass also introduces forces in the remaining spring elements, and their total spring force $\underline{F}_{S,n,i}(x\neq x_1)$ can be found as

$$\underline{F}_{S,n,i}(x \neq x_1) = -\underline{s}_{\varepsilon,n,i}(x' - x_1)\underline{u}(x')dx + \int_{L_{\text{fuzzy}}} \underline{s}_{\varepsilon,n,i}(x' - x)\underline{u}(x')dx = -\underline{s}_{\varepsilon,n,i}(x' - x_1)\underline{u}(x')dx + \underline{s}_{n,i}, \tag{12}$$

where $\underline{s}_{n,i}$ is the total stiffness of the spring elements of the oscillator. Because of force equilibrium the sum of the spring

force $\underline{F}_{S,n,i}(x \neq x_1)$ and the reaction force of the mass $\underline{F}_{M,n,i}(x')$ is equal to the spring force $\underline{F}_{S,n,i}(x_1)$ at x_1. Thus, by combination of Eqs. (10)–(12) we get

$$\underline{F}_{S,n,i}(x_1) = \underline{F}_{M,n,i}(x') + \underline{F}_{S,n,i}(x \neq x_1),$$

$$\underline{s}_{\varepsilon,n,i}(x' - x_1)(\underline{u}(x_1) - \underline{u}(x'))dx = (-\underline{s}_{\varepsilon,n,i}(x' - x_1)dx + \underline{s}_{n,i} - \omega^2 M_{n,i})\underline{u}(x'), \tag{13}$$

Rearranging Eq. (13) we find an expression for $\underline{u}(x')$ as a function of $\underline{u}(x_1)$, which yields

$$\underline{u}(x') = \frac{\underline{s}_{\varepsilon,n,i}(x' - x_1)\underline{u}(x_1)dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}. \tag{14}$$

Further, the force per unit length $\underline{F}'_{n,i}(x_0)$ at x_0 at the connection boundary is given as

$$\underline{F}'_{n,i}(x_0) = \underline{s}_{\varepsilon,n,i}(x' - x_0)(\underline{u}(x_0)\delta_{x_0,x_1} - \underline{u}(x')), \tag{15}$$

where δ_{x_0,x_1} is the Kronecker delta, that is, $\delta_{x_0,x_1}=1$ when $x_0=x_1$, and otherwise zero. Finally, if Eq. (14) is substituted in Eq. (15) we get the force per unit length $\underline{F}'_{n,i}(x_0)$ exerted onto the fuzzy connection boundary due to its displacement $\underline{u}(x_1)$:

$$\underline{F}'_{n,i}(x_0) = \underline{s}_{\varepsilon,n,i}(x' - x_0)\left(\delta_{x_0,x_1}\underline{u}(x_0) - \frac{\underline{s}_{\varepsilon,n,i}(x' - x_1)\underline{u}(x_1)dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}\right) = \underline{s}_{\varepsilon,n,i}(x' - x_0)\left(\delta_{x_0,x_1} - \frac{\underline{s}_{\varepsilon,n,i}(x' - x_1)dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}\right)\underline{u}(x_1). \tag{16}$$

According to Eq. (16) $\underline{z}_{\varepsilon,n,i}(x_0 - x_1)dx$ reads

$$\underline{z}_{\varepsilon,n,i}(x_0 - x_1)dx = \frac{\underline{s}_{\varepsilon,n,i}(x' - x_0)}{i2\pi f}\left(\delta_{x_0,x_1} - \frac{\underline{s}_{\varepsilon,n,i}(x' - x_1)dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}\right). \tag{17}$$

Now, this is only the impedance of the ith oscillator of the nth set. Assume again that there is an infinite number of *identical* oscillators overlapping one another on the fuzzy connection boundary [see Fig. 7(b)], such that a mass element $M_{n,i}$ is located at each of all positions along the x' axis. Further, each of these oscillators has a total stiffness $\underline{s}_{n,i}$. We now seek an expression for the impedance $\underline{z}_{\varepsilon,n}(x_0 - x_1)$ of the nth set of oscillators. First the oscillators are given the displacement $\underline{u}(x_1)$ at position x_1, whereas all other positions at their spring bases are locked. Then the total force per unit length $\underline{F}'_n(x_0)$ at x_0 is found by integrating the expression in Eq. (16) with respect to x', and this gives

$$\begin{aligned}
\underline{F}'_n(x_0) &= \int_{L_{\text{fuzzy}}} \underline{s}_{\varepsilon,n,i}(x' - x_0)\left(\delta_{x_0,x_1} - \frac{\underline{s}_{\varepsilon,n,i}(x' - x_1)dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}\right)\underline{u}(x_1)dx' \\
&= \left(\int_{L_{\text{fuzzy}}} \underline{s}_{\varepsilon,n,i}(x' - x_0)\delta_{x_0,x_1}dx' - \int_{L_{\text{fuzzy}}} \frac{\underline{s}_{\varepsilon,n,i}(x' - x_0)\underline{s}_{\varepsilon,n,i}(x' - x_1)dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}\right)\underline{u}(x_1) \\
&= \left(\underline{s}_{n,i}\delta_{x_0,x_1} - \int_{L_{\text{fuzzy}}} \frac{\underline{s}_{n,i}^2 g(x' - x)g(x' - x_0)dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}dx'\right)\underline{u}(x_1) \\
&= \left(\underline{s}_{n,i}\delta_{x_0,x_1} - \frac{\underline{s}_{n,i}^2 dx}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}\int_{L_{\text{fuzzy}}} g(x' - x_1)g(x' - x_0)dx'\right)\underline{u}(x_1).
\end{aligned} \tag{18}$$

The integration in the last line of Eq. (18) can be recognized as the convolution product $(g*g)(x_0-x_1)$ and the expression can therefore be simplified to

$$\underline{F}'_n(x_0) = \underline{s}_{n,i}\left(\delta_{x_0,x_1} - \frac{\underline{s}_{n,i}}{-\omega^2 M_{n,i} + \underline{s}_{n,i}}(g * g)(x_0 - x_1)dx\right)\underline{u}(x_1). \tag{19}$$

Finally, by substituting $\underline{s}_{n,i}=\omega_n^2 M_{n,i}(1+i\eta)$ and $\underline{v}(x)=i2\pi f\underline{u}(x)$ in Eq. (19) we have that

$$\underline{F}'_n(x_0) = -i2\pi f\left(\frac{f_{r,n}^2}{f^2}\right)(1 + i\eta)M_{n,i}\left(\delta_{x_0,x_1} - \frac{f_{r,n}^2(1 + i\eta)}{f_{r,n}^2(1 + i\eta) - f^2}(g * g)(x_0 - x_1)dx\right)\underline{v}(x_1). \tag{20}$$

Hence, the total impedance $\underline{z}_{\varepsilon,n}(x_0-x_1)$ of the nth set of oscillators reads

$$\underline{z}_{\varepsilon,n}(x_0 - x_1)dx = -i2\pi f\left(\frac{f_{r,n}^2}{f^2}\right)(1 + i\eta)M_{n,i}\left(\delta_{x_0,x_1} - \frac{f_{r,n}^2(1 + i\eta)}{f_{r,n}^2(1 + i\eta) - f^2}(g * g)(x_0 - x_1)dx\right). \tag{21}$$

A structural fuzzy composed of N sets of infinitely many identical oscillators as described earlier is homogenous, as the boundary impedance only depends on the distance $|x_0-x_1|$. Further, if $\varepsilon \to 0$, then Eq. (21) reduces to the boundary impedance of infinitely many identical simple oscillators as illustrated in Fig. 6(c). Examining the expression in Eq. (21), it is seen that the transfer impedances are proportional to the convolution product $(g*g)(x_0-x_1)$, as the term δ_{x_0,x_1} is zero when $x_0 \neq x$. Figure 8 shows this convolution and it is seen that the transfer impedances extends a distance of 2ε to each side of x. This means that the actual spatial memory is 4ε, and that the transfer impedances are zero for $|x_0-x| \geqslant 2\varepsilon$. Note that the areas below $g(x_0-x)$ and $(g^*g)(x_0-x_1)$ are both equal to 1.

D. Soize's local equivalent oscillator

A numerical implementation of the boundary impedance $\underline{z}_{\varepsilon,n}$ in Eq. (21) is unfortunately rather complicated due to its *nonlocal* nature. This requires for instance the use of a finite

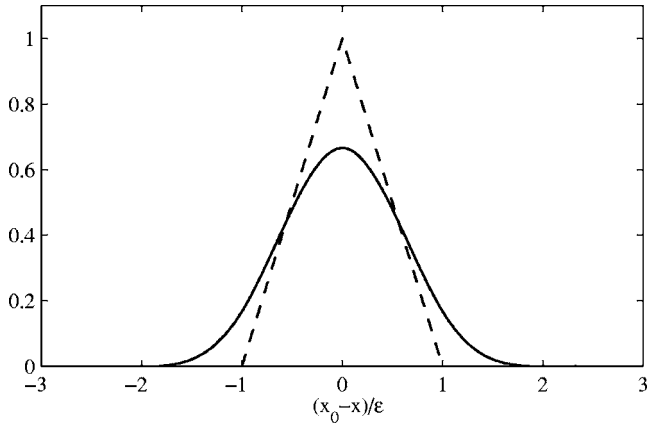


FIG. 8. Examination of the magnitude distribution of the transfer impedance. Functions: ---, $g(x_0-x)/(1/\varepsilon)$; —, $(g^* g)(x_0-x)/(1/\varepsilon)$.

element model with special fuzzy elements. As mentioned earlier, the main purpose of the fuzzy structure theory is to serve as a simple modeling tool. Therefore, Soize introduced the *equivalent local* oscillator. The idea is that the nth set of infinitely many identical equivalent oscillators, can replace the nth set of spatial oscillators. This also implies that the contact force per unit length $\underline{F}'_{\text{equ},n}(x_0)$ introduced by the local equivalent oscillator at x_0 must be equal to the contact force $\underline{F}'_{\varepsilon,n}(x_0)$ given in Eq. (8). This is achieved by introducing the so-called *equivalent coupling factor* α, which transforms all the nonlocal force contributions of the spatial oscillators into equivalent local contributions. As one can imagine, α generally varies with frequency and is both dependent on the length of the spatial memory 2ε and the motion of the master $u(x)$.

Now, let us assume that the nth set of equivalent oscillators consist of an infinite number of *identical* oscillators distributed on the fuzzy connection boundary. Further, each equivalent oscillator has a mass $M_{n,i}$, and masses are located at all positions along the x' axis. For the nth set of equivalent oscillators the boundary impedance $\underline{z}_{\text{equ},n}(x_0)=\underline{F}'_n(x_0)/\underline{v}(x_0)$ is given as[6]

$$\underline{z}_{\text{equ},n}(x_0)=\frac{\underline{s}_{n,i}}{i\omega}\left(1-\frac{\underline{s}_{n,i}}{\underline{s}_{n,i}-\omega^2 M_{n,i}}\alpha\right)=i2\pi f M_{n,i}\times\left(\frac{f_r^2(1+i\eta)(1-(f_r/f)^2(1+i\eta)(1-\alpha))}{f^2-f_r^2(1+i\eta)}\right), \tag{22}$$

where $\alpha\in]0,1]$. If α is chosen properly then the boundary impedance $\underline{z}_{\text{equ},n}(x_0)$ in Eq. (22) can replace successfully the boundary impedance $\underline{z}_{\varepsilon,n}(x_0-x_1)$ of the nth set of spatial oscillators[6] as was given in Eq. (21). The frequency variation of this equivalent impedance $\underline{z}_{\text{equ},n}(x_0)$ is shown in Fig. 9 for different values of α. As indicated in Fig. 9, we now suggest that the equivalent oscillator can be interpreted as a simple oscillator with spring stiffness $\underline{s}_1$ where the mass has been grounded by a second spring with stiffness $\underline{s}_2$. It applies that $\underline{s}_{n,i}=\underline{s}_1+\underline{s}_2$ and the mass of the grounded oscillator is $M_{n,i}$. The relationship between the impedance in Eq. (22) and the impedance of the grounded oscillator $z_{\text{ground},n}$ in Fig. 9 is $\underline{z}_{\text{equ},n}=\underline{z}_{\text{ground},n}/\alpha$ where $\alpha=\underline{s}_1/(\underline{s}_1+\underline{s}_2)$. Note that $\alpha\to 1$ when $\underline{s}_2\to 0$ and $\underline{z}_{\text{equ},n}$ will then approach the impedance of a simple oscillator, Eq. (1). Also, when $\alpha\neq 1$ the oscillator is stiffness controlled at low frequencies. It should be mentioned that the impedance of a set of spatial oscillators results in a stiffness-controlled behaviour of the master at low frequencies. The reason is that the mass-less bar supporting the point mass in Fig. 5(a) is restricted to translational motion and therefore unable to rotate. Any rotation of the master at low frequencies is therefore restricted by the springs.

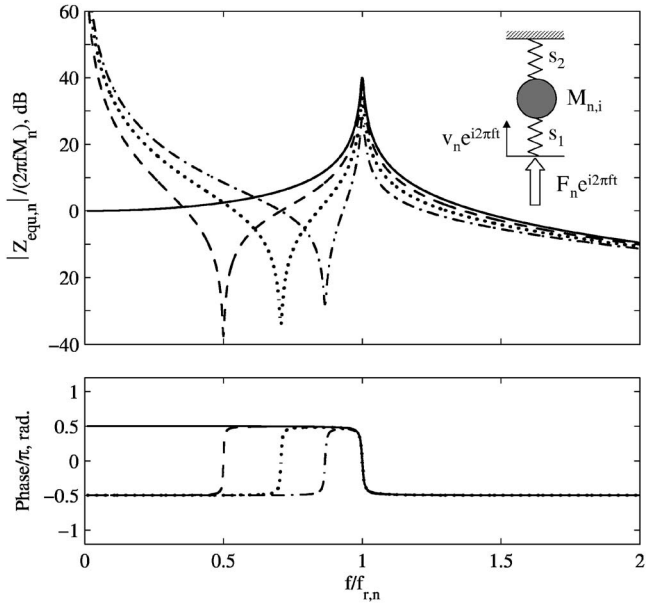


FIG. 9. Frequency variation of normalized impedance, $Z_{\text{equ},n}/(2\pi f M_n)$ for different values of the equivalent coupling factor α: —, 1; ---, 0.75; ···, 0.5; -·-, 0.25.

E. The equivalent coupling factor

Soize states that it is not self-evident that the local *equivalent oscillator* can model correctly a structural fuzzy with spatial memory.[6] And as one can imagine, α has to be chosen carefully. Finding a relationship between ε and α requires matching of boundary forces using the impedances found in Eqs. (21) and (22), frequency by frequency. Such results have been published by Soize[6] and they show α as a function of the spatial memory 2ε for different frequency bands for a simply supported beam. The authors of the present article, however, suggest that α should be determined in a more general way as a function of the ratio ε/λ where λ is the free wavelength in the master, which here is restricted to undergo one-dimensional wave motion only. By transforming Soize's data, it is revealed that a unique relationship between α and ε/λ is found. The transformed data of α as a function of ε/λ are shown in Fig. 10; these results have been fitted with a fourth-order polynomial. It should be noted that the free wavelength is defined only for sinusoidal variations, and for structures with more complicated eigenfunctions, we therefore suggest substituting λ with twice the distance between adjacent nodes.

Finally, it should be mentioned that a general and simple method of predicting α has been the subject of the authors'

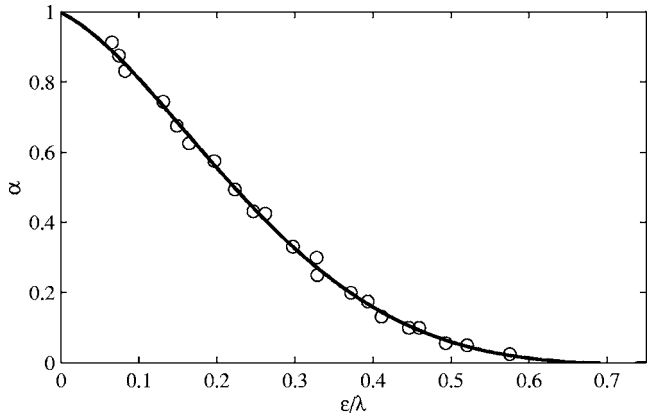


FIG. 10. Variation of the equivalent coupling factor α with ε/λ:(○), data computed from read-off results in Ref. 6; —, polynomial fit.

latest work and will shortly be submitted for publication together with a practical validation of the associated equivalent modelling method.

IV. SMOOTHED EXPRESSION FOR THE BOUNDARY IMPEDANCE OF STRUCTURAL FUZZY WITH SPATIAL MEMORY

The fuzzy structure theory developed by Soize was originally intended for finite element modeling. To determine the damping induced in the master, Soize developed his own methods based on probabilistic concepts in order to account for model uncertainties. A new proposition for a simplified deterministic method for predicting the mean damping induced by structural fuzzy with spatial memory is presented in the following. The purpose of this model is to illustrate the main effects of including spatial memory in the modeling of structural fuzzy.

As a starting point we consider a structural fuzzy consisting of N sets of spatial oscillators with different natural frequencies. The total boundary impedance of the structural fuzzy can be determined as the sum of the impedance contributions from these N sets for which the impedance of the nth set, $n \in [1,N]$, was presented in Eq. (21). The total boundary impedance $\underline{z}_{\text{fuzzy},\varepsilon}(x_0-x_1)$ thus becomes

$$\underline{z}_{\text{fuzzy},\varepsilon}(x_0-x_1) = \sum_{n=1}^{N} \underline{z}_{\varepsilon,n}(x_0-x_1). \tag{23}$$

So far, only Soize has presented a method of predicting the boundary impedance of fuzzy with spatial memory.[6] Now, applying the same approach as in Sec. II a *deterministic* expression for the boundary impedance of the structural fuzzy with spatial memory can be found. Let us approximate the expression in Eq. (23) by infinitely many sets of spatial oscillators resonating between $f_{r,\text{lower}}$ and $f_{r,\text{upper}}$. Hereby the general expression for the total fuzzy boundary impedance attached at area A of the master becomes

$$\underline{z}_{\text{fuzzy},\varepsilon}(x_0-x_1) = \int_{f_{r,\text{lower}}}^{f_{r,\text{upper}}} z_{\varepsilon}(x_0-x_1,f_r)df_r, \tag{24a}$$

or, by inserting the expression from Eq. (21) we get

$$\begin{aligned}\underline{z}_{\text{fuzzy},\varepsilon}(x_0-x_1)dx = &-\frac{i2\pi f}{A}\int_{f_{r,\text{lower}}}^{f_{r,\text{upper}}}\left(\frac{f_r^2}{f^2}\right)(1+i\eta)m_{\text{fuzzy}}(f_r)\\ &\times\left(\delta_{x_0,x_1}-\frac{f_r^2(1+i\eta)}{f_r^2(1+i\eta)-f^2}\right.\\ &\left.\times(g*g)(x_0-x_1)dx\right)df_r,\end{aligned} \tag{24b}$$

where $m_{\text{fuzzy}}(f_r)df_r$ again represents the total mass of the fuzzy resonating between the frequencies f_r and f_r+df_r. Moreover, the same simple approach can be applied to find the equivalent boundary impedance $\underline{z}_{\text{fuzzy,equ}}(x_0)$ if M_n is replaced by $m_{\text{fuzzy}}(f_r)df_r$. A smoothed version of the equivalent boundary impedance then yields

$$\begin{aligned}\underline{z}_{\text{fuzzy,equ}}(x_0-x_1) &= \int_{f_{r,\text{lower}}}^{f_{r,\text{upper}}} z_{\text{equ}}(x_0,f_r)df_r\\ &= -\frac{i2\pi f}{A}\int_{f_{r,\text{lower}}}^{f_{r,\text{upper}}}\left(\frac{f_r^2}{f^2}\right)(1+i\eta)m_{\text{fuzzy}}(f_r)\\ &\quad\times\left(1-\frac{f_r^2(1+i\eta)}{f_r^2(1+i\eta)-f^2}\alpha\right)df_r.\end{aligned} \tag{25}$$

It should be noted that the equivalent coupling factor α generally is a function of frequency. Nevertheless, according to Fig. 10 it is seen that α is constant for a specific value of ε/λ. With the new expression for the boundary impedance of structural fuzzy with spatial memory in Eq. (25) it is therefore possible to model and examine the effects of structural fuzzy with spatial memory in a simple way. For simple cases of mass distributions $m_{\text{fuzzy}}(f_r)$ the integration can be done analytically, whereas the use of more realistic mass distributions will require a numerical integration.

V. BEAM MASTER STRUCTURE WITH STRUCTURAL FUZZY

The influence of structural fuzzy with and without spatial memory will now be illustrated by a numerical example. The finite element method[13] is being used for solving the harmonically forced vibration response of a simply supported Bernoulli–Euler beam, which is considered as the master structure. A fuzzy substructure is attached on the whole length L of the beam, so that $L_{\text{fuzzy}}=L$. The damping loss factor of the beam is 0.005 and the loss factor of the fuzzy oscillator springs is 0.03. The resonating mass per unit frequency, $m_{\text{fuzzy}}(f_r)$, is taken to follow a normal distribution, giving

$$m_{\text{fuzzy}}(f_r) = \frac{M_{\text{fuzzy}}}{\text{std}\sqrt{2\pi}}e^{-(f_{r0}-f_r)^2/(2\cdot\text{std}^2)}, \tag{26}$$

where f_{r0} is the center frequency and std is the standard deviation. This chosen mass distribution is shown in Fig. 11 as a function of the beam's nondimensional frequency Ω being defined as

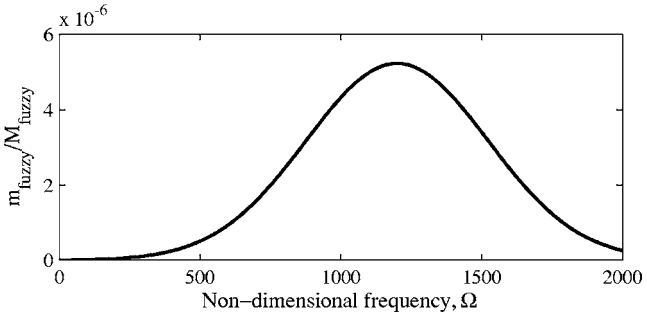


FIG. 11. Normalized resonating mass per unit frequency described by a normal distribution with a centre frequency f_{r0} corresponding to $\Omega=1200$ and a standard deviation *std* of $\Omega=750$.

$$\Omega = 2\pi f \sqrt{\frac{12\rho}{E}}\frac{L^2}{h}, \tag{27}$$

where h is the beam thickness, and ρ and E is the density and Young's modulus of the beam material, respectively. For this distribution the bounding frequencies $f_{r,\text{lower}}$ and $f_{r,\text{upper}}$ correspond to $\Omega=0$ and $\Omega=\infty$, respectively. The center frequency f_{r0} corresponds to $\Omega=1200$ and the standard deviation std is $\Omega=750$. Moreover, the total mass of the fuzzy M_{fuzzy} is taken to be one-twentieth of the beam mass, ρSL, where S is its cross-sectional area.

The boundary impedance of the fuzzy, Eq. (25), is computed by numerical integration; it is assumed that the equivalent coupling factor α is constant with frequency, which means that the ratio ε/λ is constant, whereas ε and λ decrease with frequency at the same rate. Figure 12 shows computed results of the fuzzy boundary impedance $\underline{z}_{\text{fuzzy,equ}}(x_0)$ as a function of the nondimensional frequency for different values of α.

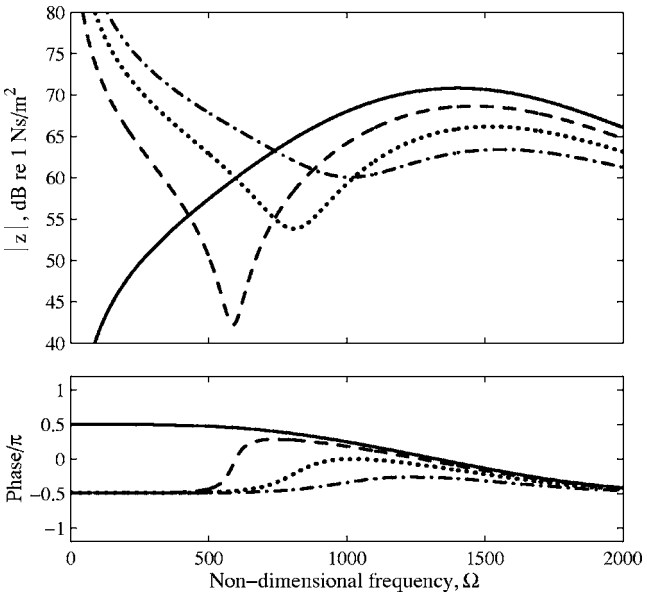


FIG. 12. Amplitude and phase of the fuzzy boundary impedance $\underline{z}_{\text{fuzzy,equ}}$. The total mass of the fuzzy is 1/20 of the master and has a normal-distributed resonating mass per unit frequency. Results are shown for different values of the equivalent coupling factor α: —, 1; ---, 0.75; ⋯, 0.5; -·-, 0.25.

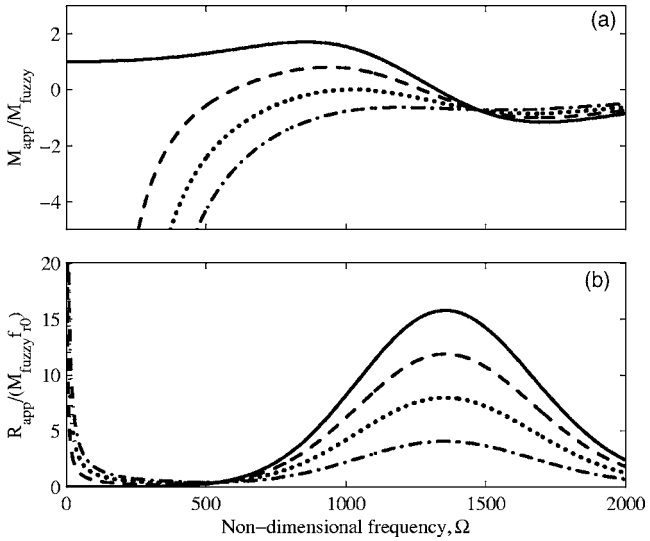


FIG. 13. (a) Normalized apparent mass, $M_{\text{app}}=\text{Im}(\underline{z}_{\text{fuzzy,equ}})/\omega$, and (b) normalized apparent damping, $R_{\text{app}}=\text{Re}(\underline{z}_{\text{fuzzy,equ}})$, of the fuzzy boundary impedance for different values of the α: —, 1; ---, 0.75; ⋯, 0.5; -·-, 0.25.

From these results a number of observations can be made. First, it is clearly seen that the structural fuzzy *without* memory is mass-controlled at frequencies below $\Omega=500$, as the amplitude slope of the boundary impedance is positive and the phase equals $\pi/2$. This is not the case for structural fuzzy *with* memory, which is clearly spring-like with a negative amplitude slope and a phase of $-\pi/2$; this will always be the case when $\alpha<1$. Second, between $\Omega=500$ and $\Omega=2000$ the phase of the fuzzy *without* memory changes smoothly from $\pi/2$ to approximately $-\pi/2$. During this change the real part of the impedance exhibits high values and the fuzzy therefore has a high damping effect. The impedance for $\alpha=0.75$ has a sudden phase change of π after $\Omega=500$ and hereafter it closely follows the phase of the fuzzy for $\alpha=1$. Examining the other two cases of smaller α values, it is found that the phase change is less pronounced and occurs at a higher frequency; therefore the fuzzy never reaches true mass-like behavior. It is also seen that the amplitude of the impedance becomes significantly lower when α is decreased, and this results in a weakening of the effect of the fuzzy. This is clearly observed for frequencies above $\Omega=1000$.

Inspired by Pierce *et al.*[8] the behavior and actual damping effect of the fuzzy is conveniently demonstrated by examining the corresponding apparent mass, $M_{\text{app}}(f)=\text{Im}(\underline{z}_{\text{fuzzy,equ}})/(2\pi f)$, and apparent damping $R_{\text{app}}(f)=\text{Re}(\underline{z}_{\text{fuzzy,equ}})$. Both quantities are shown in normalized form in Fig. 13 for different values of α. Figure 13(a) shows that the apparent mass of the fuzzy *without* memory is equal to the total mass of the fuzzy at $\Omega=0$, and up to around $\Omega=1170$ the apparent mass is higher than the total mass. Further, around $\Omega=1340$ the quantity becomes negative, which indicates a spring-controlled behavior. The three cases of fuzzy *with* memory clearly differ from this behaviour by being mostly spring-like in the whole frequency range. Above, say $\Omega=1350$, the apparent mass is very close in all cases. The apparent damping is plotted in Fig. 13(b), and it is seen that the damping effect of the fuzzy decreases significantly

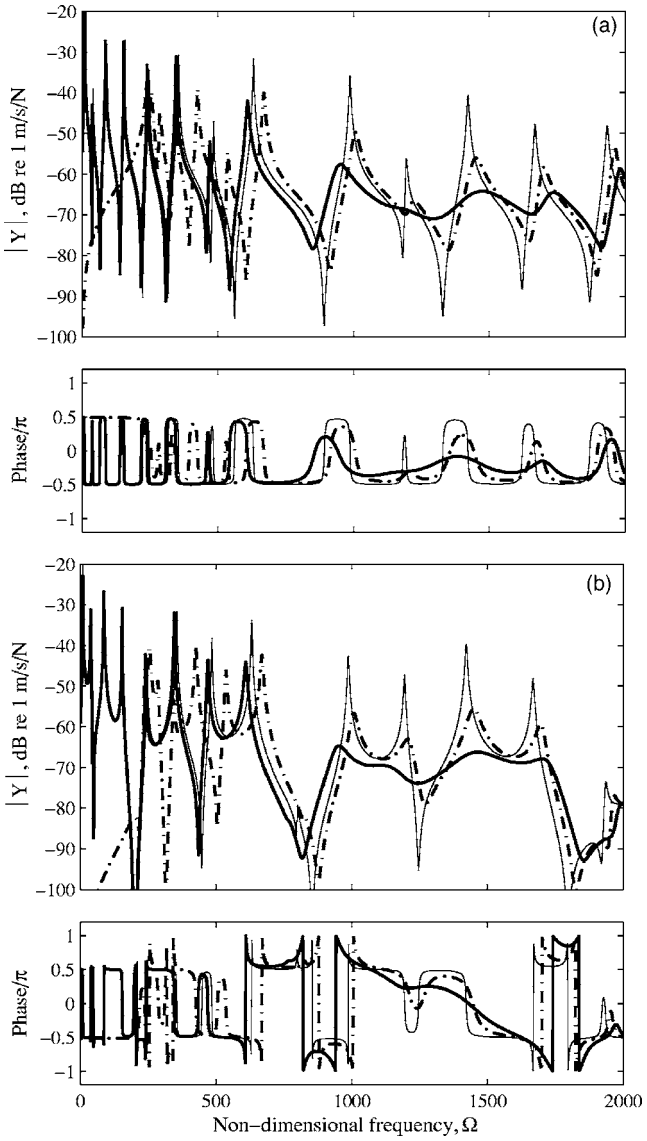


FIG. 14. Vibration velocity response per unit harmonic force, $\underline{Y}(x,x_0)=\underline{v}(x)/\underline{F}(x_0)$, at (a) $x=0.445L$ and (b) $x=0.785L$ of a simply supported beam excited at $x_0=0.445L$. Condition: —, without structural fuzzy and with structural fuzzy for α: —, 1; ---, 0.75; ···, 0.5; -·-, 0.25.

when α is reduced. The reason for this is that a reduction of α corresponds to an increase of ε. Due to the rotational motion in the master, the spring elements in the spatial oscillators counteract one another when $\varepsilon>0$. This also implies that the impedance of the structural fuzzy is reduced significantly, and this results in a lower dissipation induced in the master. For the chosen mass distribution, the maximum damping effect occurs around $\Omega=1360$, which is higher than the center frequency of $\Omega=1200$. At low frequencies below $\Omega=100$ it is seen that the apparent damping becomes very high. This nonphysical behavior is caused by the approximate modeling of the fuzzy boundary impedance, which goes toward infinity at $\Omega=0$.

Finally, Fig. 14 shows results for the vibration velocity response of the simply supported master beam, without and with the fuzzy substructure, which was discussed earlier. The vibration responses at $x=0.445L$ and $x=0.785L$ are given in terms of the beam's mobility $\underline{Y}(x,x_0)=\underline{v}(x)/\underline{F}(x_0)$ for harmonic force excitation at $x_0=0.445L$. Considering both Figs. 14(a) and 13(b) it is seen that the fuzzy *without* spatial memory, $\alpha=1$, introduces a high damping in the master beam. The effect of this stretches over a relatively wide frequency band, which covers at least six flexural modes of the beam. From around $\Omega=950$ to $\Omega=1740$ the damping effect is very pronounced and the vibration velocity of the master is dampened by up to 25 dB. Further, this structural fuzzy, causes the resonances of the beam structure to shift downwards into the region where the fuzzy is mass-like. This is the case in the range from $\Omega=0$ to $\Omega=1340$. Above $\Omega=1340$ the fuzzy becomes spring controlled and the resonances are shifted upwards.

Next, considering the case of structural fuzzy *with* a spatial memory of $\alpha=0.25$, it is evident that the spatial memory significantly decreases the damping effect of the fuzzy. In the main damping region of the fuzzy, the damping effect is reduced by almost 10 dB. This is in good agreement with the apparent damping being reduced by a factor of 3.5 by a decreasing α from 1 to 0.25, see Fig. 13(b). Further, it is seen that the strong spring-like behavior of the fuzzy at low frequencies dominates the response of the structure up to about $\Omega=140$. As the fuzzy is spring-like in the whole frequency range, it only causes the resonance frequencies of the structure to shift upwards. Fig. 14(b) shows the response at a position at some distance from the drive point. The damping effect of the fuzzy both without and with spatial memory is seen to be very similar to what was discussed earlier for the response at the drive point location. This illustrates and confirms that the influence of the structural fuzzy is global, and not specifically associated with the drive point.

VI. SUMMARY AND DISCUSSION

Soize's method of including spatial memory in structural fuzzy has been thoroughly examined and exemplified in the present paper. Additional illustrations and a derivation of the fuzzy boundary impedance have been given in order to explain the ideas governing the method. To simplify the fuzzy modeling, Soize replaces the non-local spatial oscillator with a local equivalent oscillator. In the present article this oscillator has been given a physical interpretation. Further, it has been suggested that the so-called equivalent coupling factor, which transforms the nonlocal boundary impedance into a local impedance, can be determined as a function of the ratio between spatial memory and the free wavelength in the master.

The fuzzy boundary impedance, which includes spatial memory, has been derived deterministically by using a simple smoothing approach. This method assumes that the fuzzy is described in terms of a predefined distribution of resonating mass per unit frequency. The developed method is straightforward and it has been demonstrated that a prediction of the overall vibrations of the master can be made in a simple way.

From numerical simulations of the response of a simply supported beam with structural fuzzy and different amounts of spatial memory, it has been found that the spatial memory

significantly reduces the damping introduced in the master. Further, for the case studied it can be included that the memory in some cases completely eliminates the damping effect of the fuzzy.

Various assumptions have been made in this paper in order to illustrate more clearly the effects of spatial memory in the structural fuzzy. This includes the hypothesis of modeling spatial memory by use of an equivalent spatial oscillator. A validation of this hypothesis and a discussion of its limitations clearly remain to be made. Also, a simple way of determining the equivalent coupling factor is as yet absent in the open literature. However, both of these two topics will be dealt with in a companion paper, which soon will be submitted for publication. Finally, further investigations are still required concerning practical questions of how one can determine the distribution of resonating mass per unit frequency as well as the amount of spatial memory in real-life engineering structures.

[1]J. Ormondroyd and J. P. Den Hartog, "Theory of dynamic vibration absorber," Trans. ASME **50**, APM-241 (1928).

[2]*Noise and Vibration*, edited by R. G. White and J. G. Walker (Ellis Horwood, Chichester, UK, 1982), Chap. 25.

[3]D. F. Mead, *Passive Vibration Control*, (Wiley, Chichester, UK, 1999), Chap. 8.

[4]C. Soize, "Probabilistic structural modeling in linear dynamic analysis of complex mechanical systems, Part I," Rech. Aerosp. (English edition), **5**, 23–48 (1986).

[5]F. Chabas, A. Desanti, and C. Soize, "Probabilistic structural modelling in linear dynamic analysis of complex mechanical systems, part II," Rech. Aerosp. (English edition), **5**, 49–67 (1986).

[6]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).

[7]C. Soize, "Vibration damping in low-frequency range due to structural complexity. A model based on the theory of fuzzy structures and model parameters estimation," Comput. Struct. **58**, 901–915 (1995).

[8]A. D. Pierce, V. W. Sparrow, and D. A. Russell, "Fundamental structural-acoustic idealizations for structures with fuzzy internals," J. Vibr. Acoust. **117**, 339–348 (1995).

[9]M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).

[10]C. Soize, "Estimation of fuzzy substructure model parameters using the mean power flow equation of the fuzzy structure," J. Vibr. Acoust. **120**, 279–286 (1998).

[11]C. Soize, "Estimation of fuzzy structure parameters for continuous junctions," J. Acoust. Soc. Am. **107**, 2011–2020 (2000).

[12]A. D. Pierce, "Resonant-frequency-distribution of internal mass inferred from mechanical impedance matrices, with application to fuzzy structure theory," J. Vibr. Acoust. **119**, 324–333 (1997).

[13]R. Cook, D. S. Malkus, M. F. Plesha, and R. J. Witt, *Concepts and Applications of Finite Element Analysis* (Wiley, New York, 2002).

A study of modal characteristics and the control mechanism of finite periodic and irregular ribbed plates[a)]

Tian Ran Lin[b)]
School of Engineering Systems, Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia

(Received 12 April 2007; revised 7 November 2007; accepted 5 December 2007)

An analytical solution is presented in this paper to investigate the control mechanism and modal characteristics of finite periodic and irregular ribbed plates. Peak responses of a finite periodic ribbed plate were examined where they were grouped into two sets of propagation zones according to the coupling mechanism at beam/plate interfaces. Details of modal characteristics in pass bands of the periodic ribbed plate were elucidated and the control mechanism was discussed. Modes in each pass band that are governed by shear force couplings were characterized by one of the beam flexural modes whose modal responses could be represented approximately by those of the corresponding orthotropic plate modes. Modes in the second set of pass bands were found to retain the resonance frequencies of the corresponding modes of the unribbed base plate. Higher order orthotropic plate modes were also identified, which could not be grouped into any pass bands defined by the classical periodic theory. The control mechanism leading to vibration confinement in disordered and irregular ribbed plates was also discussed. It was found that beam spacing irregularity attributes to localization of the group of modes associated with flexural wave couplings but not the group of modes associated with moment couplings.
 [DOI: 10.1121/1.2828220]

PACS number(s): 43.40.Dx, 43.40.At, 43.20.Bi, 43.20.Ks [EGW] Pages: 729–737

I. INTRODUCTION

Vibration characteristics of periodic structures have been a long-standing research topic for many decades due to their broad application and interesting wave phenomenon. Traditionally, the vibration of a periodic structure is analyzed by solving each section (bay) of the periodic system successively. The most common approaches are propagation wave approach[1–3] and transfer matrix method.[4,5] The propagation wave approach is associated with the development of the Bloch theorem in solid state physics,[6] which can be applied most efficiently to infinite or semi-infinite periodic systems. On the other hand, the transfer matrix method provides a more convenient approach to the analysis of finite periodic structures where boundary conditions of a finite periodic system can be incorporated into the transfer matrix with ease. Other methods, such as Z-transform method,[7] energy method[8] and Fourier transform method[9,10] have also been developed.

Heckl[1] affirmed that a periodic ribbed plate can be treated as an orthotropic plate when the distance between the adjacent ribs is less than a quarter of the shortest plate bending wavelength. Arising from calculation of wave propagating constants, he found that pass bands of a periodic ribbed plate can be divided into two classes, one is near the resonance frequencies of the unribbed plate and the other is near the frequencies of total transmission for a plate with one beam. Mead[3] showed that the bounding frequencies of wave propagation zones of a periodic system can be determined from element receptance matrices of the system. Furthermore, Mead[11] provided an analytical solution to predict wave propagation in infinite periodic supported beams and infinite periodic plates by using phased array receptance functions. Analytical equations were obtained for calculating wave propagation constants. The general wave propagation mechanism of pass bands of an infinite periodic stiffened plate based on wave propagation constants was also briefly discussed in his work. However, details of modal characteristics and the control mechanism of a finite periodic ribbed plate have never been examined. These are investigated in this work to improve the understanding of band pass/stop properties of periodic ribbed plates in hope that it will lead to better design of noise and vibration control systems-for instance, systems that employ either passive or active control methods and use either force or moment as an active control source.

An implicit assumption made in relation to periodic structures postulates that the structure is formed by identical periodic elements. In contrast, disorders due to manufacture imperfections, installation errors or arising directly from the design of a structure are common for most practical periodic systems. Vibration confinement resulting from disordered and irregular ribbed structures has attracted increasing attention because of its growing importance in vibration control of engineering structures. Mead and Bansal[12] studied free wave motions in an infinite mono-coupled periodic system with single disorder by utilizing the end receptance of the disorder and the characteristic wave receptance of the peri-

[a)]Some of the results were presented in "Vibration localization of finite plates with disordered and irregular rib spacing," Proceedings of Inter-Noise 2006, Hawaii, Dec. 2006.
[b)]Electronic mail: trlin@qut.edu.au

odic system. The disorder of the infinite periodic system was considered as two semi-infinite periodic systems connecting through the disordered element. They found that the disorder always results in reduced transmission of flexural waves in the propagation zones of a periodic system but can also lead to increased transmission in attenuation zones. Hodges[13] first introduced Anderson localization,[14] discovered originally for solid state physics, to the acoustic context and showed that vibration localization could be produced in a one-dimensional structure depending upon the effective dimension of the structure and the extent of structural irregularities. Hodges and Woodhouse[15] studied the phenomenon further in a weakly disordered one-dimensional chain of oscillators. Their experimental investigation on a vibrating string with irregularly spaced attaching masses showed good agreement with that of the theoretical prediction. Photiadis[16] studied Anderson localization in an infinite fluid loaded plate with an irregular array of line attachments by applying Green's function. His work indicated that the long range coupling resulting from bulk wave in the fluid is small compared to the nearest neighbor coupling of fluid loaded wave. The irregularity in a ribbed plate thus produces Anderson localization regardless of whether the array of line attachments is dense or sparse. Photiadis[17] extended the work further to vibration confinement of an infinite ribbed thin cylindrical shell in the mid-frequency range (above the ring frequency but below half of the critical frequency). He showed that localization similar to those found for fluid loaded plates can be produced for all azimuthal modes in this frequency range. The results were verified experimentally for higher order azimuthal modes ($n>10$) by Photiadis and Houston[18] using near field acoustic holography measurements.[19] At about the same time of Photiadis's work,[16] Sobnack and Crighton[20] examined the Anderson localization effect on wave transmission in an infinite fluid-loaded membrane supported by a number of irregularly spaced ribs by employing Green's function and statistical methods. The localization lengths for both small and large disorder limits were calculated and discussed. Effects of spacing irregularities on the dispersive pattern of an infinite ribbed membrane were briefly discussed by Maidanik and Dickey,[10] though no detailed explanation was provided in their study.

The study of vibration localization in infinite ribbed structures has greatly improved the understanding of the control mechanism of wave attenuation in a structure. However, when the structural system has finite extents (e.g., a few wavelengths long), the boundary wave reflection increases the complexity of the analysis. Vibration confinement of such finite systems is not well known. Furthermore, because only shear force coupling was considered in Refs. 16 and 20, their results were only valid for one-dimensional mono coupled systems. Photiadis[17] included both force and moment couplings in the analysis of infinite ribbed cylindrical shells. However, the cross coupling between the force and moment was neglected. This prevents the energy exchange between wave components governed by the force and moment couplings and can lead to different wave transmission and reflection at the interfaces. Vibration localization of finite ribbed plates considering a full coupling matrix is studied in the second half of this work.

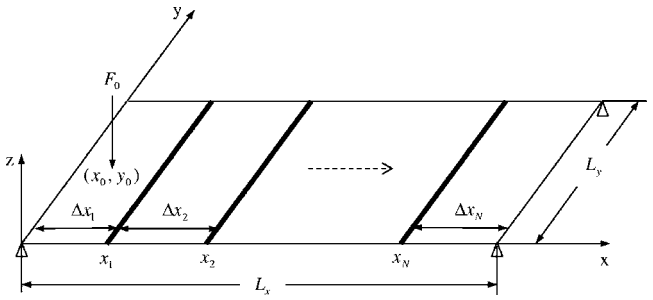


FIG. 1. Model description and the coordinate system of the finite ribbed plate.

In Sec. II, a simple analytical solution is presented to predict the vibration response of finite periodic and irregular ribbed plates by employing the modal expansion technique similar to that described by Lin and Pan.[21,22] In contrast to conventional approaches where each periodic section is solved successively, our work considers the entire plate as a single entity, and the stiffened beams are subsequently added onto the plate via the force and moment couplings at beam/plate interfaces. The advantage of this method is that the solution can be applied to both periodic and irregular ribbed plates with the same matrix formulation. Interactions of shear force couplings, moment couplings and the cross couplings between the force and the moment at the interfaces are clearly defined. Contributions of the modal coupling force and moment at each rib location to the ribbed plate response are explicitly stated in the solution. Three numerical cases are examined in Sec. III. First, modal characteristics of a finite periodic ribbed plate are classified and the control mechanism is discussed. A single rib spacing disorder is imposed to the periodic ribbed plate in the second case where the control mechanism leading to vibration attenuation in pass bands is discussed. Finally, the dependency of vibration confinement upon the coupling mechanism at the interfaces of a finite irregular ribbed plate is examined. The main findings are summarized in Sec. IV.

II. FORMULATION

A finite ribbed plate model is shown in Fig. 1 where a rectangular plate is simply supported on all its edges and is reinforced by N stiffened beams, which are either periodically or irregularly distributed on the plate surface and with their neutral axes parallel to the pair of short plate edges. It is assumed that the beams are symmetrical so that the beam flexural and torsional vibrations are uncoupled. We also assume that the beams are well separated from each other. The ribbed plate is excited by a point force (F_0) at plate location (x_0, y_0).

Using a thin plate vibration model, the governing equation of the bending displacement (W) of the base plate can be written as

$$\nabla^4 W - k_p^4 W = \frac{F_0}{D}\delta(x-x_0)\delta(y-y_0) - \sum_{i=1}^{N}\frac{q_i}{D}\delta(x-x_i) - \sum_{i=1}^{N}\frac{m_i}{D}\delta'(x-x_i), \quad (1)$$

where $k_p=(\rho_s\omega^2/D)^{1/4}$ and $D=E_p h^3/12(1-\nu^2)$ are the plate bending wave number and stiffness. E_p, h, ν and ρ_s are, respectively, the Young's modulus, thickness, Poison's ratio and surface mass of the plate; q_i and m_i are the coupling force and moment per unit length at the ith beam/plate interface; x_i is the location of the ith beam on the plate and the prime ($'$) indicates a spatial derivative.

The governing equations of the flexural and torsional displacements of the ith beam (U_i, θ_i) are given by

$$\frac{\partial^4 U_i}{\partial y^4} - k_{bi}^4 U_i = \frac{q_i}{B_{bi}}, \quad i=1,2,\ldots,N, \quad (2)$$

and

$$\frac{\partial^2 \theta_i}{\partial y^2} + k_{ti}^2\theta_i - R_i\frac{\partial^4\theta_i}{\partial y^4} = \frac{m_i}{T_i}, \quad i=1,2,\ldots,N, \quad (3)$$

where $k_{bi}=(\rho_{Li}\omega^2/B_{bi})^{1/4}$ and $k_{ti}=(\rho_{bi}I_{pi}\omega^2/T_i)^{1/2}$ are, respectively, the flexural and torsional wave numbers. The term ρ_{Li} is the mass per unit length and $\rho_{bi}I_{pi}$ is the mass moment of inertia per unit length. B_{bi} and T_i are the flexural and torsional stiffness and R_i is the warping to torsional stiffness ratio of the ith beam.

Eqs. (1)–(3) can be solved by modal expansion of the plate bending displacement (W), the flexural and torsional displacements of the beams (U_i and θ_i) as[23]

$$W = \sum_m\sum_n w_{m,n}\phi_m(x)\phi_n(y), \quad (4)$$

$$U_i = \sum_n u_{ni}\phi_n(y), \quad i=1,2,\ldots,N \quad (5)$$

and

$$\theta_i = \sum_n \theta_{ni}\phi_n(y), \quad i=1,2,\ldots,N, \quad (6)$$

where $\phi_m(x)=\sin(k_m x)$, $\phi_n(y)=\sin(k_n y)$, $k_m=m\pi/L_x$ and $k_n=n\pi/L_y$ are the mode shape functions and trace wave numbers of the simply supported rectangular plate with respect to the two orthogonal plate edge directions. The term $w_{m,n}$ is the modal coefficient of the (m,n)th plate bending mode, u_{ni} and θ_{ni} are the modal coefficients of the nth beam flexural and torsional vibration modes.

Substituting Eqs. (4)–(6) into Eqs. (1)–(3), and applying the compatibility conditions at the beam/plate interfaces $(U_i(y)=W(x_i,y)$ and $\theta_i(y)=\partial W/\partial x(x_i,y)$, $i=1,2,\ldots,N)$, for each modal index n, we have

$$\begin{Bmatrix}\{Q\}\\ \{M\}\end{Bmatrix}_n = \begin{bmatrix}[A] & [C]\\ [C]^T & [B]\end{bmatrix}_n^{-1}\begin{Bmatrix}F_0\phi_n(y_0)\{H^{ss}\}\\ F_0\phi_n(y_0)\{H^{sc}\}\end{Bmatrix}_n, \quad (7)$$

where the superscript T indicates a matrix transpose.

$$\{Q\} = \lfloor Q_1\ Q_2\ \ldots\ Q_{N-1}\ Q_N\rfloor^T \quad (8)$$

and

$$\{M\} = \lfloor M_1\ M_2\ \ldots\ M_{N-1}\ M_N\rfloor^T \quad (9)$$

are the modal coupling force and moment vectors to be determined, $Q_i=\int_0^{L_y} q_i\phi_n(y)dy$ and

$$M_i = \int_0^{L_y} m_i\phi_n(y)dy, \quad i=1,2,\ldots N.$$

In Eq. (7), $[A]$ is a $N\times N$ square submatrix representing the interactions of shear force couplings between the beam/plate interfaces, the elements of which are given by

$$A_{i,j} = \sum_m \frac{\phi_m(x_i)\phi_m(x_j)}{G_{m,n}}, \quad i,j=1,2,\ldots,N, \quad i\neq j \quad (10a)$$

and

$$A_{i,j} = \sum_m \frac{\phi_m^2(x_i)}{G_{m,n}} + \frac{1}{G_{ni}}, \quad i,j=1,2,\ldots,N, \quad i=j, \quad (10b)$$

where $G_{m,n}=D\Lambda_{m,n}(k_{m,n}^4-k_p^4)$, $\Lambda_{m,n}=L_xL_y/4$. The term $k_{m,n}$ is the modal wave number of the (m,n)th plate bending mode and is equal to $\sqrt{k_m^2+k_n^2}$. $G_{ni}=B_{bi}\Lambda_n(k_n^4-k_{bi}^4)$ and $\Lambda_n=L_y/2$.

$[B]$ is a $N\times N$ square submatrix representing the interactions of moment couplings between the beam/plate interfaces, the elements of which are given by

$$B_{i,j} = \sum_m \frac{\phi_m'(x_i)\phi_m'(x_j)}{G_{m,n}}, \quad i,j=1,2,\ldots,N, \quad i\neq j \quad (11a)$$

and

$$B_{i,j} = \sum_m \frac{\phi_m'^2(x_i)}{G_{m,n}} - \frac{1}{G_{Tni}}, \quad i,j=1,2,\ldots,N, \quad i=j, \quad (11b)$$

where $G_{Tni}=T_i\Lambda_n(R_ik_n^4+k_n^2-k_{ti}^2)$.

$[C]$ is a $N\times N$ cross coupling submatrix, which elements are given by

$$C_{i,j} = \sum_m \frac{\phi_m(x_i)\phi_m'(x_j)}{G_{m,n}}, \quad i,j=1,2,\ldots,N. \quad (12)$$

Elements of the external force vector on the right-hand side of Eq. (7) are given by

$$H_i^{ss} = \sum_m \frac{\phi_m(x_0)\phi_m(x_i)}{G_{m,n}}, \quad i=1,2,\ldots,N \quad (13)$$

and

$$H_i^{sc} = \sum_m \frac{\phi_m(x_0)\phi_m'(x_i)}{G_{m,n}}, \quad i=1,2,\ldots,N. \quad (14)$$

Caution should be taken in order to avoid the inaccuracy caused by computer rounding errors when inverting the matrix in Eq. (7) if damping is accounted for. The technique of separating the real and imaginary parts of variables prior to matrix inversion as suggested by Lin[22] can be adopted to improve the prediction accuracy in this case.

Once modal coupling forces and moments at the interfaces are determined, the modal coefficient of the plate response can be calculated by

$$w_{m,n} = \frac{F_0 \phi_m(x_0)\phi_n(y_0) - \sum_{i=1}^{N} [\phi_m(x_i) Q_i + \phi'_m(x_i) M_i]}{G_{m,n}}. \quad (15)$$

The first term in Eq. (15) is the modal response of the corresponding un-ribbed plate to the same excitation and the summation term is the contribution of modal coupling forces and moments at the beam/plate interfaces to the plate response. The plate response can now be calculated from Eq. (4) while the flexural and torsional vibration of each beam can be calculated from Eqs. (5) and (6).

By integrating the vibration response over the plate surface area of each section, the time averaged, steady state vibration energy of the ith plate section (bounded by two consecutive stiffened beams but excluding the energy in the beams) of the ribbed plate can be calculated by

$$\langle \bar{T} \rangle_P^i = \frac{\rho_s \omega^2 \Lambda_n}{2} \sum_n \int_{x_{i-1}}^{x_i} \left[\sum_{m'} w_{m',n} \phi_{m'}(x) \right] \times \left[\sum_m w_{m,n} \phi_m(x) \right]^* dx, \quad i = 1,2, \ldots (N+1), \quad (16)$$

where $x_{i-1}=0$ when $i=1$ and $x_i=L_x$ when $i=(N+1)$. For simply supported boundary conditions, the integral in Eq. (16) can be solved analytically.

If we consider the ith bay of the ribbed plate consisting of the ith plate section and half of the beam on each section boundary, the total kinetic energy of the bay becomes

$$\langle \bar{T} \rangle^i = \langle \bar{T} \rangle_P^i + \frac{\langle \bar{T} \rangle_B^{i-1} + \langle \bar{T} \rangle_B^i}{2} + \frac{\langle \bar{T} \rangle_T^{i-1} + \langle \bar{T} \rangle_T^i}{2}, \quad i = 1,2, \ldots, (N+1), \quad (17)$$

where

$$\langle \bar{T} \rangle_B^i = \frac{1}{2} \int_0^{Ly} \rho_{Li} \dot{U}_i \dot{U}_i^* dy = \frac{\rho_{Li} \omega^2 \Lambda_n}{2} \sum_n \left| \sum_m w_{m,n} \phi_m(x_i) \right|^2 \quad (18)$$

is the flexural vibration energy, and

$$\langle \bar{T} \rangle_T^i = \frac{1}{2} \int_0^{Ly} \rho_{bi} I_{Pi} \dot{\theta}_i \dot{\theta}_i^* dy = \frac{\rho_{bi} I_{Pi} \omega^2 \Lambda_n}{2} \sum_n \left| \sum_m w_{m,n} \phi'_m(x_i) \right|^2 \quad (19)$$

is the torsional vibration energy of the ith beam, $\langle \bar{T} \rangle_B^0 = \langle \bar{T} \rangle_T^0 = \langle \bar{T} \rangle_B^{N+1} = \langle \bar{T} \rangle_T^{N+1} = 0$.

For periodic ribbed plates where the beams are well separated from each other, the vibration energy of each stiffened beam contributes to only a small fraction of the total vibration energy of the associated bay. Therefore, beam vibration energies are not included in the kinetic energy distribution of the corresponding periodic bay in the subsequent analysis.

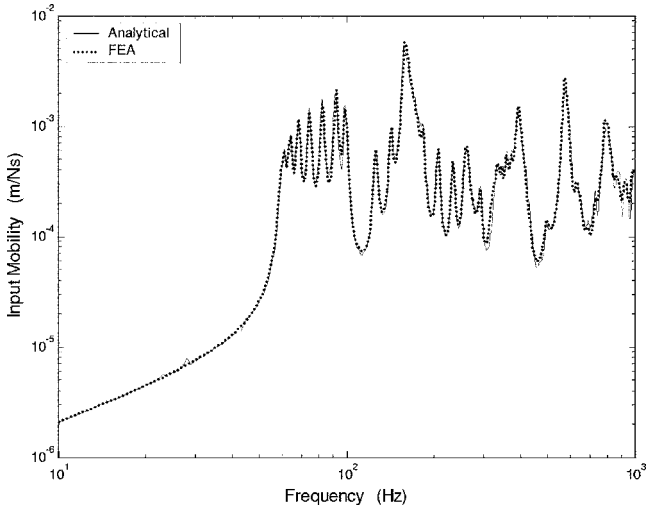


FIG. 2. Point force input mobility of the finite periodic ribbed plate.

III. RESULTS AND DISCUSSION

The finite periodic ribbed plate used in the numerical study is composed of a simply supported rectangular plate and nine stiffened beams ($N=9$). The beams are periodically distributed on the plate (with their neutral axis parallel to the pair of short plate edges) and divide the plate into ten equal sections. All structural components of the ribbed plate are made of aluminum with material properties $E=7.1 \times 10^{10}$ N/m^2, $\rho=2660$ Kg/m^3, $\nu=0.3$. The rectangular plate has a surface area of $S=5 \times 1$ m^2 and is 8 mm thick. For simplicity, beams with uniform rectangular cross section ($A_{bi}=b_i \times t_i=80 \times 5$ mm^2) are used in the study. The chosen structural properties give rise to the beam/plate flexural stiffness ratio $B_i/D=4.55$ and the plate bending/beam torsional stiffness ratio $D/T_i=38.1$. A normal unit point force is applied at plate location $(x_0, y_0)=(0.2\text{ m}, 0.3\text{ m})$ in the first section (the source section) of the ribbed plate. Structural responses due to point force excitations applied at other plate locations were also calculated but are not shown in this paper because of the similarity found in the results. A moderate damping value ($\eta=0.03$) is used for all structural components in the input mobility calculation. The result is compared to that obtained from finite element analysis as shown in Fig. 2. Good agreement is found for the entire frequency range of investigation (1–1000 Hz). Zero structural damping value is used in the subsequent analysis to clearly illustrate the "stop/pass band" properties of the finite periodic ribbed plate and the control mechanism of vibration confinement in disordered and irregular ribbed plates.

A. Vibration characteristics of finite periodic ribbed plates

Vibration characteristics of the finite periodic ribbed plate are studied by the kinetic energy distribution of periodic sections of the ribbed plate as shown in Fig. 3. For clarity, only the kinetic energy distribution of three periodic sections (the source, the fifth and the last sections) is shown

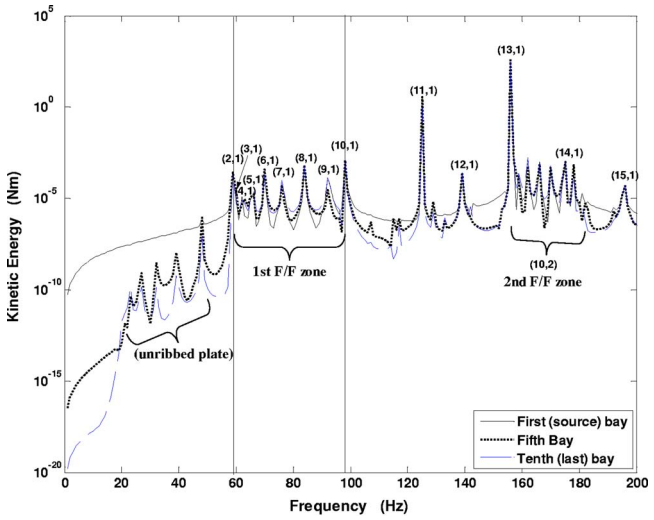


FIG. 3. (Color online) Kinetic energy distribution of three periodic bays (the source, the fifth and the last) of the finite periodic ribbed plate.

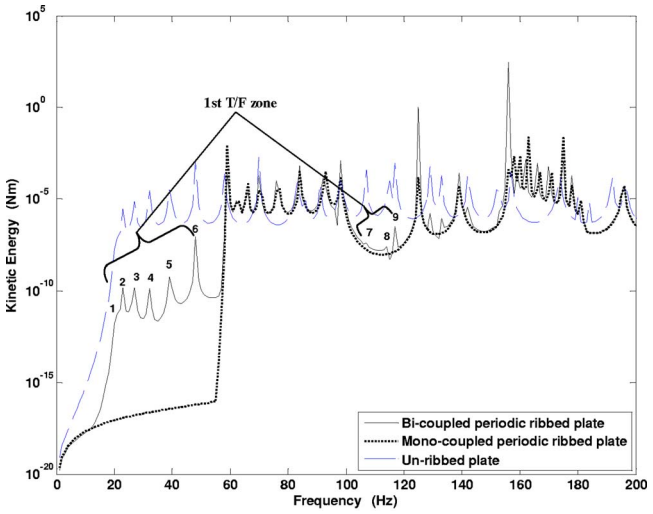


FIG. 4. (Color online) Kinetic energy distribution of the last plate section of the bi-coupled, the mono-coupled periodic ribbed plates and the corresponding section of the un-ribbed plate.

in Fig. 3. Modal indices of the major peak responses are also shown in the figure to assist the understanding of vibration characteristics of the ribbed plate.

As expected, there are two sets of wave propagation zones of the periodic ribbed plate that correspond to the two coupling mechanisms (shear force and moment couplings) considered in the analysis. This is typical for bi-coupled periodic systems.[1,3,11] The set associated with shear force couplings at the interfaces is termed as **F**/**F** (flexural/flexural couplings) wave propagation zones. The other set (associated with moment couplings) is termed as **F**/**T** (flexural/torisonal couplings) wave propagation zones. It is found that peaks in each **F**/**F** wave propagation zone are characterized by one of the beam flexural modes, i.e., $n=1,2,\ldots,\infty$. For instance, the zone enclosed by the two vertical lines (the first **F**/**F** zone) is governed by the first beam flexural mode ($n=1$) while the propagation zone enclosed by the bracket is governed by the second mode ($n=2$).

The periodic ribbed plate used in this numerical example is an asymmetrical periodic system (asymmetrical mass and stiffness distribution on the boundaries of the two end sections with respect to the center of each section),[3] therefore, there are only nine response peaks enclosed in each **F**/**F** propagation zone. It is found that the mode shape distribution for peaks in **F**/**F** zones can be approximated by modes of equivalent orthotropic plates. For instance, the nine response peaks enclosed in the first **F**/**F** zone have mode shape distributions of the distorted sine waves corresponding to the trace wave numbers $k_{m'}$ ($m'=2,3,\ldots,10$) where m' is the number of distorted half wavelength along the long plate edge (L_x). An approximation method to describe the resonant modes of a multiple-ribbed plate by those of an equivalent orthotropic plate is given by Wah.[24] One particular mode of interest in the first **F**/**F** zone is the peak at the upper bounding frequency of the zone ($k_{m'(m'=10)}$) whose modal frequency overlaps with the frequency of the coincidence condition between the plate trace wavelength λ_x and the span of the periodic section Δx_i ($\lambda_{x(n=1)}=2\Delta x_i$). The plate trace wavelength λ_x is related to the trace wave number $k_x=\sqrt{k_p^2-k_n^2}$ by $\lambda_x=2\pi/k_x$. For this mode, all stiffened beams are located exactly at the nodal locations of the mode. Therefore, effects of beam flexural stiffness on modal vibration of the plate at this frequency are negligible and the modal response of the periodic ribbed plate is identical to that of the corresponding un-ribbed plate.

Besides the nine peak responses enclosed in the zone, there is another peak (also governed by flexural wave couplings) that lies just below the lower bounding frequency of the propagation zone due to the asymmetry of the structure.[3] This peak (at 55 Hz as found in other simulations) is found to be a nonpropagating mode corresponding to the trace wave number $k_{m'(m'=1)}$ where most modal vibration energy is confined in the source section of the ribbed plate. Therefore, its modal response is only a minor response whose vibration amplitude is not much different than those of the forced responses at neighboring frequencies.

All peaks in the second **F**/**F** wave propagation zone assume a distorted mode shape distribution similar to that of the $(m',n)=(10,2)$ orthotropic plate mode except for the mode marked by $(m',n)=(14,1)$ in Fig. 3. A common characteristic for peaks in this pass band is that they are all governed by the $n=2$ beam flexural mode. As a result, mode (14,1) does not belong to modes in the second **F**/**F** zone, its modal frequency appears to overlap with the frequency bandwidth of the pass band.

Modes in **F**/**T** zones can be identified by comparing the response of the bi-coupled periodic ribbed plate to that of the corresponding mono-coupled periodic ribbed plate as shown in Fig. 4. Only shear force couplings are considered for the mono-coupled periodic ribbed plate (which is obtained by letting the submatrices $[B]$ and $[C]$ in Eq. (7) equal zeros and subsequently cancel out from the matrix formulation). Because the modal stiffness for modes in the **F**/**T** zones is dominated by the plate bending stiffness due to the smaller beam torsional stiffness, modes in these zones retain the resonance frequencies of the corresponding modes of the unribbed base plate (note: modes in **F**/**T** zones would lie at or

near the resonance frequencies of the un-ribbed plate[1] if the beams have substantial large torsional stiffness, e.g., *I* or *T* frames). The first few peak responses of the first **F**/**T** zone occur at very low frequencies owing to the same reason (i.e., the frequencies of the first six peak responses marked by 1–6 in Fig. 4 coincide with the resonance frequencies of the first six modes of the un-ribbed base plate). The small beam torsional stiffness also leads to wider frequency bandwidths of **F**/**T** propagation zones.[11] For example, the frequency band of the first **F**/**T** zone spreads across the frequency band of the first **F**/**F** zone (see Fig. 4).

The forming of peak responses in **F**/**T** zones is attributed to the inclusion of moment couplings in the mathematical model of the finite periodic ribbed plate. This results in the nonzero numerator in Eq. (15) when $G_{m,n}$ becomes zero (i.e., resonance of the un-ribbed plate). In contrast, the numerator in Eq. (15) would also become zero when $G_{m,n}$ is zero if only shear force couplings are considered (with the exception where all beams are located exactly at nodal positions of a vibrating mode of the plate such as the peak response at the upper bounding frequency of the first **F**/**F** zone). Therefore, for a mono-coupled periodic ribbed plate considering only shear force couplings, peak responses will not be formed at these frequencies resulting from l'Hospital's rule.

The inclusion of moment couplings in the mathematical model also causes the vibration energy redistribution between peak responses in **F**/**F** zones through the cross coupling terms (submatrix $[C]$) in Eq. (7). Some peak frequencies in **F**/**F** zones are also affected by the cross coupling terms as illustrated in Fig. 4.

The mode shape distribution of peak responses in **F**/**T** zones can be traced back to the modal response of the un-ribbed base plate occurring at approximately the same frequency. However, due to the large flexural impedance mismatch at beam/plate interfaces, only the vibration energy associated with the wave component propagating at the incident angle $\pm\sin^{-1}(k_t/k_p)$ can propagate freely in the periodic structure (k_t and k_p are, respectively, the beam torsional and plate bending wave numbers). Therefore, peak responses for modes in **F**/**T** zones in the periodic case are minor responses (i.e., small peak amplitudes).

There are other major peaks in the ribbed plate response in addition to modes enclosed in the two sets of wave propagation zones. These peaks cannot be grouped into any pass band defined by the classical periodic theory. The peaks are found to have the modal indices (11, 1), (12, 1), (13, 1), (14, 1) and (15, 1) of an equivalent orthotropic plate. Appearance of these higher order orthotropic plate modes in the ribbed plate response discloses that the classical periodic theory alone is inadequate in predicting the rich vibration characteristics of finite periodic ribbed plates.

B. Vibration confinement of disordered and irregular ribbed plates

Vibration confinement of finite disordered and irregular ribbed plates is studied in this section for the two cases: (a) single beam spacing disorder to simulate the disorder of a periodic ribbed plate due to installation errors or manufacturing imperfections, and (b) irregular beam spacing to study vibration confinement of an irregular ribbed plate.

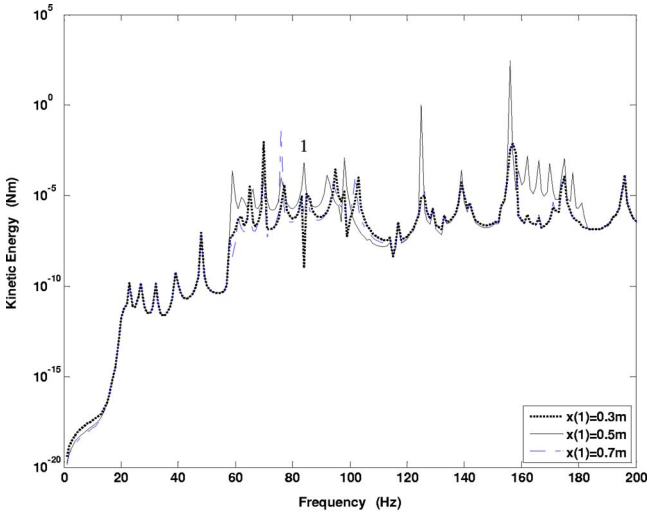


FIG. 5. (Color online) Kinetic energy distribution of the last plate sections of the periodic and the disordered ribbed plates due to the beam spacing disorder of the first beam.

1. Single beam spacing disorder

In this simulation, the location of the first beam of the ribbed plate is shifted from $x_1=0.5$ m to two disordered locations at $x_1=0.3$ m and $x_1=0.7$ m in two separate simulations. The other properties of the ribbed plate remain the same as in the periodic case. Effects of such disorder to vibration energy propagation in the ribbed plate are studied by comparing the kinetic energy distribution of the last plate section in both periodic and disordered cases. These are shown in Fig. 5.

It is found that the disorder mainly affects modes whose modal response is governed by flexural wave couplings. Although the disorder results in general vibration reduction in **F**/**F** wave propagation zones, it also leads to increased peak amplitude for some individual modes in the zones. Such interesting phenomenon is attributed to the effect of beam locations to the modal vibration of the ribbed plate. For example, when the location of the disordered beam is moved to or close to an anti-nodal position of a mode, its flexural stiffness would have greater influence on the modal stiffness, which results in increased modal frequency and decreased modal vibration amplitude. In contrast, the modal stiffness decreases when the location of the disordered beam is shifted to or close to a nodal location of a vibrating mode. The dependency of modal response of a ribbed plate on the beam stiffness has been studied analytically by Lin and Pan[21] and verified experimentally by Nightingale and Bosmans.[25] An example given here to illustrate this effect is the modal response of the chosen peak in Fig. 5 (marked by 1). The mode shape distribution of this mode in the periodic and in the two disordered cases is shown in Fig. 6. It is illustrated that because the location of the disordered beam is shifted from the position close to a nodal position (in the periodic case) to a position close to an anti-nodal position (in the disordered

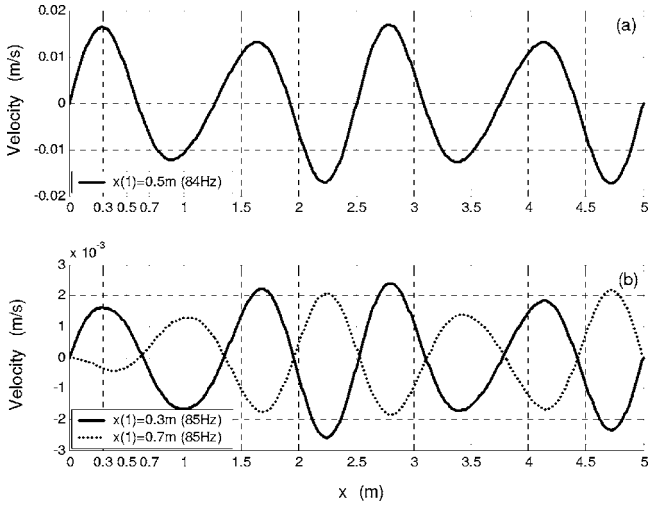


FIG. 6. Mode shape distribution of the $(m',n)=(8,1)$ mode of the ribbed plate calculated at $y=0.3$ m. (a) Periodic ribbed plate, (b) disordered ribbed plates.

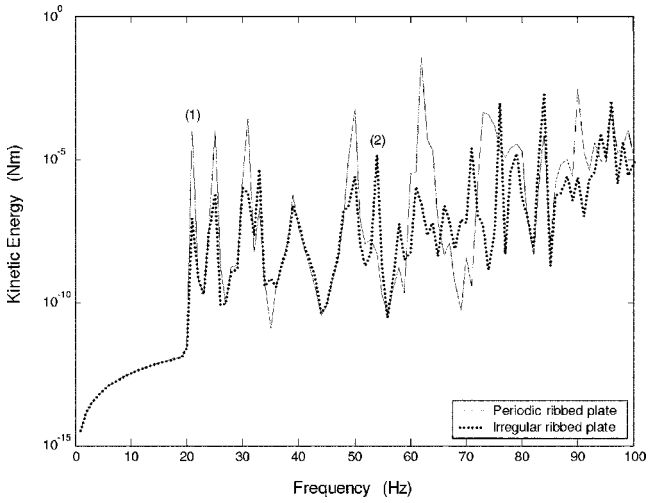


FIG. 7. Kinetic energy distribution of the last plate sections of the irregular and the corresponding periodic ribbed plates.

cases) of the mode, the effect of flexural stiffness of the disordered beam to the modal stiffness increases, which leads to the increased peak frequency from 84 to 85 Hz and the decreased modal vibration amplitude. The frequency bandwidth of the first **F**/**F** zone also increases after the disorder due to the increased modal frequency of the $(m',n)=(10,1)$ mode (note: the disordered beam is no longer located exactly at the nodal position of the mode) leading to higher upper bounding frequency of the zone. In contrast, the disorder has little effect on modal responses of **F**/**T** zones since the governing mechanism (moment couplings) for modes in these zones is not affected by the disorder.

Similar results could also be found if the disorder is originated from perturbations of beam cross sectional areas or mechanical properties. The discussion presented here illuminates the principle of vibration confinement in a ribbed plate.

2. Irregular beam spacing

In order to clarify vibration confinement in a finite irregular ribbed plate, the surface area of the plate is enlarged to $S=40\times 1$ m^2 while the nine stiffened beams are irregularly distributed on the plate surface and have a mean beam spacing of 4 m. The fluctuation of beam locations is restricted to no more than half of the mean beam spacing away from the corresponding periodic locations and to ensure a random phase factor $\exp(jk_p\Delta x_i)$[13,20] in the simulation. Furthermore, the span of the last plate section of the irregularly ribbed plate is deliberately kept to be the same as in the periodic case so that the vibration energy of this section in the periodic and irregular cases is comparable. Several ribbed plate configurations with different beam spacing irregularities are simulated. Nevertheless, only one simulation result is presented here due to close similarity found in others. The exact beam locations of the irregular ribbed plate in the simulation are displayed in the subsequent mode shape plots.

Figure 7 shows the kinetic energy distribution of the last plate section of the irregular and the corresponding periodic ribbed plates. In contrast to vibration localization exposed by other researchers,[16,17,20] it is found that the rib spacing irregularity only results in vibration confinement for some modes of the finite ribbed plate. The modes confined by the irregularity are found to be modes in **F**/**F** wave propagation zones and the un-confined modes are modes in **F**/**T** wave propagation zones. The mechanism of this interesting phenomenon is the same as in the previous case where only modes governed by the flexural wave coupling are affected by spatial irregularities but not modes governed by the moment coupling.

The mode shape distribution of two chosen modes of the ribbed plate is given here to illustrate this phenomenon in more detail. The first is the modal response of a localized mode [mode (1) in Fig. 7] and the other is the modal response of an un-confined mode [mode (2) in Fig. 7]. The mode shape distributions of these two modes in the periodic and irregular cases are shown in Figs. 8 and 9, respectively. It is shown that modal vibration of mode (1) is largely confined in the source section after the irregularity where the vibration energy decays exponentially as waves propagate away from the source section (a typical phenomenon of Anderson localization). The slight increase of vibration amplitude in the last two plate sections in the irregular case is attributed to the effect of boundary wave reflection resulting from the finite plate dimension where a small portion of modal vibration energy can still reach the end boundary of the plate.

In contrast, the modal response of mode (2) is not confined by the irregularity [see Fig. 9(b)]. This is because for modes in **F**/**T** zones, vibration energy of the wave component propagating at the incident angle $\pm\sin^{-1}(k_t/k_p)$ can propagate to a long distance via moment couplings [see Fig. 9(a)], which is not affected by beam spatial irregularities. Furthermore, due to local resonance[13] arising from the spatial matching between the plate bending trace wavelength λ_x and the span of a plate section in the irregular case, higher modal vibration energy can install in the resonance plate sec-

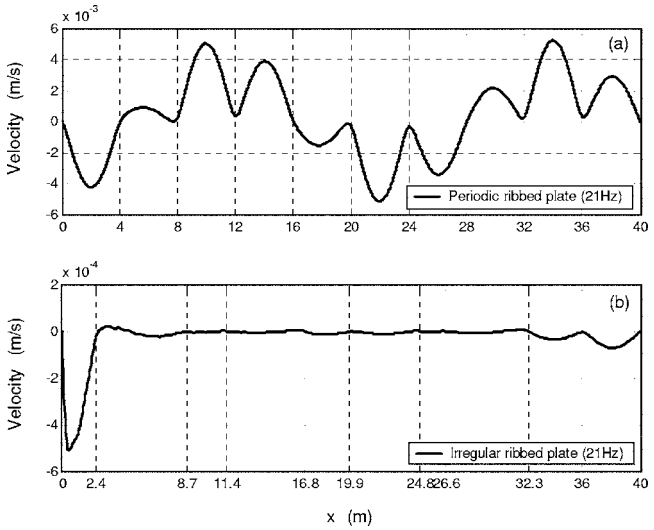


FIG. 8. Mode shape distribution of the periodic and irregular ribbed plates at 21 Hz calculated at $y=0.3$ m. (a) Periodic ribbed plate, (b) irregular ribbed plate.

tion and propagates to its neighboring sections. Therefore, plate sections away from the source section but close to the resonance section can have higher modal vibration energy than plate sections close to the source section but away from the resonance section in the irregular case. This explains the reason why the last plate section in the irregular case displays higher modal vibration energy than that in the periodic case (see Fig. 7).

Notably, although modes in **F**/**T** zones cannot be confined by the beam spatial irregularity, they can be effectively attenuated by applying traditional damping treatments to the structure due to the relatively short beam torsional wavelength. This is coupled with the nonzero denominator $G_{m,n}$ in Eq. (15) at resonance frequencies of the un-ribbed base plate when damping is included.

IV. CONCLUSIONS

An analytical solution is presented in this paper to predict the vibration response of finite periodic and irregular ribbed plates by employing the well-known modal expansion solution. The results are used to examine the control mechanism and modal vibration properties of a finite periodic ribbed plate as well as vibration confinement in disordered and irregular ribbed plates.

The peak responses of a finite periodic ribbed plate were classified into two groups of wave propagation zones according to the two coupling mechanisms at beam/plate interfaces. The group of zones governed by shear force couplings was termed as **F**/**F** propagation zones, and the group whose modal response is controlled by moment couplings was termed as **F**/**T** propagation zones. Modes in each **F**/**F** zone are characterized by one of the beam flexural modes. For instance, modes in the first **F**/**F** zone of the periodic ribbed plate are governed by the fundamental beam flexural mode $(n=1)$ whose modal vibration was represented approximately by the corresponding modes of an equivalent orthotropic plate. Modes in the second **F**/**F** zone all have the mode shape distribution similar to the $(m',n)=(10,2)$ orthotropic plate mode. Modes in **F**/**T** propagation zones retain the resonance frequencies of the corresponding modes of the unribbed base plate attributing to the small beam torsional stiffness, which also leads to wider frequency bandwidth of **F**/**T** zones.

Besides modes enclosed in the two sets of propagation zones, higher order orthotropic plate modes were also identified in the ribbed plate response. These modes could not be grouped into any pass band defined by the classical periodic theory. Appearance of these higher order orthotropic plate modes in the ribbed plate response indicates the limitation of the classical periodic theory in predicting the rich vibration characteristics of finite periodic ribbed plates.

It was found that general vibration reduction can be achieved in the major pass bands (the set of bands controlled by flexural wave couplings) by imposing a single disorder to a finite periodic ribbed plate. Irregular beam spacing can only confine the group of modes associated with flexural wave couplings but not the group of modes associated with moment couplings.

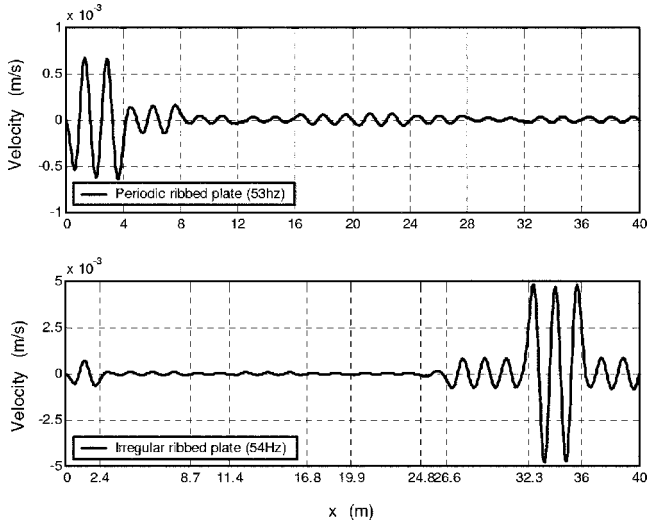


FIG. 9. Mode shape distribution of the periodic and irregular ribbed plates at 53 Hz calculated at $y=0.3$ m. (a) Periodic ribbed plate, (b) irregular ribbed plate.

ACKNOWLEDGMENTS

The author wishes to thank Jie Pan of the University of Western Australia for useful advice in the early part of this work. Financial support from the Australian Research Council and Strategic Marine Pty. Ltd., Western Australia, in the early part of this work is also gratefully acknowledged. The author thanks the reviewer of the manuscript and thanks the associate editor Earl G. Williams for the helpful comments.

[1]M. Heckl, "Wave propagation on beam-plate systems," J. Acoust. Soc. Am. **33**, 640–651 (1961).
[2]D. J. Mead, "A general theory of harmonic wave propagation in linear periodic system with multiple coupling," J. Sound Vib. **27**, 235–260 (1973).
[3]D. J. Mead, "Wave propagation and natural modes in periodic systems: II. Multi-coupled systems, with and without damping," J. Sound Vib. **40**, 19–39 (1975).
[4]A. K. Roy and R. Plunkett, "Wave attenuation in periodic structures," J. Sound Vib. **104**, 395–410 (1986).
[5]Y. Yong and Y. K. Lin, "Propagation of decaying waves in periodic and piecewise periodic structures of finite length," J. Sound Vib. **129**, 99–118

(1989).
[6]L. Brillouin, *Wave Propagation in Periodic Structures* (Dover, New York, 1953), Chap. 7, pp. 140, 171.
[7]L. Meirovitch and R. C. Engels, "Response of periodic structures by the Z-transform method," AIAA J. **15**, 167–174 (1977).
[8]R. S. Langley, "A variation principle for periodic structures," J. Sound Vib. **135**, 135–142 (1989).
[9]V. N. Romanov, "Sound radiation from an infinite plate reinforced with a finite set of beams and driven by a point force," Sov. Phys. Acoust. **23**, 63–68 (1977).
[10]G. Maidanik and J. Dickey, "Velocity distributions on unloaded finitely and regularly ribbed membranes," J. Sound Vib. **149**, 43–70 (1991).
[11]D. J. Mead, "A new method of analyzing wave propagation in periodic structures: Applications to periodic Timoshenko beams and stiffened plates," J. Sound Vib. **104**, 9–27 (1986).
[12]D. J. Mead and A. S. Bansal, "Mono-coupled periodic systems with a single disorder: Free wave propagation," J. Sound Vib. **61**, 481–496 (1978).
[13]C. H. Hodges, "Confinement of vibration by structural irregularity," J. Sound Vib. **82**, 411–424 (1982).
[14]P. W. Anderson, "Absence of diffusion in certain random lattices," Phys. Rev. **109**, 1492–1505 (1958).
[15]C. H. Hodges and J. Woodhouse, "Vibration isolation from irregularity in a nearly periodic structure: Theory and measurements," J. Acoust. Soc. Am. **74**, 894–905 (1983).
[16]D. M. Photiadis, "Anderson localization of one-dimensional wave propagation on a fluid-loaded plate," J. Acoust. Soc. Am. **91**, 771–780 (1992).
[17]D. M. Photiadis, "Localization of helical flexural waves by irregularity," J. Acoust. Soc. Am. **96**, 2291–2301 (1994).
[18]D. M. Photiadis and B. H. Houston, "Anderson localization of vibration on a framed cylindrical shell," J. Acoust. Soc. Am. **106**, 1377–1391 (1999).
[19]E. G. Williams, H. D. Dardy, and K. B. Washburn, "Generalized nearfield acoustic holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).
[20]M. B. Sobnack and D. G. Crighton, "Anderson localization effects in the transmission of energy down an irregularly ribbed fluid-loaded structure," Proc. R. Soc. London, Ser. A **444**, 185–200 (1994).
[21]T. R. Lin and J. Pan, "A closed form solution for the vibration response of finite ribbed plates," J. Acoust. Soc. Am. **119**, 917–925 (2006).
[22]T. R. Lin, "Vibration of finite coupled structures, with applications to ship structures," Ph.D. thesis, the University of Western Australia, 2006.
[23]M. C. Junger and D. Feit, *Sound, Structures and Their Interaction* (MIT Press, Cambridge, MA, 1972), Chap. 6, pp. 143–148.
[24]T. Wah, "Vibration of stiffened plates," Aeronaut. Q. **15**, 185–198 (1964).
[25]T. R. T. Nightingale and I. Bosmans, "On the drive-point mobility of a periodic rib-stiffened plate," *Proc. Inter-Noise 2006*, Hawaii, Dec. 2006.

Broadband acoustic scattering measurements of underwater unexploded ordnance (UXO)

J. A. Bucaro,[a] B. H. Houston, M. Saniga, and L. R. Dragonette
Naval Research Laboratory, Washington, D.C. 20375, USA

T. Yoder, S. Dey, and L. Kraus
SFA, Inc. Crofton, Maryland 21114, USA

L. Carin
Duke University, Durham, North Carolina 27708, USA

(Received 8 May 2007; accepted 9 November 2007)

In order to evaluate the potential for detection and identification of underwater unexploded ordnance (UXO) by exploiting their structural acoustic response, we carried out broadband monostatic scattering measurements over a full 360° on UXO's (two mortar rounds, an artillery shell, and a rocket warhead) and false targets (a cinder block and a large rock). The measurement band, 1–140 kHz, includes a low frequency structural acoustics region in which the wavelengths are comparable to or larger than the target characteristic dimensions. In general, there are aspects that provide relatively high target strength levels (~−10 to −15 dB), and from our experience the targets should be detectable in this structural acoustics band in most acoustic environments. The rigid body scattering was also calculated for one UXO in order to highlight the measured scattering features involving elastic responses. The broadband scattering data should be able to support feature-based separation of UXO versus false targets and identification of various classes of UXO as well.
 [DOI: 10.1121/1.2821794]

PACS number(s): 43.40.Fz, 43.20.Fn, 43.60.Lq, 43.30.Xm [DF] Pages: 738–746

I. INTRODUCTION

Many active and former military installations have ordnance ranges and training areas with adjacent water environments in which unexploded ordnance (UXO) now exists due to wartime activities, dumping, and accidents. These contaminated areas include coastal and inland waters both in the United States and abroad. Over time, such geographic areas are becoming less and less remote as the adjacent lands become further developed, and the potential hazard to the public from encounters with such UXO has begun to rise. Presently there exists no sufficiently effective capability to survey such underwater areas to detect, identify, and map UXO locations.

We have been exploring the potential for developing a structural acoustics based sonar methodology[1] for wide area search and identification of underwater UXO. This new structural acoustics approach may have significant advantages over more conventional acoustic methods which rely on the formation of high resolution images. These advantages include: diverse set of "fingerprints" leading to low false alarm rates; longer range operation leading to wide area coverage; and low frequency sediment penetration leading to buried target prosecution.[2]

Conventional sonar approaches (see Fig. 1) which form images must operate at relatively high frequencies since the limiting image resolution is directly proportional to the acoustic wavelength. In this regime, acoustic wavelengths λ are short compared to the target dimensions and the waves are scattered for the most part from the external boundary of the target (geometric scattering). In contrast, in the structural acoustic regime acoustic wavelengths are comparable to, or longer than, the target dimensions and energy can readily penetrate the target. The acoustic scattering is now also related to the vibrational dynamics and elastic wave bearing properties of the object, both of its outer and internal structure. The time-frequency features[3,4] in the scattered echoes can then be used to "fingerprint" the target without the need to form an image.

The first phase of the current effort is focused on an examination of the scattering exhibited by typical UXO targets using the Naval Research Laboratory's (NRL) state-of-the-art underwater scattering facilities, both laboratory-based[5] and at-sea.[6] Not surprisingly, to date very little is known about acoustic scattering from such targets especially in the structural acoustic regime. Although UXO can be expected to come in many sizes, the typical UXO might be relatively small such as an 80 mm mortar round or 5 in. rocket shell; and considering that such objects tend to be tapered as well—often over a large fraction of their length, scattering highlights might be expected to be relatively low. In addition, the example targets mentioned above and many other UXO's as well have thick metal casings which might have a significant impact on elastic scattering associated with both coupling in and out of the interior as well as acoustically excited global resonance mechanisms. The scattering measurements we report here, to our knowledge the first broadband scattering measurements on such targets, were

[a]Author to whom correspondence should be addressed. Electronic mail: bucaro@pa.nrl.navy.mil.

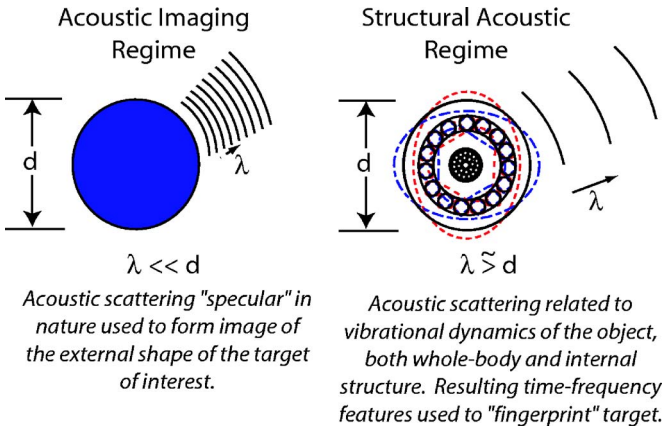


FIG. 1. (Color online) Structural Acoustic ID versus Imaging.

carried out to answer the following questions:

(a) In the structural acoustic domain, what target strength levels are associated with scattering from these relatively small, typically tapered, thick-walled bodies of revolution? Are the levels consistent with detection at reasonably long range (~ tens of meters)?
(b) Are there sufficient levels of scattering associated with *elastic* mechanisms which we know from experience can provide a rich set of features spread over frequency-angle space, a characteristic which facilitates identification? Over what band of frequencies are these elastic mechanisms observed?
(c) Is the scattering from the UXO targets *as a class* sufficiently different from that associated with typical false targets?

We have recently completed the first phase of this data collection using the laboratory based facility, and these results are reported and discussed here. The experimental details are presented in Sec. II including a description of the targets chosen for this study, the source and receiver devices, the measurement configuration, and the measured broadband multi-aspect monostatic target strength. The results are discussed in Sec. III.

II. EXPERIMENTAL DETAILS

An initial, representative UXO target set was identified, and the specific targets obtained from the Aberdeen Proving Grounds. This four target set included a 155 mm artillery shell, a 5 in. rocket warhead, an 80 mm mortar round, and a 120 mm mortar round. Each target was then filled with a two-component epoxy material whose density, bulk modulus, and Poisson ratio are 1500 kg/m^3, 8.1×10^9 Pa, and 0.34 respectively.

The acoustic scattering measurements were carried out at the Laboratory for Structural Acoustics (LSA)[5] at NRL (see Fig. 2) which is a state-of-the-art underwater acoustic research laboratory. The LSA infrastructure includes a large cylindrical one million gallon (17 m diameter × 15 m deep) de-ionized water tank which is vibration isolated, temperature controlled, and heavily instrumented with in-water precision robotically controlled receivers that support background clutter removal for compact range measurements and high spatial density sampling for near-field acoustic holography.

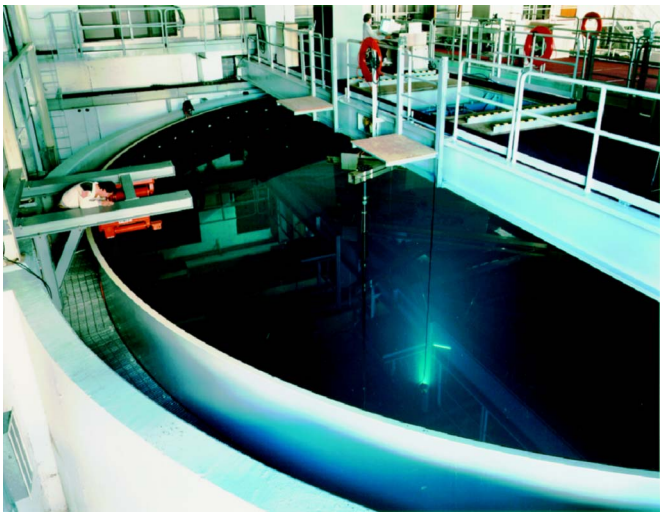


FIG. 2. (Color online) NRL Laboratory for Structural Acoustics. One million gallon pool facility with adiabatic walls and acoustic coatings, vibration isolation system, and complex acoustic scanners, sources, and processing algorithms.

The measurements reported here were conducted with the facility in its compact scattering range mode as shown in Fig. 3. Each UXO target was suspended at middepth in the tank together with the source and receiver. Two sets of sources and receivers were used for these experiments which covered an overall bandwidth of 1–140 kHz. The first source is a 3-m-long near-field line array mounted horizontally. The transducer elements are shaded in such a way as to produce a plane-wave sound field in the near field of the array throughout a limited volume centered at the target position. The line source generates a pseudo-unipolar broadband pulse approximately 20 μs in duration and covers a band of frequencies, f, from 1–25 kHz. The receiver used with this source is a short, vertical line array that was also suspended at the middepth of the tank. A second piston-like source (diameter 0.27 m) which generates a broadband pulse ~5 μs long was used to collect data in the band from 8–140 kHz. A small, standard hydrophone was used as a receiver with this source. The measurement system is designed for collection of both monostatic and bistatic scattering data. However, for the measurements reported here, only the monostatic configuration was used, i.e., the source and receiver fall along the same bisector to the target center. The scattered echo response was measured 2.7 m from the target as a function of aspect angle, θ, in 1° increments over 360°. The data were processed to recover full complex scattering cross sections expressible as target strength referenced to 1 m in the following way. In order to obtain the target strength, three quantities are measured: the incident acoustic pressure, the pool clutter (background) pressure, and the scattered pressure. First, the source is excited and the incident pressure measured at the location corresponding to the target center for the scattering measurement. Second, the source is excited and the clutter pressure field measured at the location at

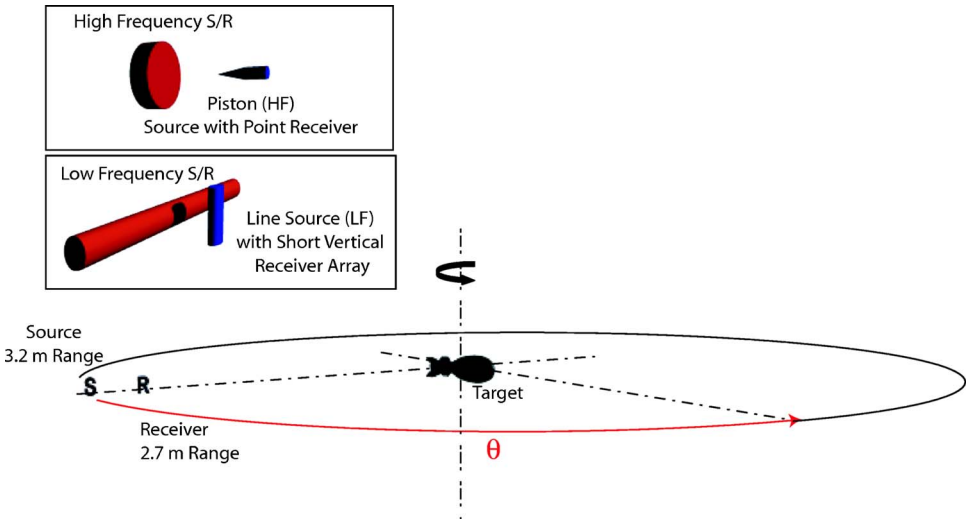


FIG. 3. (Color online) Experimental Measurement Geometry. The target is placed in the plane-wave region of (1) a near-field cylindrical source at low frequency (LF) or (2) a far-field piston source at high frequency (HF) with a nearly co-located broadband short vertical receiver array with the former and a small point-like hydrophone receiver with the latter. The target (which is ~2.7 m from the receiver) is rotated over a full 360° in increments of 1°.

which the receiver will be placed for the scattering experiment. Last, the target is inserted and the scattered pressure field measured.

The target strength is obtained by first subtracting the clutter measurement from the scattering measurement. This process removes energy from any indirect paths due to reflections from the finite - sized pool boundaries or submerged equipment. This step is only possible through precise control of the locations of the acoustic elements and only if fluctuations in the acoustic medium are sufficiently small. For our facility, robotic control of the source and receiver position is approximately 30 μm, and the iso-velocity water is maintained to within 0.01° C for more than a 24 h period. With the clutter removed from the scattered signal, the parameter $X(f,\theta)$ is formed in terms of the scattered signal, $P_{\text{scat}}(f,\theta)$, and the incident field measured at the target center, $P_{\text{inc}}(f)$:

$$X(f,\theta) = \frac{P_{\text{scat}}(f,\theta)}{P_{\text{inc}}(f)} \frac{r_{\text{scat}}}{e^{ikr_{\text{scat}}}}, \tag{1}$$

where r_{scat} is the distance from the target center to the receiver. The target strength (TS) is then defined and displayed as $10\log_{10}(|X(f,\theta)|^2)$.

The measured data over the composite band from 1 to 140 kHz are displayed in Figs. 4–7 as a function of frequency and target aspect for the four UXO targets with the color scale mapping target strength levels in dB. Shown in Figs. 8 and 9 are the measurements made on the two representative false targets, the large rock and the cinder block.

III. DISCUSSION

A. Scattering levels

We address first the question regarding the target strength levels associated with scattering from the UXO targets and whether the levels could support detection at reasonably long range (~ tens of meters).

As can be seen, the highest target strength level observed for each UXO target ranges from about −15 dB for the smallest target (80 mm mortar round) to >−10 dB for the largest (155 mm artillery shell). Taken as a set, at any particular aspect we can readily observe scattering features at some frequencies in the structural acoustics band with levels >−20 dB. Based on our experience and as we establish below, such levels should *generally* provide sufficient signal to noise for detection out to modest ranges.

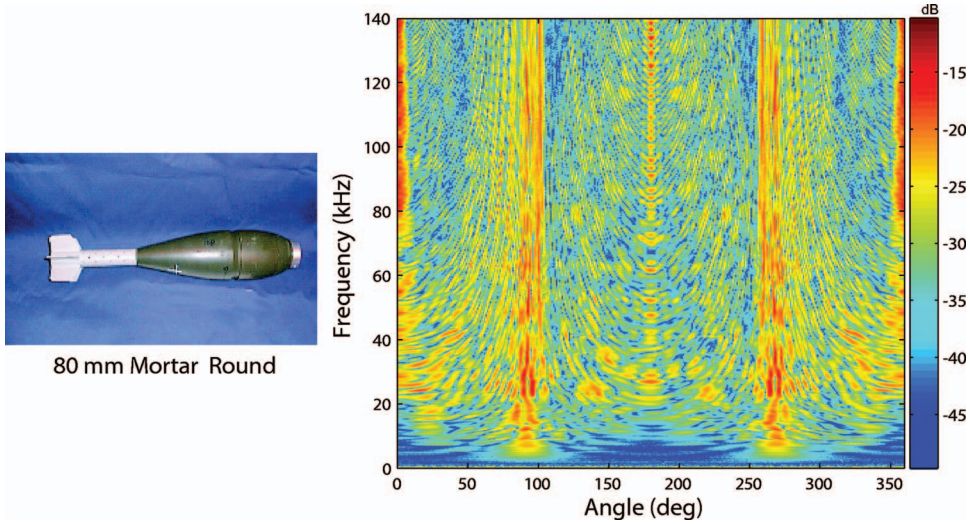


FIG. 4. Measured Target Strength & Target Photo. Magnitude of the target strength coded in color map versus frequency and target aspect for the 80 mm mortar round.

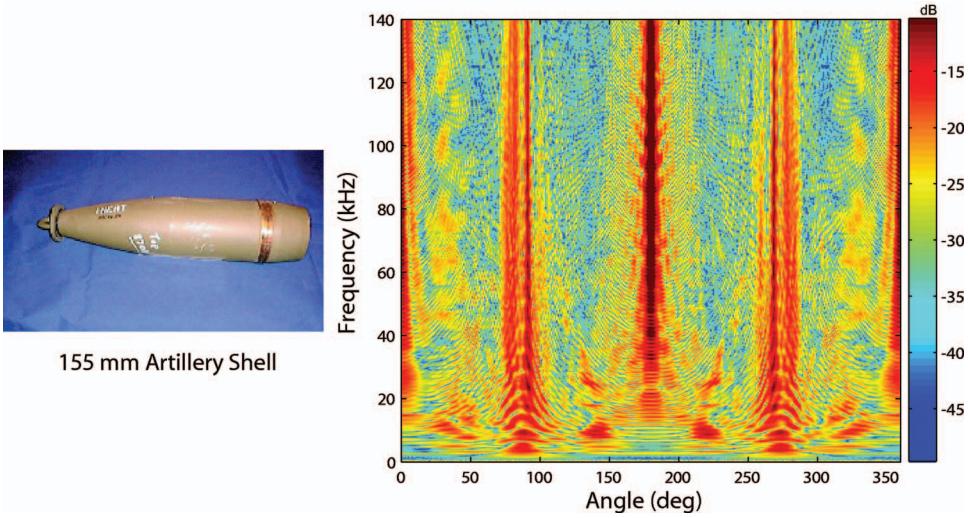


FIG. 5. Measured Target Strength & Target Photo. Magnitude of the target strength coded in color map versus frequency and target aspect for the 155 mm artillery shell.

In Fig. 10 we show the result of a simple estimate based on the sonar equation[7] of the maximum detection range for the following simple case. We assume a source level of 170 dB re: μPa, a medium with no boundaries, two-way propagation loss based only on spherical spreading and a frequency dependent absorption,[8] and a frequency dependent noise level derived from measurements in the San Diego harbor.[9] This frequency dependent, relatively high noise spectral level is about 60 dB re: μPa^2/Hz at 5 kHz falling nonmonotonically to about 30 dB re: μPa^2/Hz at 100 kHz. The predicted range (defined by the range for which the signal/noise=1) is plotted versus frequency for both a 0 dB target and a −20 dB target. As can be seen, for these conditions the −20 dB target features would be detectable at the low end of the band (5 kHz) beyond 100 m ranges and at 20 kHz out to 200 m. Further, −10 dB scattering levels would be detectable beyond even these ranges. We have assumed in the estimate relatively low source levels so that reductions in detection range resulting from boundary effects not considered here should be able to be countered by increases in source level. We conclude that the features present in the structural acoustic band from the UXO should be detectable out to ranges sufficiently long for the intended application viz. acoustic detection and identification of UXO objects in coastal and inland waters.

B. Geometric (rigid) scattering response

The geometric (rigid) scattering[10] is that scattering which arises only from the acoustic impedance mismatch at the interface between the fluid and the target surface and which is caused by reflection which at high frequencies obeys the laws of geometric optics. The geometric response is important for two reasons. First, it provides a "floor" TS level whose angular highlight pattern can be simply inferred from the target's shape. Further, interference with any elastic scattering effects also taking place would produce TS decreases from this floor only over limited bandwidths if at all. Second, knowledge of the geometric response when compared to the actual measurements allows one to determine the degree of participation in the scattering of elastic mechanisms.

The casings of our targets are much stiffer than the surrounding fluid so that we can obtain the geometric scattering by taking the target to be rigid. Accordingly, the rigid-body

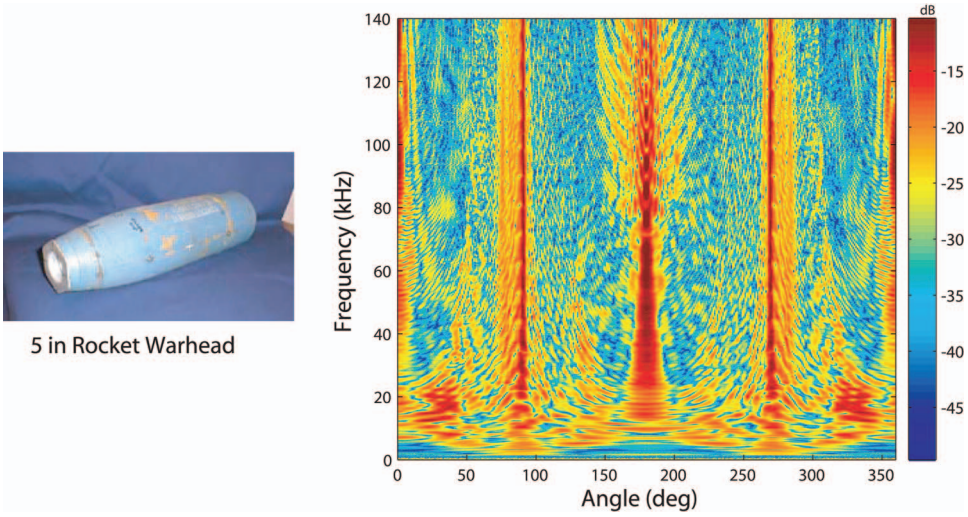


FIG. 6. Measured Target Strength & Target Photo. Magnitude of the target strength coded in color map versus frequency and target aspect for the 5 in. rocket warhead.

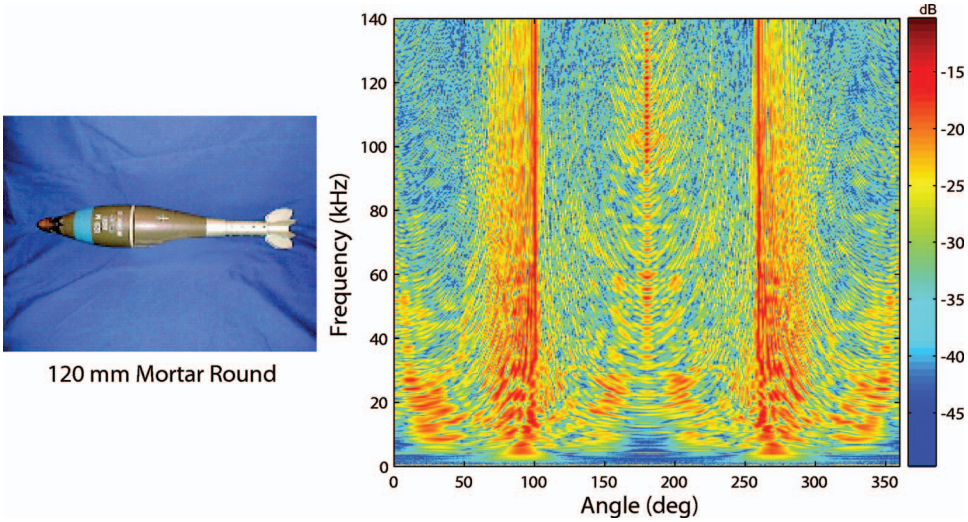


FIG. 7. Measured Target Strength & Target Photo. Magnitude of the target strength coded in color map versus frequency and target aspect for the 120 mm mortar round.

target strength was calculated for the precise shape of one of the targets, the 5 in. rocket, using a highly parallelized finite element-based code[11] over as high a frequency band as was practical (up to 40 kHz). This in turn was determined by the largest grid resolution that could be run on our available parallel computing resource. The results are shown in a semi-log plot in Fig. 11 along with the measured result. In the following, we attempt to put this geometric scattering result into some perspective.

1. Beam aspect TS for the 5 *in. rocket (and the* 155 *mm shell)*

If one approximates the lateral surface of both the 5 in. rocket and the 155 mm shell as a finite circular cylinder, the beam aspect target strengths (TS) of these targets can then be estimated directly from simple expressions. In particular, the far field TS of a finite cylinder can be closely approximated[10,12,13] as

$$TS = 10\log\{a\ell^2/2\lambda\} + 20\log|f_{\infty,}| \tag{2}$$

where "a" is the radius, "ℓ" is the length of the cylinder, "λ" is the wavelength of the incident sound, and "f_∞" is the form function. The first term of Eq. (2) represents the geometric response[10] of the target, which provides the TS floor. At beam aspect, this geometric floor provides most of the back-scattered energy over most of the frequency band. Inserting the length and radius (maximum value) for the 5 in. rocket gives −13.6 dB at 10 kHz, which is close to that computed for the actual shape using the finite element model (see Fig. 6). We also point out that the additional response at aspects just before beam seen in both the finite element and measured responses are the highlights from the tapered section.

2. Stern aspect TS for the 5 *in. rocket (and* 155 *mm shell)*

The stern of the 5 in. rocket (and the 155 mm shell) ends in a circular disk shape and should lend itself to TS estimation by computing the response of a circular plate. For a rigid circular plate, the TS is given by[10]

$$\mathrm{TS} = 20\log(A/\lambda) + 20\log\{(2J_1(\beta)/\beta)\cos(\vartheta)\}, \tag{3a}$$

where "A" is the area of the circular disk, "β" is $\{ka\sin(\vartheta)\}$ where "k" is $2\pi/\lambda$, "a" is the radius of the disk, and "ϑ" is the angle measured from the normal to the disk. At normal incidence the second term in Eq. (3) is 0 dB. If we assume a frequency independent reflection coefficient of 0.7, the re-

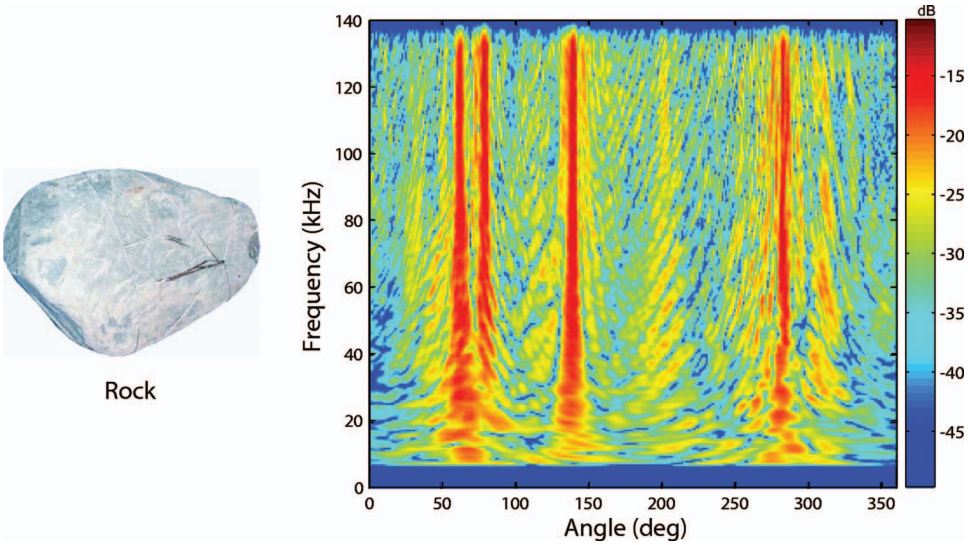


FIG. 8. Measured Target Strength & Target Photo. Magnitude of the target strength coded in color map versus frequency and target aspect for the large rock.

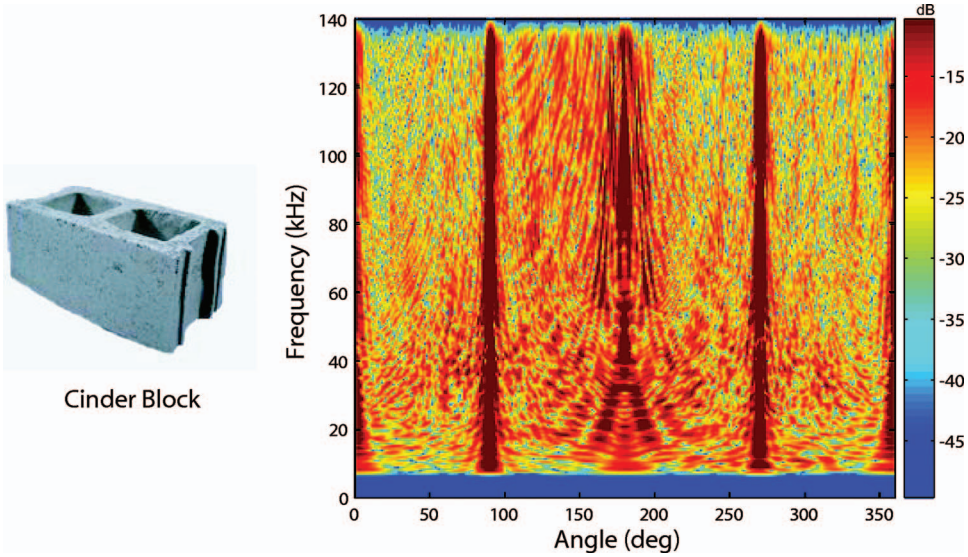


FIG. 9. Measured Target Strength & Target Photo. Magnitude of the target strength coded in color map versus frequency and target aspect for the cinder block.

sponse of the plate at exactly stern incidence would be approximated by

$$\mathrm{TS} = 20\log(A/\lambda) + 20\log(.7). \tag{3b}$$

For the size of the 5 in. rocket stern end, at 10 kHz this would predict a TS of about −21 dB which should then rise at the rate of 6 dB per octave (first term in Eq. (3b)) which is close to what is observed in the finite element calculation (Fig. 11).

3. Conical response prediction

The mortar rounds with their strong lateral curvature, sharp bow taper, and fin-like stern are targets for which no strong beam, stern, or bow responses are expected. There are, however, truncated conical portions of the target which should produce a modest TS near beam aspect. Considering the 120 mm round, the largest of these extends from the vicinity of the indentation (driving band), located near the center of the mortar shell, towards the stern. This conical feature should produce a maximum backscattered response at 10° beyond beam aspect (100°). Observation of the measured scattering response for this target (see Fig. 7) clearly shows this feature as well as the lack of strong responses at bow and stern. As can be seen, the response at 100° is indeed larger than the response at beam aspect (90°) for this target, again indicating that geometric scattering from the cutoff cone portion is a dominant mechanism. The TS of a truncated, circular cone illuminated at normal incidence to its lateral surface is given by[14]

$$\mathrm{TS} = 10\log[2/9\lambda][L_2^{3/2} - L_1^{3/2}]^2[\sin(\alpha)/\cos^4(\alpha)], \tag{4}$$

where L_2 is the length that the full cone would extend if continued to its pointed end, L_1 is the length of the missing portion of the cone, and α is the half angle of the cone. Figure 12 compares the response of the 120 mm mortar round measured at 100° with the computation made from Eq. (4). Above 20 kHz, where the geometric response appears to be the dominant scattering mechanism, the conical approximation is generally within 5 dB of the measured response. Carrying out this analysis for the smaller mortar round produced similar results.

C. Elastic response

We now return to the comparison in Fig. 8 between the measured response versus the computed geometric result for the 5 in. rocket. The comparison clearly shows the impact of elastic effects on, and contributions to, the scattered response in the structural acoustic frequency domain. Notwithstanding these significant elastic effects, the imprint of the geometric response at beam and stern is still visible, although modulated in frequency through the interference of geometric and elastic responses. Even though limited resources prevented us from carrying out the rigid finite-element computations for the other three targets, comparing the general character in the measured responses for all four UXO targets in Figs. 4–7 and considering our discussion of the rigid responses based on the simple TS formulas, we can infer that the other targets have significant contributions from elastic mechanisms as well. In general, the elastic scattering could be related to a

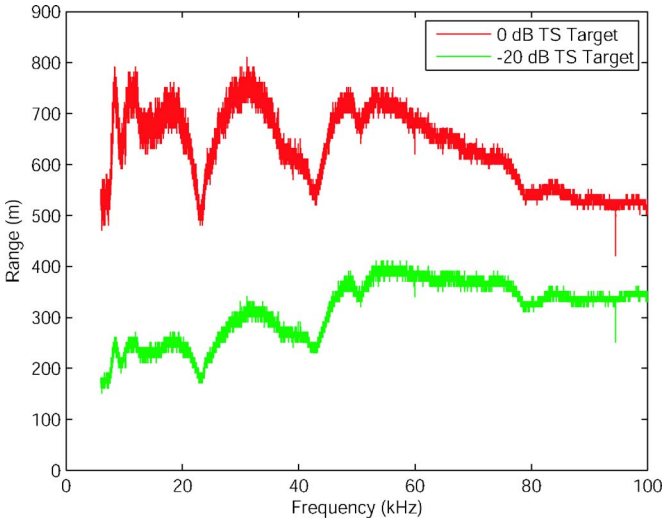


FIG. 10. (Color online) Maximum Detection Range for two target strength levels, 0 and −20 dB, versus frequency.

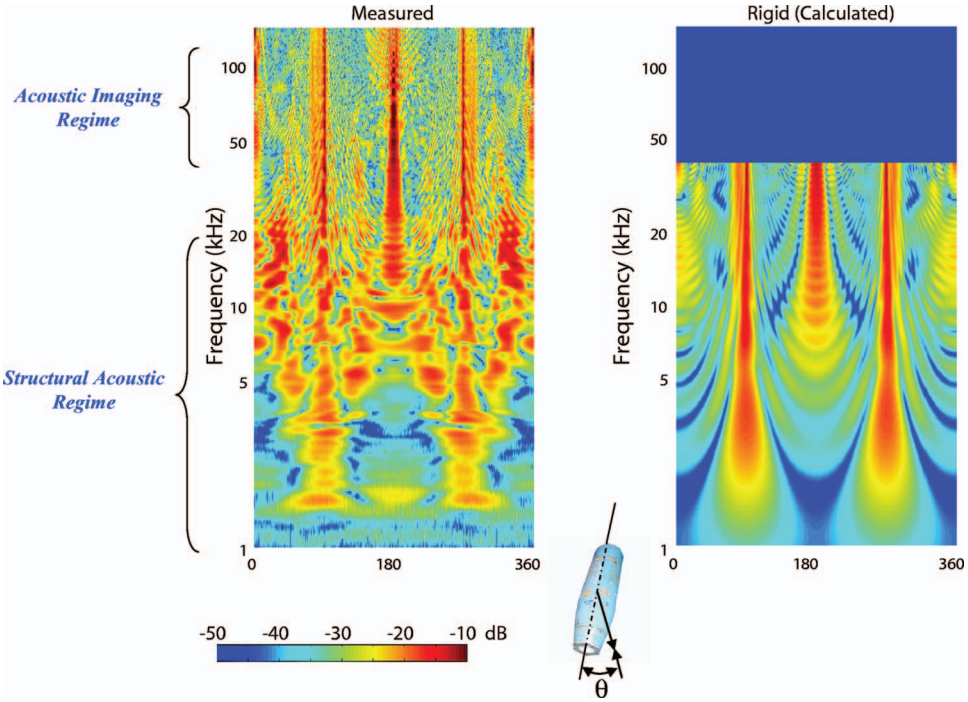


FIG. 11. Measured Target Strength & Calculated Rigid Response. Semilog plot of the magnitude of the target strength coded in color map versus frequency and target aspect for the 5 in. rocket. (Left: measured; right: computed using finite-element-based scattering code.)

number of mechanisms which include shell dynamics, reflection and mode conversion of elastic shell-borne waves, air-cavity responses, penetration and re-radiation from the epoxy material, etc.

For all four UXO targets, we find evidence over broad aspects that at least some of the elastic energy involves scattering from the internal structure. To support this fact, in Fig. 13 we show a standard time/angle plot for the scattered echo from the 5 in. rocket along with an outline based on what we know about the target geometric details properly converted to travel time (using the sound speed in water). On this outline we have labeled the two loci of points on the target profile representing the minimum (A) and maximum (B) acoustic arrival times. At 0° aspect, (A) is associated with the target front and (B) with the target rear. In producing this result, we used a frequency domain filter to suppress the low and high end of the band where the signal to noise was low due to source level roll off. The associated impulse response of the filter was such that the ring-down time for the return to drop more than 30 dB is 0.1 ms. (One can see this ring-down artifact in Fig. 12 where there is some received signal level about 0.1 ms before the minimum acoustic arrival time.) In the figure we can see resonant-like signals originating beyond the A contour and extending well beyond the contour labeled B. These are most likely due to elastic responses from the target interior from some of the mechanisms listed above. These responses are most visible in roughly 30° angular bands centered at 35°, 90°, 135° and the corresponding angles on the other side of the target. These internal scattering effects, whose specific source we have not yet identified, could be related to the dynamics of local as well as more global structural detail.

D. Summary of UXO scattering returns

Based on the observed structural acoustic scattering details and the above discussions, the four UXO targets appear to break into two classes consistent with their rough shapes: Class (1) includes the 5 in. rocket warhead and the 155 mm shell and Class (2) the two mortar rounds. Class (1) has a strong broadband beam (90°) highlight due to the cylindrical-like shape; this feature is scalloped with frequency due to the interference of radiation from circumferential elastic waves with the specular return. There is also a broadband return just below 90° from the shorter tapered side and a strong stern response from the flat end. The latter is also scalloped in frequency and spread in angle due to interference between it and a return from the air cavity/internal structure boundary. Finally, there is a low frequency response region centered roughly around quartering aspects associated with the internal structure.

Class (2) has no strong beam (90°) response; however, there are similar broadband highlights near beam from the

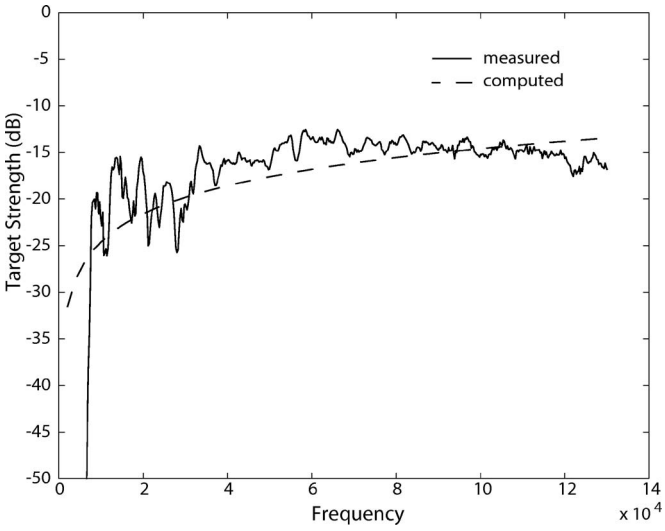


FIG. 12. TS (dB) at 100° versus frequency for the 120 mm mortar round: measured (solid) and as computed (dashed) using Eq. (4).

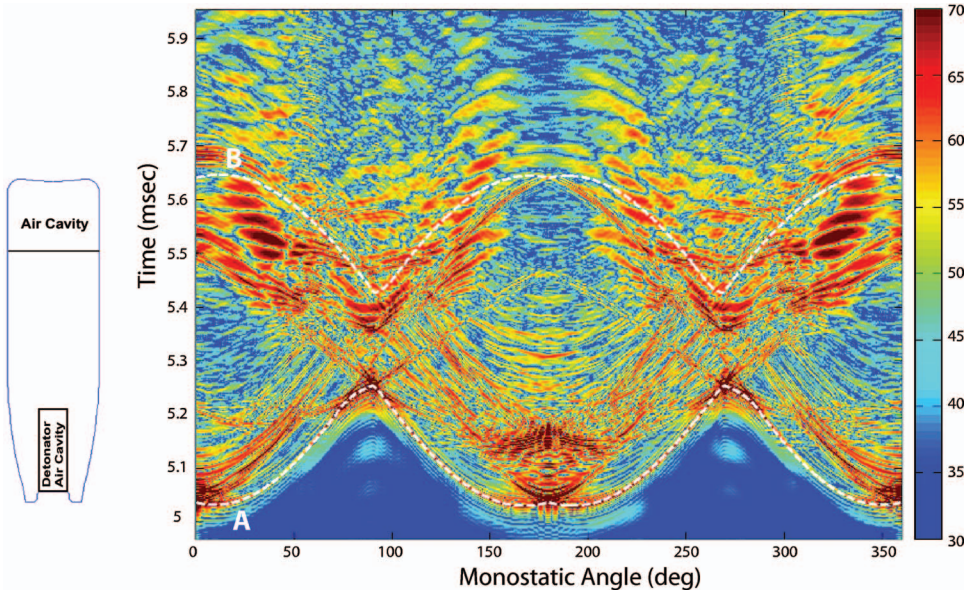


FIG. 13. Time-angle plots for the scattering returns from the 5 in. rocket. The contours labeled A and B represent the minimum and maximum acoustic arrival times, respectively. They are associated with the front (A) and back (B) of the target.

two cone-like sections that make up the round. For the 120 mm mortar, the scattering from the front and stern is low owing to the sharp taper (detonator) and fin-like structure, respectively. The modest return seen at bow in the 80 mm mortar TS plot is actually an artifact due to our use of a flat disk to seal off and waterproof the device which was absent its sharp tapered detonator. As a class, we expect low bow and stern returns in general. As with Class 1, there also exists a modest response region at low frequencies roughly centered around quartering aspects associated with the internal structure.

E. Potential for feature-based target identification

Finally, we address the question as to whether the observed scattering returns would lend themselves to feature-based target identification. Runkle *et al.*[4] considered the scattering from a set of five similarly sized, submerged cylindrical shell targets with identical shapes but differing internal structure. Those authors were able to train discrete hidden Markov models (HMMs) on these measured shell scattering patterns, and the HMMs were subsequently shown to be very effective at identifying the individual shell target when presented with test data from a target randomly selected without knowing which target aspects were included in the data set. Even in cases with additive noise, the HMM algorithm provided robust identification performance despite the high degree of similarity in the scattered patterns among the five shell targets. Further, using these same data bases, Krishnapuram and Carin[15] achieved improved discrimination using a support vector machine classifier. Dasgupta *et al.*[16] demonstrated *class-based* identification using measured data from both the same set of air-filled shells as in Ref. 4 and ellipsoidal shells of different sizes and different material properties for which a finite element-based structural acoustic code was used to generate the scattering data. We find that the *general* character, *gross* features, and echo levels observed in the frequency aspect-dependent UXO scattering patterns presented here are qualitatively similar to the scattering reported from the shells in all the cases referenced above. We argue that it is therefore reasonable to conclude that similar results could be obtained for identification algorithms trained and tested on the UXO scattering returns measured here, although this has not yet been demonstrated.

We also point out that even a cursory comparison of the UXO target data taken as a class to that of the two false targets (the rock and the cinder block) suggests that feature-based separation of the UXO from these types of false targets should be straightforward. Furthermore, an additional, gross feature discriminate might be the symmetry of the scattering patterns. Unlike many false targets, UXOs tend to be symmetric bodies of revolution; and the resulting almost perfect right/left symmetry in the UXO scattering patterns should be directly exploitable for false target separation.

IV. CONCLUDING REMARKS

We have carried out monostatic broadband acoustic scattering measurements on a set of four UXO targets which included an 81 mm mortar round, a 120 mm mortar round, a 155 mm artillery shell, and a 5 in. rocket warhead. To our knowledge, these data sets represent the first of their kind for actual UXO targets.

The next step in the project (and work now under way) involves postprocessing of the measured data bases to extract the structural acoustic features around which the identification algorithms will be designed, trained, and tested.

ACKNOWLEDGMENT

This research was supported wholly by the U.S. Department of Defense, through the Strategic Environmental Research and Development Program (SERDP).

[1]B. H. Houston, J. A. Bucaro, T. Yoder, L. Kraus, J. Tressler, J. Fernandez, T. Montgomery, and T. Howarth, "Broadband low frequency sonar for non-imaging based identification," *Oceans 2002 IEEE/NTS Proceedings*, 383–387 (2002).
[2]H. J. Simpson, B. H. Houston, and R. Lim, "Laboratory measurements of sound scattering from a buried sphere above and below the critical angle (L)," J. Acoust. Soc. Am. **113**(1), 39–42 (2003).
[3]P. Runkle, L. Carin, L. Couchman, T. Yoder, and J. Bucaro, "Multi-aspect

identification of submerged elastic targets via wave-based matching pursuits and hidden Markov models," J. Acoust. Soc. Am. **106**, 605–616 (1999).
[4]P. Runkle, L. Carin, L. Couchman, T. J. Yoder, and J. A. Bucaro, "Multi-aspect target identification with wave-based matching pursuits and continuous hidden Markov models," IEEE Trans. Pattern Anal. Mach. Intell. **21**, 1371–1378 (1999).
[5]B. H. Houston, "Structural acoustic laboratories at NRL in Washington, D.C," J. Acoust. Soc. Am. **92**(4), 2399–2400 (1992).
[6]H. J. Simpson, C. K. Frederickson, E. C. Porse, B. H. Houston, L. A. Kraus, A. R. Berdoz, P. A. Frank, and S. W. Liskey, "Very-low-frequency scattering experiments from proud targets in a littoral environment using a 55 meter rail," J. Acoust. Soc. Am. **114**(4), 2313 (2003).
[7]R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw–Hill, New York, 1983), pp. 17–30.
[8]C. S. Clay and H. Medlin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977), pp. 96–102.
[9]S. Stanic, C. K. Kirkendall, A. B. Tveten, and T. Barock, "Passive swimmer detection," NRL Rev. 97–98 (2004).
[10]R. J. Urick, op. cit., pp. 291–327.
[11]S. Dey and D. K. Datta, "A parrallel *hp*-FEM infrastructure for three-dimensional structural acoustics," Int. J. Numer. Methods Eng. **68**, 583–603 (2006).
[12]R. D. Doolittle and H. Uberall, "Sound scattering by elastic cylindrical shells," J. Acoust. Soc. Am. **39**, 272–275 (1966).
[13]L. Flax and W. Neubauer, "Acoustic reflection from layered elastic absorptive cylinders," J. Acoust. Soc. Am. **61**(2), 307–312 (1977).
[14]"*A Handbook of Sound and Vibration Parameters*," prepared for the Naval Sea Systems Command, by the Systems Technology Department, General Dynamics Electric Boat Division, September 18, 1978.
[15]B. Krishnapuram and L. Carin, "Support vector machines for improved multiaspect target recognition using the Fisher kernal scores of hidden Markov models," *Proceedings of IEEE International Conference on Signal Processing*, **3**, 2989–2992 (2002).
[16]N. Dasgupta, P. Runkle, L. Carin, L. Couchman, T. Yoder, J. Bucaro, and G. J. Dobeck, "Class-based target identification with multiaspect scattering data," IEEE J. Ocean. Eng. **28**, 271–282 (2003).

Characterizing noise and perceived work environment in a neurological intensive care unit

Erica E. Ryherd[a] and Kerstin Persson Waye
Occupational and Environmental Medicine, Department of Public Health and Community, Göteborg University, Box 414, 405 30 Göteborg, Sweden

Linda Ljungkvist
Institute of Health and Care Sciences, Göteborg University, Box 457, 405 30 Göteborg, Sweden

(Received 29 June 2007; accepted 17 November 2007)

The hospital sound environment is complex. Alarms, medical equipment, activities, and ventilation generate noise that may present occupational problems as well as hinder recovery among patients. In this study, sound measurements and occupant evaluations were conducted in a neurological intensive care unit. Staff completed questionnaires regarding psychological and physiological reactions to the sound environment. A-weighted equivalent, minimum, and maximum (L_{Aeq},L_{AFMin},L_{AFMax}) and C-weighted peak (L_{CPeak}) sound pressure levels were measured over five days at patient and staff locations. Acoustical descriptors that may be explored further were investigated, including level distributions, restorative periods, and spectral content. Measurements near the patients showed average L_{Aeq} values of 53–58 dB. The mean length of restorative periods (L_{Aeq} below 50 dB for more than 5 min) was 9 and 13 min for day and night, respectively. Ninety percent of the time, the L_{AFMax} levels exceeded 50 dB and L_{CPeak} exceeded 70 dB. Dosimeters worn by the staff revealed higher noise levels. Personnel perceived the noise as contributing to stress symptoms. Compared to the majority of previous studies, this study provides a more thorough description of intensive care noise and aids in understanding how the sound environment may be disruptive to occupants. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2822661]

PACS number(s): 43.50.Jh, 43.55.Gx [KA] Pages: 747–756

I. INTRODUCTION

The sound environment in hospitals is complex and diverse. Staff activities, medical equipment, alarms, portable carts, communication/paging systems, and high load heating, ventilation, and air-conditioning systems are just a few examples of the types of noise sources often present. Although the noise in hospitals is often a complaint among patients, staff, and visitors,[1] the pool of research on hospital noise remains limited.[2] In particular, the specific acoustical characteristics of the noise and occupant reaction have not been well established—a fundamental step in determining the current hospital "soundscape." The three central aims of this project were: (a) to identify acoustic descriptors relevant for the sound environment in intensive care units, (b) to investigate the staff's perception of the sound environment, and (c) to identify areas for future research.

II. PREVIOUS RESEARCH

The World Health Organization (WHO) published guidelines of recommended hospital noise levels;[3] yet, a recent landmark survey of hospital noise research revealed that not one result indicated compliance with these guidelines.[2] The study also found that hospital noise levels have been rising consistently since 1960, with the average day-time L_{Aeq} levels rising from 57 dB in 1960 to 72 dB, and nighttime L_{Aeq} levels rising from 42 to 60 dB. Based on this survey of previous research, it is apparent that overall noise levels in hospitals are problematic.

The effect of these high levels of noise on patients and staff is a major concern, as hospitals should be conducive to patient recovery and safety as well as employee health and productivity. However, most of the previous work has focused only on the overall levels of noise with minimal examination of the frequency distribution, tonality, time-variance, and other detailed properties of sound. Previous work has shown that these detailed qualities of sound can impact occupant perception, annoyance, and performance.[4–8] A few studies do provide some information about the frequency spectrum of health care noise,[2,9–12] but provide limited or no psychological or physiological evaluations.

It has been shown that noise can generally have both psychological and physiological effects on humans.[13] Previous research on hospital patients have documented negative physiological effects of the acoustics such as a reduction in the recuperative properties of sleep,[14] cardiovascular response,[15,16] increased incidence of rehospitalization,[16] extended hospital stay,[17] and increased dosages of pain medication.[18] Animal testing has also revealed that wound healing may be slowed.[19] Intensive care unit (ICU) patients are also at risk for developing ICU syndrome, a condition which may be partially attributable to environmental conditions. Patients with ICU syndrome may exhibit symptoms of

[a]Currently at: Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 124 Love Bldg., 771 Ferst Dr., Atlanta, GA 30332-0405. Electronic mail: eryherd@hotmail.com

distress, bewilderment, and hallucinations,[20] and report feeling extreme instability, vulnerability, and fear.[21]

The impact of hospital noise on staff members is also important, yet a lower number of studies have been undertaken on this topic. There is some evidence that overall levels of hospital noise may decrease staff mental efficiency and short-term memory[22] and contribute to stress,[23,24] burn-out,[25] and hearing loss.[26] Increased sound absorption has also been correlated with improvement in the staff psychosocial environment[27] and perception of noise.[28] In addition to their own well being, the reaction of staff members to the hospital sound environment could be important for the safety of patients. For example, speech interference and increased medical errors are two potentially hazardous effects of hospital noise.[2,3]

Although the previous work provides a good basis in understanding the hospital sound environment, information remains limited with regard to the reaction of staff members and the specific acoustical characteristics of the sound. This study expands upon the previous work by using a variety of acoustical descriptors to provide a more complete impression of the quality of sound in a neurological ICU environment. Timely new data on the hospital noise soundscape are also provided through an investigation of the response of nursing staff members to ICU noise. Taken together, this information can be used to pursue future targeted noise mitigation research.

III. METHODOLOGY

A. ICU environment

The research was conducted at a neurological ICU at a major research hospital in western Sweden. The neurological intensive care unit is a critical care setting where patients are sedated most of the time and require constant monitoring by the nurses. It may generally be considered a high pressure working environment. Previous intensive and acute care unit research related to noise has been conducted in neonatal units,[29–35] pediatric units,[2,24] coronary or cardiac units,[15,16,27,36] respiratory or medical units,[37,38] general surgical units,[39,40] and unspecified units,[41–43] but few have examined a dedicated neurological unit.[44]

The ICU examined has several individual rooms containing two patients each. One room was selected for the noise measurements. This room had a similar function, layout, furnishings, medical equipment, typical staff activities, and overall noise levels as compared to the other rooms in the ICU. A view of the floor plan is shown in Fig. 1. As in most modern hospital environments, limited absorption exists in the room. It has linoleum tile flooring, gypsum board walls, and a lay-in acoustical tile ceiling. All rooms in the ICU were occupied during the study period and reverberation time measurements were not allowed.

The staff was split into two categories of lower and higher rank depending on their level of education. During the noise measurement period, two nurses were always on staff in the measurement room, with one being lower and one higher rank. One to two patients were present during the entire measurement period. The patient in bed one (shown in Fig. 1) was on a respirator and sedated for the first three measurement days. On the fourth day, the sedation was stopped. The first patient in bed two was awake, not sedated, and not on a respirator. This patient was transferred to another ward on the second day. The second patient in bed two was sedated and on a respirator.

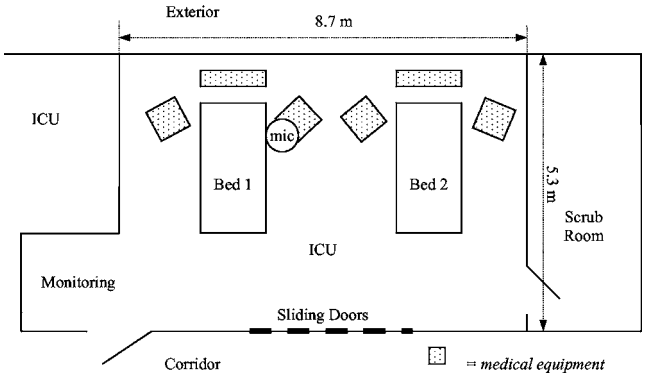


FIG. 1. Floor plan of the ICU room where stationary and dosimeter sound measurements were collected (not to scale). Floor to ceiling height=2.7 m.

B. Noise measurements

Two types of sound pressure level measurements were collected over the course of five days: (a) stationary sound level meter measurements near a patient, and (b) staff dosimeter measurements. Data were collected continuously using a fast response time for maximum and minimum levels (0.125 s)[45] as recommended by WHO.[3] Note that much of the previous research has measured noise based on a slow response time (1 s)[45] which may have resulted in decreased sensitivity to rapid changes in sound level. During all measurements, patients, staff, and visitors continued with their normal activities.

1. Stationary room measurements

Stationary sound level meter measurements were conducted to gain an indication of the ambient noise in the room at the patient location. The goal was to locate the microphone as close to the patient's head as possible. This proved difficult as the microphone could not impede the movement of staff or the moving arms of the medical towers surrounding the patients, so the microphone was cantilevered off of the top of one medical tower. The primary measurements were made with the microphone located approximately 0.5 m from the patient's head at a height of 1.7 m.

Data were collected with a Brüel & Kjær (B&K) type 2260 sound level meter and corresponding analysis was conducted using B&K Evaluator 7820 software. One minute averaging intervals and a range of 40.8–120.8 dB were used. Octave band and third octave band spectra were obtained from a 15 min quiet period during the night to gain a description of the background noise, and over a period of 5 days to obtain an average description of the sound environment in the frequency domain.

TABLE I. Subject questionnaire demographics. NR=no response.

	Job rank	
	Lower—16 total persons	*Higher—31 total persons*
Gender	15 female (1 NR)	25 female, 5 male, (1 NR)
Age	Mean=43.0; median=44.5	Mean=40.4; median=44.5
Avg. length of time working in ICU	12 years (1 NR)	10 years (2 NR)
Full-time/ Part-time status	9 full-time, 7 part-time	21 full-time, 10 part-time

2. Staff dosimeter measurements

Dosimeters were worn by the nursing staff to gain an approximation of their noise exposure. Data were collected with a Larson Davis Spark 705+ dosimeter and corresponding analysis was conducted using BLAZE 5.06 software. Thirty second averaging intervals, a gain of 30 dB, and a range of 40–113 dB were used.

Two nurses at a time wore dosimeters. One dosimeter was always worn by a higher rank nurse and the other by a lower rank nurse. After each nursing shift ended, the dosimeters were passed onto the next set of nurses, thereby allowing continual monitoring of the general exposure levels of the nurses in the ICU environment. Each nurse wearing a dosimeter kept a log documenting their activities for the work shift. They would indicate when they had to leave the ICU room and these time periods were later excluded from the overall measurements. The dosimeter microphone was worn on the midtop of the wearer's shoulder and oriented parallel to the plane of the shoulder in accordance with ANSI S12.19.[46] All nurses were given instructions for positioning and mounting of the microphones to minimize differences between wearers.

C. Staff questionnaire

A structured questionnaire was administered to the staff members just prior to the measurements. Forty-seven ICU nursing staff members completed the questionnaire, corresponding to 68% of the nurses working in the ICU. Additional demographic information about the sample is presented in Table I.

The questionnaire items could generally be split into the following categories: perception and handling of alarms, informing relatives about alarms, other noise sources, perceived risks for personnel and patients, perceived importance of research on the ICU sound environment, and possible interventions. The questions were designed based on the experiences that one of the authors had as an intensive care nurse, and also on previous research in occupational and environmental medicine.[47,48]

The questionnaire consisted of 32 closed-ended and 7 open-ended response questions. In the closed-ended response questions, staff members had to indicate their level of agreement with various statements based on a four-point discrete scale. They were also given the option of checking "I don't know." In the open-ended response questions, staff members described some of their positive and negative experiences with regards to the sound environment. Questionnaires were administered in Swedish.

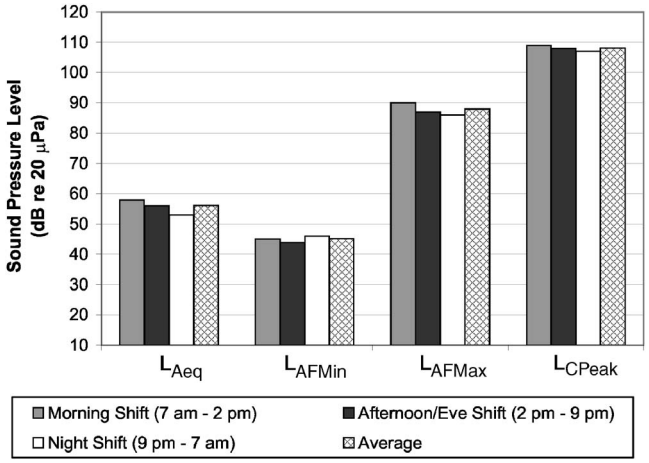


FIG. 2. Measurements of A-weighted equivalent, minimum, and maximum (L_{Aeq},L_{AFMin},L_{AFMax}) and C-weighted peak (L_{CPeak}) sound pressure levels as a function of work shift. Values are from the 1 min stationary room measurements averaged over the five measurement days.

IV. RESULTS

A. Stationary room noise measurement results

1. Overall levels

The overall maximum, peak, and equivalent sound pressure levels measured with the stationary sound level meter at the patient are shown in Fig. 2. Values shown are averaged over the 5 day period of observation. Note that all averaged values presented in this study reflect the logarithmic or energy averages unless otherwise noted. Figure 2 shows that the L_{Aeq} in the room ranges from 53 to 58 dB, depending on time of day. These values are comparable with those measured in other hospital studies over the last 10 years which generally reported L_{Aeq} values ranging from 40 to 80 dB depending on time of day.[2]

Figure 2 indicates slightly higher equivalent, maximum, and peak sound pressure levels during the day-time shifts compared to the night shift; however, these differences are generally minor. The average difference in L_{Aeq} from day to night in this study was 4 dB. These results are in general agreement with two recent ICU studies.[2,40] In a pediatric ICU, the variation in L_{Aeq} as a function of time of day was found to be less than 10 dB,[2] and in a general ICU the variation was on the order of 5 dB.[40] As shown in Fig. 3, the L_{Aeq} also did not significantly change as a function of day of the week.

The data presented in Fig. 2 and 3 indicate that the overall levels are not changing substantially over relatively large periods of time. However, beeping alarms, medical equipment, rolling carts, footfall, and closing doors are just a few of the dynamic and impulsive sounds that create potentially offensive short-term level fluctuations. These short-term fluctuations are apparent from investigation of Fig. 4, which shows 1 min measurement intervals over a portion of the Tuesday night shift.

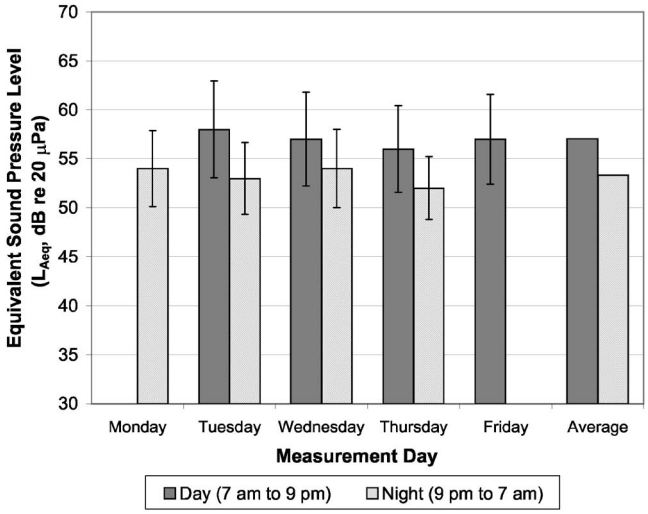


FIG. 3. Day and night A-weighted equivalent sound pressure levels (L_{Aeq}) averaged as a function of measurement day with standard deviation. Results are from the 1 min stationary room measurements.

Additional information about the sound environment was obtained through a statistical distribution of levels. The percent of time the maximum and peak levels exceeded values ranging from 50 to 100 dB was calculated. Results are presented in Fig. 5. From Fig. 5(a), it can be seen that L_{AFMax} exceeds 50 dB most of the time. The interpretation of these results is that the maximum A-weighted sound pressure levels (L_{AFMax}) recorded every 1 min over 5 working days and nights exceeded 50 dB more than 95% of the time. Furthermore, L_{AFMax} exceeds 80 dB 3.1% of the time during the day and 1.3% of the time at night. WHO recommends a maximum L_{AFMax} of 40 dB at night.[3] From Fig. 5(b), it can be seen that L_{CPeak} exceeds 70 dB more than 90% of the time, but reaches levels above 100 dB 2.0% of the time during the day and 0.4% of the time at night.

The behavior of noise over time can also be related to the amount of quieter or "restorative" periods in the background noise. It was noted that there were brief periods of time where the background noise level stabilized to a levels

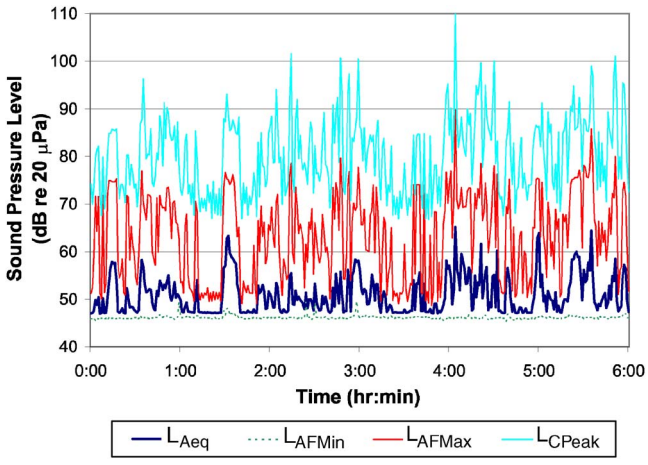


FIG. 4. (Color online) A-weighted equivalent, minimum, and maximum (L_{Aeq}, L_{AFMin}, L_{AFMax}) and C-weighted peak (L_{CPeak}) sound pressure levels measured over 1 min intervals during Tuesday night. Values are from the stationary room measurements.

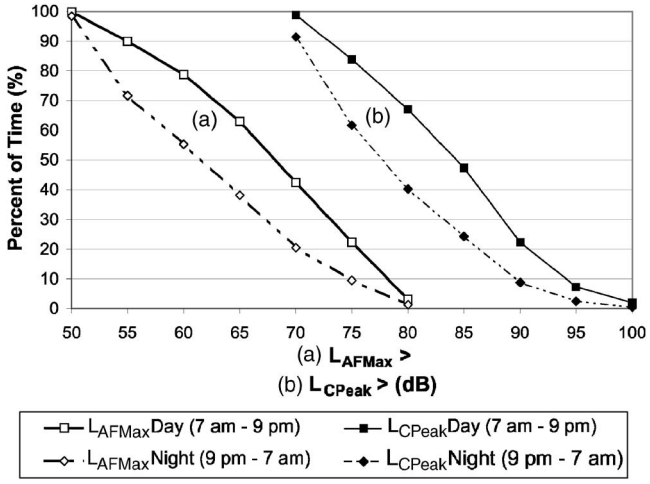


FIG. 5. Statistical level distributions of peak and maximum levels. *Y* axis represents the percent of time that (a) L_{AFMax} and (b) L_{CPeak} exceed values shown on the *x* axis. Values are from the 1 min stationary room measurements averaged over the five measurement days.

of approximately 47 to 48 dB L_{Aeq}, 52 to 53 dB L_{AFMax}, and 71–72 dB L_{CPeak}. These brief periods could be restorative to occupants as the background noise was not fluctuating as much during these times, although the levels are still higher than recommended. It is expected that the noise during these periods was primarily due to the ventilation systems and more steadily operating medical equipment. A subsequent spectral analysis of one of these periods is provided in Sec. IV A 2. Table II provides an analysis of the average occurrence and length of periods of time greater than or equal to 5 min with an L_{Aeq} below 50 dB. Note that there were no 5 min periods of time below 45 dB L_{Aeq} during the entire measurement week. The mean restorative length (L_{Aeq} $<$50 dBA) was 9 and 13 min for the day and night, respectively. The maximum length ranged from 13 to 19 min during the day and from 40 to 50 min at night.

A similar analysis was conducted for L_{AFMax}. Table II shows the average occurrence and length of periods of time greater than or equal to 5 min with an L_{AFMax} below 55 dB. Previous research has shown that sleep disturbance can occur from traffic noise at levels from 45 dB L_{AFMax}.[49] An L_{AFMax} level of 55 dB was chosen in the present analysis, because there were no 5 min periods of time below either 45 or 50 dB L_{AFMax} during the entire measurement week. The mean restorative length (L_{AFMax} $<$ 55 dBA) was 10 and 8 min for the day and night, respectively. The maximum length ranged from 9 to 15 min during the day and from 10 to 26 min at night.

Likewise, Table II shows the average occurrence and length of periods of time greater than or equal to 5 min with an L_{CPeak} below 75 dB. Note that there were no 5 min periods of time below 70 dB L_{CPeak} during the entire measurement week. The mean length (L_{CPeak} $<$ 75 dBA) was 6 and 8 min for the day and night, respectively. The maximum length ranged from 5 to 15 min during the day and from 14 to 17 min at night.

2. Spectral analysis

Also of interest is the spectral content of the background noise. Figure 6 shows results of octave band and one-third

TABLE II. Analysis of restorative periods where the L_{Aeq} was less than 50 dB, $L_{AFMax}<55$ dB, or $L_{CPeak}<75$ dB for 5 min or longer during the day (7 a.m.–9 p.m.) and night (9 p.m.–7 a.m.).

Measurement day and time	Number of periods ⩾5 min			Mean length of periods ⩾5 min (in min)			Maximum length of period ⩾5 min (in min)		
	L_{Aeq} <50	L_{AFMax} <55	L_{CPeak} <75	L_{Aeq} <50	L_{AFMax} <55	L_{CPeak} <75	L_{Aeq} <50	L_{AFMax} <55	L_{Peak} <75
Tuesday day	9	1	3	8.4	13.0	5.7	18	13	7
Wednesday day	10	3	4	8.4	7.3	7.5	19	9	13
Thursday day	19	4	11	9.3	11.3	7.5	15	15	15
Friday day[a]	5	1	2	9.6	10.0	5.0	13	10	5
Monday night	7	10	16	16.3	11.9	7.0	50	26	17
Tuesday night	18	8	13	11.7	7.5	7.8	40	16	16
Wednesday night	20	8	10	10.4	7.8	8.9	42	13	15
Thursday night	21	10	18	13.8	6.4	7.5	44	10	14

Measurement period 7 a.m.–2 p.m.

octave band analyses of the measured linear equivalent sound pressure levels (dB L_{eq}). Two periods of time are shown—the overall average taken over the 5 measurement days, and one of the quieter restorative periods occurring during the night (15 min). The dip at 160 Hz is assumed to reflect a mode in the room. As the main purpose was to measure the response at the patient and not the room average, detailed measurements were carried out at a single position.

Noise criteria ratings were calculated based on the average background noise levels shown in Fig. 6, and are summarized in Table III.[50] All of the criteria systems rate the noise as sounding hissy for both the overall average and the restorative period (note that the noise criteria method does not allow for spectral quality descriptors). This would indicate that high frequency noise dominates the ICU. Note that some debate exists over the appropriateness of these criteria under a variety of spectral conditions.[7,8,50]

B. Dosimeter results

The sound pressure levels measured with the dosimeters are shown in Fig. 7. Values shown are averaged over the 5 day period of observation. It can be seen that the average L_{Aeq} ranges from 65 to 71 dB depending on the time of day. Hence, the dosimeter measured higher equivalent, maximum, and peak sound pressure levels than those made by the stationary sound level meter, shown in Fig. 2.

Slightly higher equivalent sound pressure levels were measured during the day-time shifts compared to the night shift, as shown in Fig. 7. Again, these differences are minor. There was minimal difference in the maximum and peak levels measured by the dosimeters based on shift.

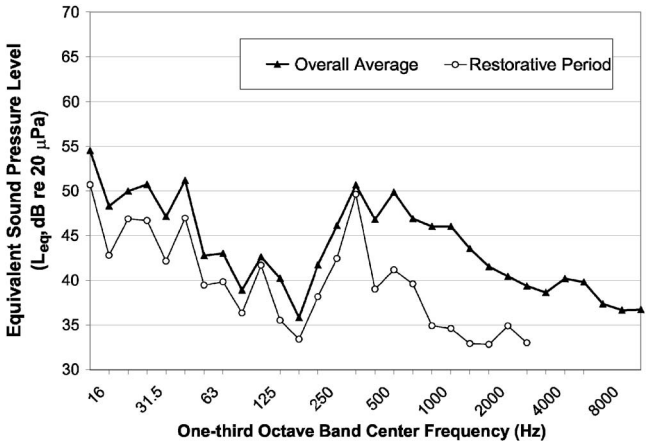


FIG. 6. Measured background noise spectra in one-third octave bands. Values shown are from the 1 min stationary room measurements averaged over the five measurement days (labeled as overall average) and averaged over a 15 min sample quieter restorative period at night (labeled as restorative period).

C. Questionnaire results

Figures 8 and 9 show example results from the questionnaire with regard to the perceived risks for patients and personnel, perception and handling of alarms, and possible interventions. For a full report on the perception evaluations of staff in this study refer to Ljungkvist.[51]

As shown in Fig. 8, most of the nurses surveyed (91%) felt that noise may negatively affect them in their daily work environment. Many nurses reported experiencing irritation (66%), fatigue (66%), concentration problems (43%), and tension headaches due to the sound environment (40%). In addition to their own reactions, the nurses also perceived risks for the patients. Most of the nurses (96%) felt that noise contributed to the development of ICU syndrome in patients. Approximately half of the nurses (49%) said they often discussed the sound environment with their colleagues, but 70% said they did not discuss it with their managers. This is in-

TABLE III. Criteria ratings based on the average noise levels from the stationary measurements.

Criteria method[a]	Overall average ratings (5 days)	Restorative period ratings (15 min)
Noise criteria (NC)	49	41
Balanced noise criteria (NCB)	48 hissy	38 rumbly, hissy
Room criteria (RC)	49 hissy	41 hissy
Room criteria Mark II (RC Mark II)	49 hissy, objectionable	41 hissy, objectionable

Reference 50.

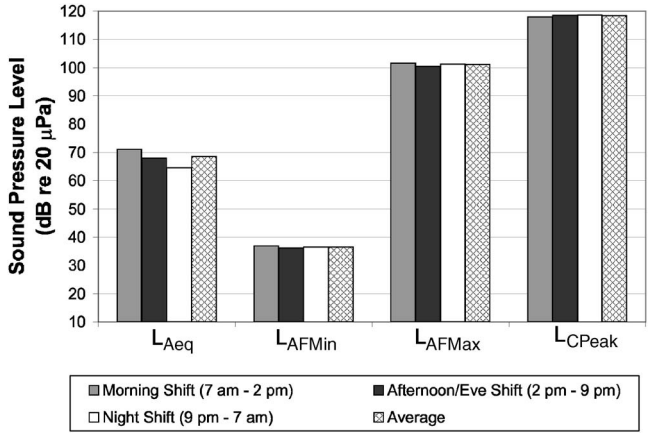


FIG. 7. Measurements of A-weighted equivalent, minimum, and maximum (L_{Aeq}, L_{AFMin}, L_{AFMax}) and C-weighted peak (L_{CPeak}) sound pressure levels as a function of work shift. Values are from the 30 s dosimeter measurements averaged over the five measurement days.

teresting considering that most of the nurses (96%) felt that it was important that sound measurements were being made.

Figure 9 presents the nurses' opinions on the specific questions about alarms. Some of the nurses (38%) thought that an intensive work day with many alarms could affect their sleep. Many (43%) also thought that the alarm noises influenced their ability to perform their job tasks. Almost half of the nurses surveyed (49%) said they sometimes adjusted the alarm levels so that they would not hear them, although 70% conceded that all adjustments to the alarms should be documented on the patient's observation sheet.

V. DISCUSSION

A. Noise measurement results

Noise exposure in the ward measured in this study was analyzed using both stationary measurements at the patient and dosimeters worn by the staff. Dosimeters are most widely used in industrial settings to estimate occupational exposure to high levels of noise which may induce hearing loss. However, dosimeters are useful tools in lower level noise settings such as schools and hospitals as they can be worn directly on the person, thereby feasibly gaining a better approximation of a person's noise exposure than stationary measurements.

In this study, higher levels were measured with the dosimeters as compared to the stationary sound level meters. The average difference in L_{Aeq} was found to be between 12 and 13 dB depending on the time of day. In post-study lab measurements, it was established that the differences between the two measuring instruments were not a calibration issue. Accuracy of the meters are ± 1 dBA for the stationary sound level meter and ± 2 dBA for the dosimeters. However, these differences are not large enough to account for the level differences measured in this study. A more plausible explanation is the contribution of the wearer's voice and activities that are recorded by the dosimeter close to the wearer. The influence of the wearer's voice in midrange noise level settings such as hospitals could thus add significantly to the equivalent noise level recorded by the dosimeter. For example, it has been found that in midrange noise levels from 45 dB and above a human tends to increase his/her voice in order to overcome background noise, referred to as the Lom-

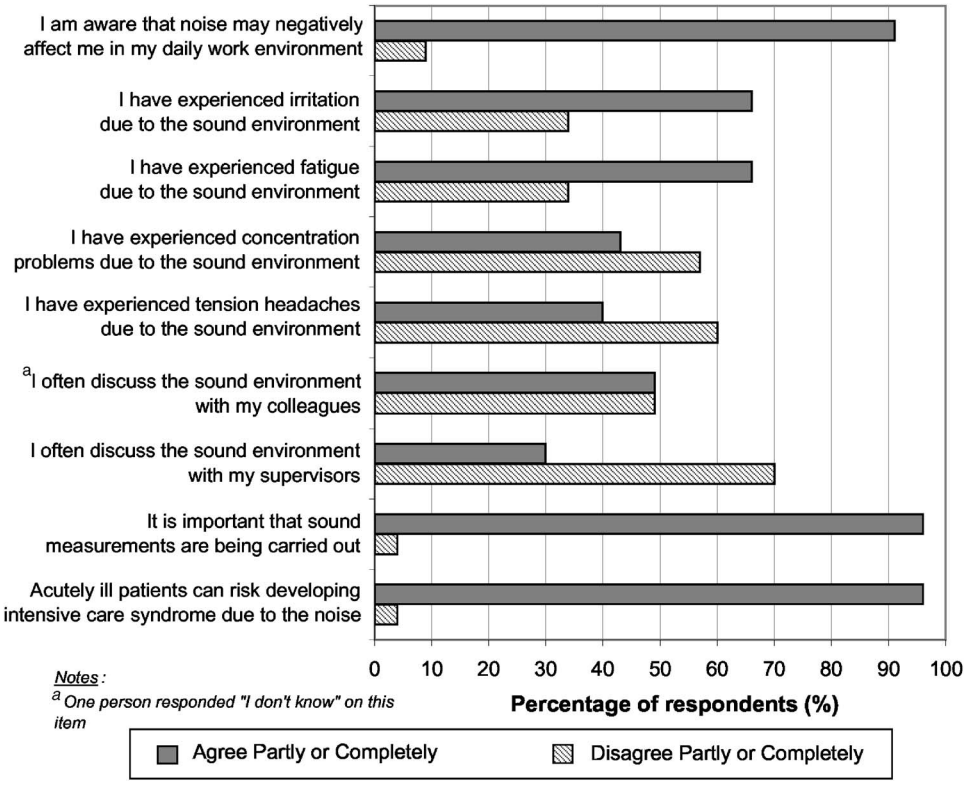


FIG. 8. Results from questionnaire items regarding perceived risks for patients and personnel. Statements are translated from Swedish to convey the general intention.

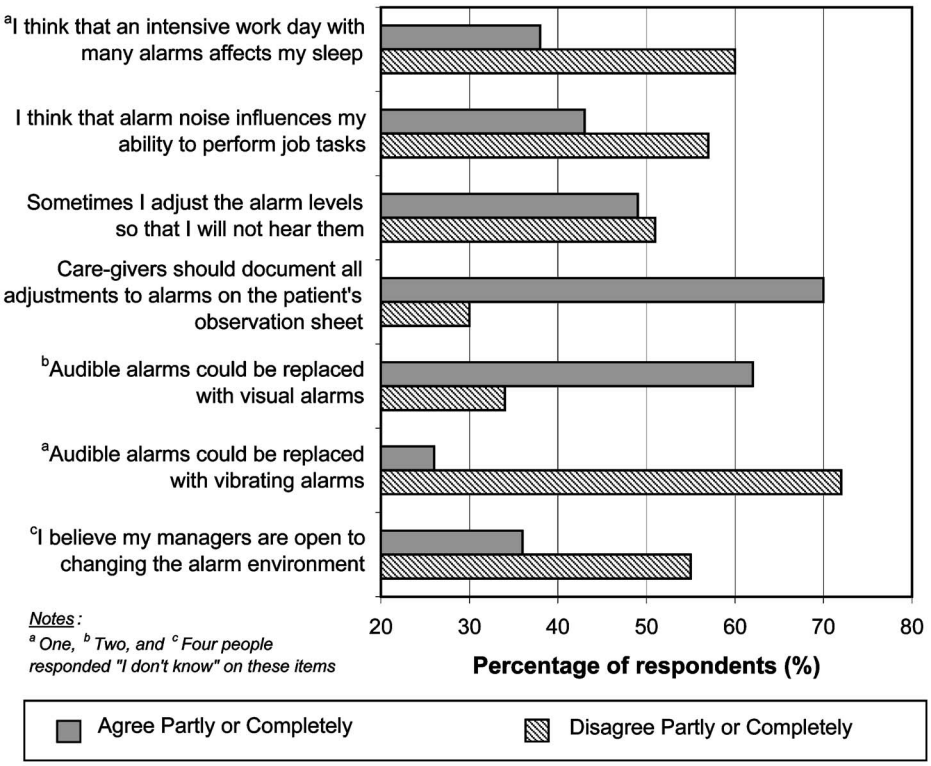


FIG. 9. Results from questionnaire items regarding alarms. Statements are translated from Swedish to convey the general intention.

bard effect.[52] A separate detailed study is currently being conducted by the authors to further examine the contribution of a wearer's voice on body-mounted microphone measurements in a range of noise environments from 45 to 75 dB L_{Aeq}.

Regardless of measurement method, it should be noted that the overall noise levels presented in Figs. 2 and 7 are well above the noise levels recommended by WHO and others. WHO recommends a maximum L_{Aeq} of 35 dB in rooms where patients are treated or observed, 30 dB in ward rooms, and a L_{AFMax} of no more than 40 dB at night (11 p.m.–7 a.m.).[3] Exceeding these levels in the intensive care unit measured in this study was not unexpected as this has been shown in many of the studies previously mentioned.[2]

The current measurements were made in an occupied state. The WHO guidelines do not specifically state whether conformance should be checked in an occupied or unoccupied condition, although community noise is defined as noise emitting from all sources except noise at the industrial workplace.[3] Other guidelines for indoor noise may specify unoccupied levels.[50] This subtle difference has been neglected by many previous studies, which often either present occupied levels or do not state the occupancy. In the current project, there was never an opportunity to measure the ICU in an unoccupied condition. However, guidance can be given by the restorative intervals between the alarms, when it is likely that the background noise was primarily due to steadily operating medical equipment and ventilation systems. Staff speech, distress noises from patients, doors closing, alarms, etc., were probably not as prevalent during these periods. These periods can therefore be viewed as more representative of what the "unoccupied" levels might be, and yet they also do not meet the WHO recommendations. It should also be noted that although the acoustic properties measured in unoccupied rooms play a role for the resulting sounds, the occupied sounds that the patients and the personnel are exposed to undoubtedly need to be attended to from a health standpoint. It is therefore highly important that relevant descriptors of the occupied rooms are developed and related to the human response.

It is often suggested that the maximum sound pressure level should be reported alongside the number of events,[3] though this information is typically omitted in the previous literature. This can lead to a misleading impression of the sound environment. The maximum and peak levels reported in Fig. 2 and in previous work could possibly be more indicative of isolated events rather than typical maximum and peak levels heard in the unit. For example, in Fig. 2, L_{AFMax} values are greater than 85 dB and L_{CPeak} values exceed 105 dB, yet without additional analysis it is unknown how often over a typical work day or week these values are exceeded. Although it is true that the L_{Aeq} would shift to a higher value if the L_{AFMax} exceeded 85 dB most of the time, it is still difficult to understand the offensiveness of maximum and peak values without further analysis. The statistical distribution analysis of the maximum and peak levels in this study provides a more sensitive gauge of the severity of the background noise than would have been achieved by relying on the data in Fig. 2. From the statistical distribution of levels (Fig. 5), one can see that the maximum and peak levels shown in Fig. 2 were reached less than 5% of the time. Three recent studies also highlight the importance of related types of analyses.[12,35,40] Kracht *et al.* presented peak values in operating rooms as a function of the percent of time during surgery that various levels were exceeded.[12] Williams *et al.* reported the percentage of time that recommended

equivalent, maximum, and peak levels were exceeded in neonatal intensive care units.[35] MacKenzie and Galbrun presented percentage of occurrences of maximum levels as a function of sound pressure level range.[40]

In this study, the use of statistical distributions of level and restorative periods also highlights the differences between the sound environments during the day versus the night. Figure 5 shows more impulsive-type noises occurring during the day, as reflected by higher maximum and peak levels' percentages during the day. In examining the restorative periods shown in Table II, differences between day and night are highlighted. In general, there are more restoration periods, and the mean and maximum lengths are larger at night.

Information about the spectral content of the ambient sound was obtained through octave and third-octave band analyses and indoor noise criteria ratings. A number of these criteria systems are currently used to quantify indoor background noise, including those shown in Table III.[50] These systems generally provide an indication of the overall loudness of noise and the spectral character. The exact frequency ranges associated with each spectral descriptor vary somewhat depending on the criteria method being used. In general, a "rumbly" descriptor indicates excessive low frequency energy (<500 Hz), "roaring" indicates midfrequencies (125–500 Hz), and "hissy" indicates high frequencies (1–8 kHz). In this study the noise was rated as "hissy" according to the criteria considered. Previous research has shown that perception of hiss in background noise can negatively influence productivity on certain types of tasks.[7] Although the overall level was found to decrease during the quieter restorative periods in the current study, the spectral quality remained imbalanced.

B. Questionnaire results

The alarm environment was believed to be related to negative reactions by some of the nurses in this study. Previous studies have also raised concerns over the density of alarms in the ICU and the potential effects on occupants.[53–55] Studies have shown that no action is taken by the staff for the majority of audible alarms,[2,56,57] and that many alarms are incorrectly identified even by experienced staff members.[58] Other options such as visual and vibrating alarm systems are available, and some hospitals have implemented these successfully. One study indicated better identification rates using vibro-tactile alarms as compared to auditory alarms or a combination of auditory/vibro-tactile systems.[59] In this study, 62% of the nurses surveyed felt that audible alarms could be replaced with visual alarms, while only 26% felt that vibrating alarms were an option. Although a majority of the nurses are willing to change to a visual alarm system, 55% of the nurses did not think that their managers were open to changing the alarm environment.

A large majority of the nurses also stated the risk for patients developing ICU syndrome due to the noise. The clinical ramifications of ICU syndrome are still under investigation, but in one study the duration of delirium was the single strongest predictor of length of stay in the ICU and the hospital.[60]

VI. CONCLUSIONS

This article has provided detailed information about the noise in a neurological intensive care unit. Several relevant descriptors of the sound environment were investigated including L_{Aeq}, L_{AFMax}, and L_{CPeak}. Statistical distributions of peak and maximum levels, restorative periods, and spectral quality analysis including criteria ratings provide a more sensitive gauge of the nature of the background noise than would have been achieved by relying only on L_{Aeq}, L_{AFMax}, and L_{CPeak}. Future work should consider these and other aspects of the sound environment in addition to the overall noise levels, such as tonality. Additionally, measurements of both the patient and staff exposures revealed interesting differences that the authors are currently investigating.

Given the prevalence of annoyance to alarms and the nurses' apparent willingness to modify the alarm environment in this study, further research in this area is also warranted. Research is currently being conducted by the authors to investigate occupants' physiological reactions to certain ICU stimuli, including alarms. Further research into the usefulness and practicalities of visual and vibrating alarm systems are also needed.

Overall, more research is required on the impact of noise in health care settings on occupants, including patients, staff, and visitors. The current findings indicated that many of the staff perceived noise as a problem that may contribute to stress symptoms such as irritation, fatigue, tension headaches, and difficulties concentrating. Additional research is needed to correlate self-reported stress with physiological reactions, and to understand the relative contribution of noise to stress response versus other environmental or occupational conditions. The authors are currently pursuing further research on the physiological effects of hospital noise in occupants, including the development of patient ICU syndrome as related to noise.

One important aspect to consider in hospital or health care facility noise research is the uniqueness of each type of space. The architectural design, types of equipment, condition of patients, staff work environment, presence of visitors, and expectations of occupants will vary depending on what type of space is being considered. Additional research on the acoustical characteristics of a variety of medical spaces, as well as evaluations of occupant psychological and physiological reactions, will give further understanding of how to improve the modern hospital soundscape.

ACKNOWLEDGMENTS

This work has been supported by the Swedish Council for Working Life and Social Research Diarienumer 2001-2558 and the Acoustical Society of America F.V. Hunt Fellowship. We are indebted to Agneta Agge for her work in the data collection and analysis, and to the intensive care unit staff for their cooperation. We appreciate the advisement of Dr. Hans Ragneskog of the Sahlgrenska Academy with re-

gards to the questionnaire study. We are also grateful to Dr. Ilene Busch-Vishniac and Dr. James West of McMaster and Johns Hopkins University for their advice.

[1]C. Baker, "Sensory overload and noise in the ICU: Sources of environmental stress," Crit. Care Nurs. Q. **6**, 66–79 (1984).
[2]I. Busch-Vishniac, J. West, C. Barnhill, T. Hunter, D. Orellana, and R. Chivukula, "Noise levels in Johns Hopkins Hospital," J. Acoust. Soc. Am. **118**, 3629–3645 (2005).
[3]"Guidelines for community noise," edited by B. Berglund and T. Lindvall, Technical Report, World Health Organization, Geneva, Switzerland, 1999.
[4]B. Berglund, U. Berglund, and T. Lindvall, "Scaling loudness, noisiness and annoyance of community noise," J. Acoust. Soc. Am. **60**, 1119–1125 (1976).
[5]U. Landström, A. Kjellberg, L. Söderberg, and B. Nordström, "The effects of broadband, tonal, and masked ventilation noise on performance, wakefulness, and annoyance," Low Freq. Noise, Vib., Act. Control **10**, 112–122 (1991).
[6]K. Persson Waye, R. Rylander, S. Benton, and H. G. Leventhall, "Effects on performance and work quality due to low frequency ventilation noise," J. Sound Vib. **205**, 467–474 (1997).
[7]E. Bowden and L. Wang, "Relating human productivity and annoyance to indoor noise criteria systems: A low frequency analysis," Symposium Transactions SYMP-0076–2004. R1, ASHRAE Winter Meeting, Orlando, FL, 5–9 February (2005).
[8]E. Ryherd and L. Wang, "Effects of exposure duration and type of task on subjective performance and perception in noise," Noise Control Eng. J. **55**, 334–347 (2007).
[9]R. Aitken, "Quantitative noise analysis in a modern hospital," Arch. Environ. Health **37**, 361–364 (1982).
[10]S. Falk and N. Woods, "Hospital noise: Levels and potential health hazards," N. Engl. J. Med. **289**, 774–781 (1973).
[11]D. Orellana, I. Busch-Vishniac, and J. West, "Noise in the adult emergency department of Johns Hopkins Hospital," J. Acoust. Soc. Am. **121**, 1996–1999 (2007).
[12]J. Kracht, I. Busch-Vishniac, and J. West, "Noise in the operating rooms of Johns Hopkins Hospital," J. Acoust. Soc. Am. **121**, 2673–2680 (2007).
[13]W. Passchier-Vermeer and W. Passchier, "Noise exposure and public health," Environ. Health Perspect. **108**, 123–131 (2000).
[14]S. Parthasarathy and M. Tobin, "Sleep in the intensive care unit," Intensive Care Med. **30**, 197–206 (2004).
[15]C. Baker, B. Garvin, C. Kennedy, and B. Polivka, "The effect of environmental sound and communication on CCU patients' heart rate and blood pressure," Res. Nurs. Health **16**, 415–421 (1993).
[16]J. Hagerman, G. Rasmanis, V. Blomkvist, R. Ulrich, C. Eriksen, and T. Theorell, "Influence of intensive coronary care acoustics on the quality of care and physiological state of patients," Int. J. Cardiol. **98**, 267–270 (2005).
[17]D. Fife and E. Rappaport, "Noise and hospital stay," Am. J. Public Health **66**, 680–681 (1976).
[18]B. Minkley, "A study of noise and its relationship to patient discomfort in the recovery room," Nurs. Res. **17**, 247–250 (1968).
[19]A. Wysocki, "The effect of intermittent noise exposure on wound healing," Adv. Wound Care **9**, 35–39 (1996).
[20]I. Bennun, "Intensive care unit syndrome: A consideration of psychological interventions," Br. J. Med. Psychol. **74**, 369–377 (2001).
[21]A. Granberg, I. Bergbom Engberg, and D. Lundberg, "Patients' experience of being critically ill or severely injured and cared for in an intensive care unit in relation to the ICU syndrome. I," Intensive Crit. Care Nurs. **14**, 294–307 (1998).
[22]V. Murthy, S. Malhotra, I. Bala, and M. Raghunathan, "Detrimental effects of noise on anaesthetists," Can. J. Anaesth. **42**, 608–611 (1995).
[23]M. Topf, "Noise-induced occupational stress and health in critical care nurses," Hosp. Top. **66**, 30–34 (1988).
[24]W. Morrison, E. Haas, D. Shaffner, E. Garrett, and J. Fackler, "Noise stress and annoyance in a pediatric intensive care unit," Crit. Care Med. **31**, 113–119 (2003).
[25]M. Topf and E. Dillon, "Noise-induced stress as a predictor of burnout in critical care nurses," Heart Lung **17**, 247–250 (1988).
[26]G. Holmes, K. Goodman, D. Hang, and V. McCorvey, "Noise levels of orthopedic instruments and their potential health risks," Orthopedics **19**, 35–37 (1996).
[27]V. Blomkvist, C. Eriksen, T. Theorell, R. Ulrich, and G. Rasmanis, "Acoustics and psychosocial environment in intensive coronary care," Occup. Environ. Med. **62**, 1–8 (2005).
[28]M. MacLeod, J. Dunn, I. J. Busch-Vishniac, and J. E. West, "Quieting Weinberg 5C: A case study in hospital noise control," J. Acoust. Soc. Am. **121**, 3501–3508 (2007).
[29]K. Thomas and P. Martin, "The acoustic environment of hospital nurseries: NICU sound environment and the potential problems for caregivers," J. Perinatol **20**, S94–S99 (2000).
[30]M. K. Philbin and L. Gray, "Changing levels of quiet in an intensive care nursery," J. Perinatol **22**, 455–460 (2002).
[31]P. Bremmer, J. Byers, and E. Kiehl, "Noise and the premature infant: Physiological effects and practice implications," J. Obstet. Gynecol. Neonatal Nurs. **32**, 447–454 (2003).
[32]C. Krueger, S. Wall, L. Parker, and R. Nealis, "Elevated sound levels within a busy NICU," Neonatal Netw. **24**, 33–37 (2005).
[33]E. Barreto, B. Morris, M. K. Philbin, L. Gray, and R. Lasky, "Do former preterm infants remember and respond to neonatal intensive care unit noise?," Early Hum. Dev. **82**, 703–707 (2006).
[34]J. Byers, W. Waugh, and L. Lowman, "Sound level exposure of high-risk infants in different environmental conditions," Neonatal Netw. **25**, 25–32 (2006).
[35]A. Williams, W. van Drongelen, and R. Lasky, "Noise in contemporary neonatal intensive care," J. Acoust. Soc. Am. **121**, 2681–2690 (2007).
[36]P. Gast and C. Baker, "The CCU patient: Anxiety and annoyance to noise," Crit. Care Nurs. Q. **12**, 39–54 (1989).
[37]T. Meyer, S. Eveloff, M. Bauer, W. Schwartz, N. Hill, and R. Millman, "Adverse environmental conditions in the respiratory and medical ICU settings," Chest **105**, 1211–1216 (1994).
[38]D. Kahn, T. Cook, C. Carlisle, D. Nelson, N. Kramer, and R. Millman, "Identification and modification of environmental noise in an ICU setting," Chest **114**, 535–540 (1998).
[39]B. Walder, D. Francioli, J. Meyer, M. Lancon, and J. Romand, "Effects of guidelines implementation in a surgical intensive care unit to control night time light and noise levels," Crit. Care Med. **28**, 2242–2247 (2000).
[40]D. MacKenzie and L. Galbrun, "Noise levels and noise sources in acute care hospital wards," Build. Services Eng. Res. Technol. **28**, 117–131 (2007).
[41]B. Hilton, "Noise in acute patient care areas," Res. Nurs. Health **8**, 283–291 (1985).
[42]D. Balogh, E. Kittinger, A. Benzer, and J. Hackl, "Noise in the ICU," Intensive Care Med. **19**, 343–346 (1993).
[43]C. Tsiou, D. Eftymiatos, E. Theodossopoulou, P. Notis, and K. Kiriakou, "Noise sources and levels in the Evgenidian Hospital intensive care unit," Intensive Care Med. **24**, 845–847 (1998).
[44]M. Monsén and U. Edéll-Gustafsson, "Noise and sleep disturbance factors before and after implementation of a behavioural modification programme," Intensive Crit. Care Nurs. **21**, 208–219 (2005).
[45]British Standard Institution, "Electroacoustics-Sound level meters—Part 1: Specifications," BS EN 61672-1, British Standards, London, UK.
[46]American National Standards Institute, "Measurement of occupational noise exposure," ANSI S 12.19, 1996, Acoust. Soc. Am., Melville, NY.
[47]K. Persson Waye, J. Bengtsson, A. Agge, and M. Björkman, "A descriptive cross-sectional study of annoyance from low frequency installations in an urban environment," Noise Health **5**, 35–46 (2003).
[48]E. Pedersen and K. Persson Waye, "Perception and annoyance due to wind turbine noise—A dose-response relationship," J. Acoust. Soc. Am. **116**, 3460–3470 (2004).
[49]E. Öhrström, "Effects of low levels of road traffic noise during the night: A laboratory study on number of events, maximum noise levels and sensitivity," J. Sound Vib. **179**, 603–615 (1995).
[50]American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc., "Sound and vibration control," *2007 ASHRAE HVAC Applications Handbook* (Atlanta, GA, 2003), Chap. 47, pp. 47.30–47.35.
[51]L. Ljungkvist, "Sound and noise in the ICU, a study in nursing care," (in Swedish), Masters thesis, Göteborg University, Göteborg Sweden, 2007.
[52]H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709 (1971).
[53]J. Kerr and B. Hayes, "An 'alarming' situation in the intensive therapy unit," Intensive Care Med. **9**, 103–104 (1983).
[54]C. Meredith and J. Edworthy, "Are there too many alarms in the intensive care unit? An overview of the problems," J. Adv. Nurs. **21**, 15–20 (1995).
[55]J. Phillips and J. Barnsteiner, "Clinical alarms: Improving efficiency and effectiveness," Crit. Care Nurs. Q. **28**, 317–323 (2005).

[56]M. Wallace, M. Ashman, and M. Matjasko, "Hearing acuity of anesthesiologists and alarm detection," Anaesthesiology **81**, 13–28 (1994).
[57]L. Stanford, J. McIntyre, and J. Hogan, "Audible alarm signals for anesthesia montitoring equipment," Int. J. Clin. Monit Comput. **1**, 251–256 (1985).
[58]A. Cropp, L. Woods, and D. Raney, "Name that tone: The proliferation of alarms in the intensive care unit," Chest **105**, 1217–1220 (1994).
[59]J. Ng, J. Man, S. Fels, G. Dumont, and J. Ansermino, "An evaluation of vibro-tactile display prototype for physiological monitoring," Anesth. Analg. (Baltimore) **101**, 1719–1724 (2005).
[60]E. Ely, S. Gautam, R. Margolin, J. Francis, L. May, T. Speroff, B. Truman, R. Dittus, R. Bernard, and S. K. Inouye, "The impact of delirium in the intensive care unit on hospital length of stay," Intensive Care Med. **27**, 1892–1900 (2001).

Noise in the operating rooms of Greek hospitals

Chrisoula Tsiou[a)]
Department of Nursing, Technological Education Institute of Athens, A.g. Spiridonos 1 and Palikaridi, Egaleo 12210, Athens, Greece

Gerasimos Efthymiatos
Consultant in Acoustics and Audio Systems in Greece, Sound & Acoustics, Chiou 48, Ag. Paraskevi, 15343, Greece

Theophanis Katostaras
Department of Nursing, University of Athens, Papadiamantopoulou 123, Goudi 11527, Athens, Greece

(Received 12 March 2007; accepted 13 November 2007)

This study is an evaluation of the problem of noise pollution in operating rooms. The high sound pressure level of noise in the operating theatre has a negative impact on communication between operating room personnel. The research took place at nine Greek public hospitals with more than 400 beds. The objective evaluation consisted of sound pressure level measurements in terms of L_{eq}, as well as peak sound pressure levels in recordings during 43 surgeries in order to identify sources of noise. The subjective evaluation consisted of a questionnaire answered by 684 operating room personnel. The views of operating room personnel were studied using Pearson's X^2 Test and Fisher's Exact Test (SPSS Version 10.00), a t-test comparison was made of mean sound pressure levels, and the relationship of measurement duration and sound pressure level was examined using linear regression analysis (SPSS Version 13.00). The sound pressure levels of noise per operation and the sources of noise varied. The maximum measured level of noise during the main procedure of an operation was measured at L_{eq}=71.9 dB(A), L_1=84.7 dB(A), L_{10}=76.2 dB(A), and L_{99} =56.7 dB(A). The hospital building, machinery, tools, and people in the operating room were the main noise factors. In order to eliminate excess noise in the operating room it may be necessary to adopt a multidisciplinary approach. An improvement in environment (background noise levels), the implementation of effective standards, and the focusing of the surgical team on noise matters are considered necessary changes. [DOI: 10.1121/1.2821972]

PACS number(s): 43.50.Qp, 43.50.Ki, 43.50.Jh, 43.50.Hg [KA] Pages: 757–765

I. INTRODUCTION

The organization and the function of the operating room are directly related to the use of contemporary machinery and tools, which form a noise source.[1–3] In addition, there are other administrative and organizational factors together with significant surgical programs.[4–6] These include the construction of the building and human activities.

The effects of noise in the operating room are inner acoustic and outer acoustic.[7–9] For personnel in the orthopedic operating rooms there is a possibility of hearing loss.[10–12] The degree of risk increases depending on role of the person in the operating theatre, the duration that they are exposed to the noise, and the coexistence of various predisposed factors.[1,9,13] The high pressure level of noise in the operating room has a negative impact on communication between operating room personnel and on their mood. It also prevents machine alarm signals from being noticed.[14,15] Operating room staff are more likely to be influenced by noise when there are inadequate working conditions.[5]

For patients undergoing surgery, the consequences of noise are not clear.[16] It is assumed that the type of operation, the type of anaesthesia, and the depth of the general anaesthetic determine the degree of noise to which the patients are exposed.[16–18]

The resolution of the problem of noise pollution involves the use of noise-free machinery, the implementation of effective administrative systems, measures regarding hygienic working conditions,[6] and the socio-psychological study of the surgical team.[4,16,19] The serious and important work undertaken in an operating room and the need for high levels of concentration undoubtedly make it necessary to decrease noise pollution.

II. METHODS

A. The study and the place where the study was conducted

This study is an evaluation of the problem of the effects of high sound pressure levels of noise in operating rooms that took place at nine Greek public hospitals with more than 400 beds. It intends to locate the sources and the levels of noise in the operating room and to present the opinions of the relevant personnel in relation to this matter. Understanding noise pollution is likely to contribute to the improvement of the acoustic environment in operating rooms via a more scientific approach.

[a)]Author to whom correspondence should be addressed. Electronic mail: ctsiou@teiath.gr.

TABLE I. Different opinions of surgeons, anaesthetists, and nursing personnel regarding noise pollution in the operating room. This study used 391 surgeons, 74 anaesthetists, and 219 operating room nursing personnel.

Questions	Affirmative answers			p value
	Surgeons[a] %	Anaesthetists[b] %	Nursing personnel[c] %	
Is there any noise in the operating room?	65.6	85.1	78.2	0.000
Do you feel that noise has a negative impact on your job?	65.8	84.1	71.8	0.006
Does noise in the operating room disturb you?	53.0	54.0	53.9	0.100
Which are the main sources of noise in the operating room?				
Conversation?	74.2	81.1	68.5	0.082
Louder conversation (such as arguments)?	30.4	60.8	48.4	0.000
Machines being operated?	21.5	41.9	57.1	0.000
External noise?	33.0	29.7	22.4	0.021
Air-conditioning systems?	16.4	25.7	30.1	0.000

[a]Consisting of surgeons and nonspecialist surgeons.
[b]Consisting of anaesthetists and nonspecialist anaesthetists.
[c]Consisting of operating room nurses or assistants.

From our observations we found that (a) only one hospital, built after 1995, had sound insulation and (b) in the other eight hospitals (i) there was no ancillary room between the corridor and the operating room where preparations for surgery could be made, and which would also serve to isolate the operating room from noise from the corridor, (ii) air conditioning was inadequate or nonexistent and doors were constantly left open because of the heat, and (iii) there were no internal communication systems or computers in the operating rooms, requiring personnel to come and go from the operating room to perform their duties.

B. Objective measurements of noise in the operating room

The basic methodology of the study is based on 129 objective measurements of noise in operating rooms, as well as subjective evaluation via questionnaires (see Tables I and II for sample of questions). Sound measurements were acquired during a total of 43 surgical operations comprising 26 nonorthopedic and 17 orthopedic surgeries (Table III), in various operating rooms. All the surgeries took place in the operating rooms of old hospitals (built before 1950), with the

TABLE II. Different opinions of "head surgeons" and the "rest of the surgical team" regarding noise pollution in the operating room. This study used 684 operating room personnel.

Questions	Affirmative answers		p value
	Head surgeons[a] %	Rest of the surgical team[b] %	
Is there any noise in the operating room?	67.2	73.7	0.094
Do you feel that noise has a negative impact on your job?	67.0	70.6	0.381
Does noise in the operating room disturb you?	58.6	48.8	0.052
Which are the main sources of noise in the operating room?			
Conversation?	78.5	71.3	0.057
Louder conversation (such as arguments)?	34.9	40.9	0.157
Machines being operated?	18.3	41.3	0.000
External noise?	35.5	26.5	0.021
Air-conditioning systems?	16.7	23.5	0.054

[a]Consisting of 186 specialist surgeons.
[b]Consisting of 498 nonsurgeons—nonspecialist surgeons, anaesthetists, nonspecialist anaesthetists, operating room nurses or assistants.

TABLE III. Types of surgeries and duration of sound measurement (in min) per operation.

No.[a]	NonOrthopedic surgeries	Presurgical Time[b]	Surgical Time[b]	Postsurgical time[b]	Total time per surgery
1	Appendectomy	15′	10′	10′	35′
2	TUR of abdominal walls	20′	10′	07′	37′
3	Tonsillotomy	10′	15′	15′	40′
4	Laparoscopic cholecystectomy	20′	25′	10′	55′
5	Deficiency of skin-plastic surgery	45′	25′	15′	85′
6	Bubonocele	10′	30′	15′	55′
7	Tonsillotomy	20′	30′	10′	60′
8	Bubonocele	15′	30′	10′	55′
9	Appendectomy	25′	30′	15′	70′
10	Appendectomy	20′	35′	10′	65′
11	Varicocele	20′	35′	05′	60′
12	Cataract	10′	35′	30′	75′
13	Cataract	10′	40′	10′	60′
14	Tonsillotomy	15′	45′	10′	70′
15	Spleenectomy	30′	45′	25′	100′
16	Colectomy	40′	50′	30′	120′
17	Abdominal hysterectomy	20′	55′	30′	105′
18	Gastrectomy	20′	75′	15′	110′
19	Diskectomy	40′	90′	20′	150′
20	Gastroenteroanastomosis	30′	100′	10′	140′
21	Gastrectomy	30′	105′	45′	180′
22	Colectomy	15′	110′	30′	155′
23	Thyroidectomy	15′	120′	15′	150′
24	Vasectomy	20′	150′	25′	195′
25	Aorto-coronary bypass	60′	165′	60′	285′
26	Vagotomy	30′	245′	45′	320′
	Orthopaedic surgeries				
27	Diatrochanteric fracture—Osteosynthesis Richard's	30′	20′	15′	65′
28	Hip fracture (under head)—Nontotal arthoplasty hip	25′	30′	20′	75′
29	Hip fracture (under head)—Osteosynthesis By pollar	15′	40′	35′	90′
30	Total arthroplasty hip	20′	40′	50′	110′
31	Total arthroplasty knee	35′	40′	25′	100′
32	Forearm fracture—Osteosynthesis	20′	40′	30′	90′
33	Total arthroplasty knee	25′	45′	35′	105′
34	Revision of total arthroplasty of hip	30′	50′	40′	120′
35	Forearm fracture—Osteosynthesis	20′	55′	20′	95′
36	Total arthroplasty hip	30′	60′	20′	110′
37	Leg fracture—Internal osteosynthesis	50′	60′	20′	130′
38	Total arthroplasty knee	25′	60′	25′	110′
39	Total arthroplasty knee	45′	60′	30′	135′
40	Total arthroplasty knee	35′	70′	20′	125′
41	Total arthroplasty hip	20′	70′	30′	120′
42	Hip fracture—Internal osteosynthesis	25′	80′	30′	135′
43	Total arthroplasty hip	40′	120′	30′	190′
		1100′	2645′	997′	4742′
	Total time of measurements 79 h and **2 min**				

[a]The number for each surgery is used in all figures.
[b]Three successive sound measurements are taken per surgery. The first is presurgical time, the second is surgical time, and the third is postsurgical time.

exception of numbers 22 and 39, which were carried out in two operating rooms at a new hospital built after 1995 (Table III).

The researcher monitored the sound level meter, which was located directly behind the surgical team, recording high noise levels [in excess of 75 dB(A)] and correlating them with sources of noise originating from personnel, equipment, etc.

In this study, each operation was divided into three chronological stages (Fig. 1). The criteria for this three-stage division were the medical–surgical activities taking place in the operating room and the need to identify which stage of the operation was associated with the highest level of noise pollution. Sound measurements began as soon as the patient was placed on the operating table. The first measurement stage took place during the presurgical preparatory stage ("presurgical time"). The second stage was during the main part of the operation ("surgical time"). In the third stage of

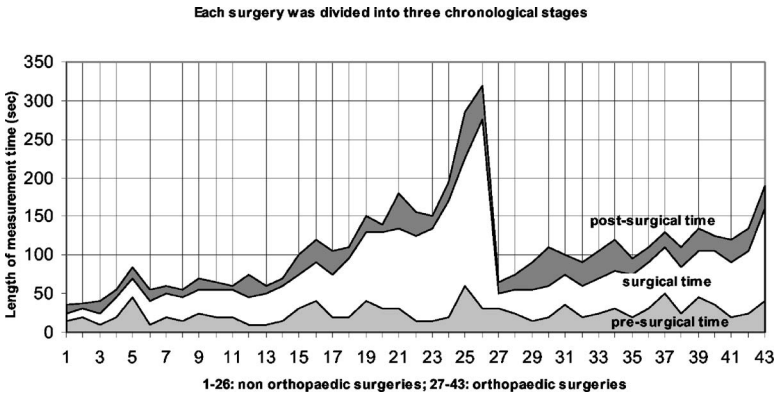


FIG. 1. The length of time of measurements for presurgical time, surgical time, and postsurgical time.

measurement ("postsurgical time") the sound pressure level was correlated with activities such as final suturing and the removal of all surgical material. The awakening and transportation of the patient out of the operating room marked the conclusion of the sound-measurement process. The duration of each stage of measurement varied depending on the type of surgery (Table III) and the associated surgical activities. The mean duration of the first measurement (presurgical time) was approximately 25 min and that of the third measurement (postsurgical time) approximately 23 min. In contrast, surgical time, which corresponds to the most critical stage of the surgery, was measured for a longer period (up to a maximum of 245 min). The observations and the 129 measurements (3 measurements per operation) of the study were made during the almost 79 h duration of the operations as a whole. The results of these sound measurements and the related observations were entered on a specially designed form. For sound measurements we used the Bruel & Kjaer type 2231 sound level meter set to slow mode. All measurements were recordings of the sound pressure level (SPL) in decibel A or dB(A). At the end of each measurement the measured data were analyzed in L_{eq}, L_1, L_{10}, L_{50}, L_{90}, L_{99}, and MAXL.

C. Subjective evaluation of the noise in the operating room via questionnaires

For the subjective evaluation of the noise in the operating room, the completion of a pre-formulated questionnaire was used. The pilot study involved 75 doctors and nursing personnel working in operating rooms. The questionnaire included general questions, for example medical specialization, age, and sex, and specific questions regarding noise and the environment within the operating room. A total of 684 doctors and nursing personnel working in operating rooms (surgeons, nonspecialist surgeons, anaesthetists, nonspecialist anaesthetists, operating room nurses or assistants), completed the questionnaire, having been given some preliminary information about the aim of the study. The views of operating room personnel were studied using Pearson's X^2 Test and Fisher's Exact Test (SPSS Version 10.00). A comparison was made of the views expressed via the questionnaire first in terms of professional role (Table I and then in terms of level of responsibility (Table II).

Table I shows the differences in the opinions of 391 surgeons, 74 anaesthetists, and 219 nursing personnel. The term "surgeons" refers to specialist and nonspecialist surgeons, the term "anaesthetists" refers to specialist and nonspecialist anaesthetists, while the term "nursing personnel" refers to operating room nurses or assistants.

Table II shows the differences in the opinions of 186 "head surgeons" and 498 nonsurgeons. The term head surgeons refers to specialist surgeons, while the term "rest of the surgical team" refers to nonspecialist surgeons, anaesthetists, nonspecialist anaesthetists, and operating room nurses or assistants.

III. RESULTS

A. Sound pressure level in operating rooms

The summary results of noise measurements in operating rooms (see Tables IV and V) indicate that the minimum level of noise in the operating room was 46.7 dB(A) and the maximum level 106 dB(A). The measurement parameters indicate that at times, noise in the operating rooms was significant (Fig. 2).

Regarding the sound measurements it was found that all the noise variables both in orthopedic and nonorthopedic surgeries follow a normal distribution (Kolmogorov-Smirnov test, $p > 0.05$. The t-test (SPSS Version 13.00) was thus applied to compare mean (SPL) in 12 different situations measured in orthopedic and nonorthopedic surgeries. Orthopedic and nonorthopedic surgeries were compared in terms of parameters L_1, L_{10}, L_{eq}, and L_{90}. Each parameter was checked

TABLE IV. Total results of sound measurements in 43 operations. Maximum and minimum levels of sound in L_{eq},[a] L_1, L_{10}, L_{50}, L_{90}, L_{99}, and MAXL in dB(A).

Sound parameters	Sound pressure level in decibel-A for each operating stage		
	Presurgical time	Surgical time	Postsurgical time
L_{eq}	61.1–78.2	57.4–70.1	60.5–74.1
L_1	70.7–90.2	68.2–83.2	69.7–81.7
L_{10}	63.2–79.7	59.7–73.2	60.7–75.7
L_{50}	53.7–71.2	51.7–64.2	53.7–66.7
L_{90}	49.2–61.2	48.2–58.7	49.7–60.7
L_{99}	47.2–59.2	46.7–56.7	47.2–58.2
MAXL	83.6–99.4	84.7–100	78.8–106

[a]L_{eq}: Equivalent sound pressure level—the steady sound level that, over a specified period of time, would produce the same energy equivalence as the fluctuating sound level actually occurring.

TABLE V. Sources of noise during the operation and maximum instantaneous sound pressure level (SPL) per noise source in dB(A).

Sources of noise during the operation	dB(A)
Connection/disconnection of gas supply	**106**
Displacement of equipment (wooden pedestals, seats etc.).	**100**
Tools	**94.8**
Objects falling on the floor	**94.5**
Handling of metal boxes containing sterilized material (drums or other)	**94.0**
Loud talking, voices, or laughter	**90.2**
Opening, closing, or knocking on door	**87.8**
Sneeze	**87.4**
Handling of paper packing or paper	**85.8**
Suction apparatus	**85.5**
The anaesthesiologist's tray	**84.7**
Coughing	**84.2**
Unpacking	**84.0**
Taking gloves	**84.0**
Anaesthesiological machine's alarm	**84.0**
Putting on gloves	**83.0**
Diathermy	**77.0**
Necessary discussion	**75.0**
Other factors (e.g., external noise, ashbin, etc.)	**>90**

in all three stages of surgery: (a) Presurgical time, (b) surgical time, and (c) postsurgical time. Table VI shows the results of the comparisons.

From the sound measurements in nonorthopedic surgeries it was established that pre-surgical time was the noisiest stage and that surgical time was the quietest stage (Table VI). Even though L_{eq} for surgical time was in excess of 57 dB(A), it can still be characterized, without exception, as the quietest stage of the operation (Table IV).

In terms of sources of noise in the operating room, there were various causes of distinct sudden noises with a sound pressure level in excess of 75 dB(A). Most sources were instantaneous, sudden, distinct sounds of high sound pressure level in excess of 90 dB(A). In summary, the tools, machinery, equipment, and surgical material used, together with human activities, were confirmed as the causes of increased noise levels in the operating room (Table V).

TABLE VI. Comparison of average sound measurements of orthopedic and nonorthopedic surgeries.

Sound measurements	Type of surgery: Orthopedic surgeries $\bar{X} \pm$ s.d.	Type of surgery: Nonorthopedic surgeries $\bar{X} \pm$ s.d.	p value	95% confidence interval
$L_{1\ in}$ dB(A)				
Presurgical time	78.31 ± 1.67	77.81 ± 3.72	0.604	−2.44356 1.43904
Surgical time	**79.73±2.97**	**73.78±2.72**	**0.000**	**−7.72755 −4.17743**
Postsurgical time	74.88 ± 2.69	75.86 ± 3.14	0.297	−0.89295 2.85539
$L_{10\ in}$ dB(A)				
Presurgical time	70.72 ± 1.88	70.35 ± 3.64	0.702	−2.30673 1.56636
Surgical time	**70.55±3.00**	**66.53±2.55**	**0.000**	**−5.57164 −2.48040**
Postsurgical time	66.73 ± 3.34	68.63 ± 3.22	0.069	−0.15249 3.96289
$L_{eq\ in}$ dB(A)				
Presurgical time	67.94 ± 1.30	67.82 ± 3.58	0.894	−1.95764 1.71374
Surgical time	**68.31±2.11**	**63.86±2.41**	**0.000**	**−5.89226 −2.99643**
Postsurgical time	64.74 ± 2.78	66.11 ± 3.31	0.165	−0.58928 3.33407
$L_{90\ in}$ dB(A)				
Presurgical time	56.46 ± 2.15	56.62 ± 3.09	0.805	−1.58179 1.89853
Surgical time	54.94 ± 1.66	54.74 ± 2.37	0.768	−1.53495 1.14128
Postsurgical time	55.20 ± 2.62	55.92 ± 2.49	0.367	−0.87637 2.32253

It can be seen from this study that mean noise intensity is related to the type and stage of surgery. Table VI shows that the two categories of surgery examined in this study (orthopedic and nonorthopedic) have very different noise levels during surgical time. L_1, L_{10}, and L_{eq} were far higher during orthopedic surgeries with a "significant" degree of difference (p=0.000). There was no difference between the

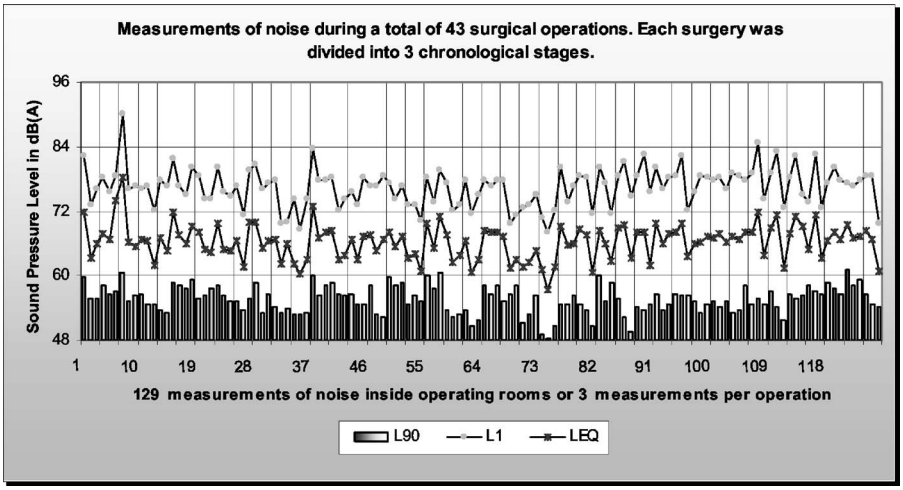


FIG. 2. Sound pressure level in dB(A) inside operating rooms.

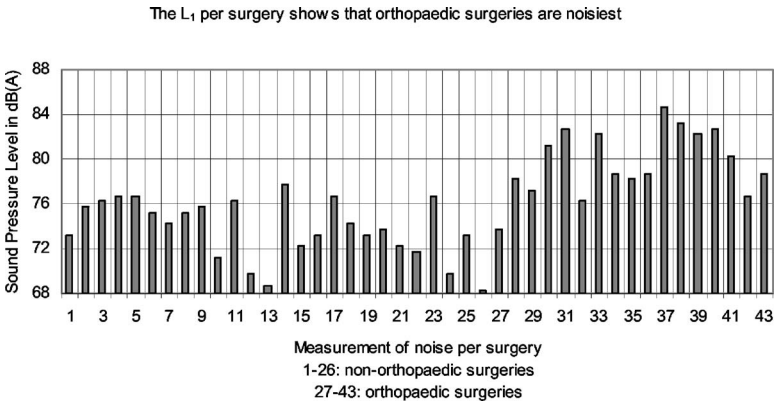


FIG. 3. L_1 during surgical time in nonorthopedic and orthopedic surgeries.

two groups in terms of L_{90} during surgical time and L_1, L_{10}, L_{eq}, and L_{90} during presurgical time and postsurgical time ($p > 0.05$).

From Fig. 3, it is clear that in surgical time for most nonorthopedic surgeries, L_1 was between 72 and 77 dB(A), whereas in orthopedic surgeries it was between 77 and 82 dB(A). Figure 4 shows L_{10} in surgical time per operation: In most nonorthopedic surgeries it was between 65 and 70 dB(A) but in orthopedic surgeries it was between 67 and 76.2 dB(A). In most operations in surgical time L_{90} measured from 54 to 57 dB(A). Figures 3 and 4 make it clear that the chief distinction between orthopedic and nonorthopedic surgeries was instantaneous noise (L_1).

Linear regression analysis (SPSS Version 13.00) was also used to measure the effect of length of time of measurement on sound pressure level. The results are shown in Table VII.

From Table VII it can be seen that as duration of measurement increased, noise levels decreased. This finding demonstrates that longer surgeries are quieter than shorter surgeries. From Figs. 3 and 4, it can also be observed that the quietest operations were appendectomy, cataract, colectomy, vasectomy, and vagotomy (Table III, Nos. 10, 12, 13, 22, 24, 26). Furthermore, Figs. 3 and 4 show that during surgical time: (i) Cataract operations (Nos. 12 and 13) were very quiet and (ii) longer surgeries (which we know to be more serious) were quieter. This is to be expected, as longer operations include more time for averaging sound pressures in the calculation of L_{eq}. Long operations may include considerable periods of quiet that reduce the overall L_{eq} for the surgery period.

From both the sound level meter and the simultaneous observations it is confirmed that the noise of orthopedic tools achieves such high pressure levels that it obliterates all other noise in the orthopedic operating rooms. The surgical tools and the suction process were the main sources of noise in the orthopedic operating room, while anaesthesiological and other machinery, people, and external noise factors were mainly responsible for the level of noise in nonorthopedic operating rooms.

B. The opinions of personnel regarding noise in the operating room

Subjective assessments of noise in the operating room can be seen in Tables I and II. Table I shows clearly that the professional role of individuals in the operating room influences their perception of noise. We found that in response to the questions "Is there any noise in the operating room?" and "Do you feel that noise has a negative impact on your job?,"

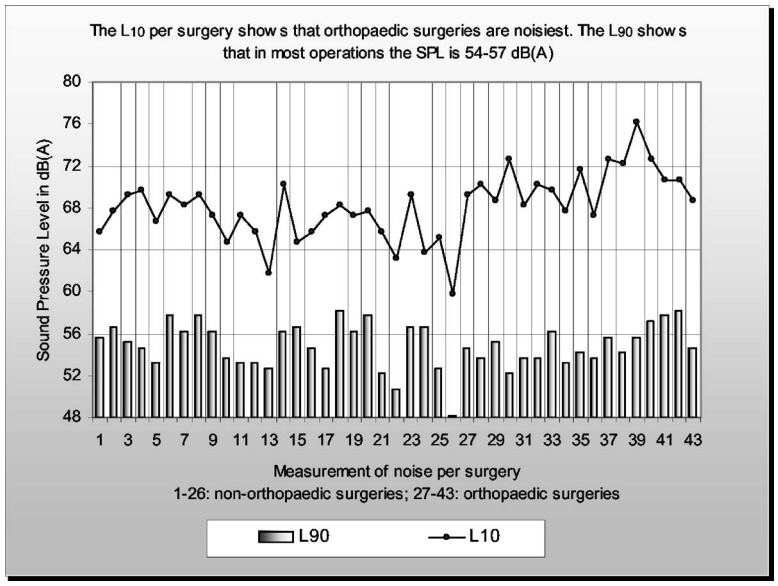


FIG. 4. L_{10} and L_{90} during surgical time in nonorthopedic and orthopedic surgeries.

TABLE VII. Dependence of sound pressure level from length of time of measurement (Linear Univariate Regression Analysis).[a]

Independent Length of time of measurement	Dependent Sound pressure level	Regression coefficient	p value
A—Presurgical time	L_1	−0.097	0.016
	L_{10}	−0.034	0.050
	L_{eq}	−0.100	0.008
	L_{90}	−0.096	0.008
B—Surgical time	L_1	−0.026	0.056
	L_{10}	−0.026	0.011
	L_{eq}	−0.024	0.023
	L_{90}	−0.014	0.053
C—Postsurgical time	L_1	−0.071	0.052
	L_{10}	−0.089	0.030
	L_{eq}	−0.083	0.032
	L_{90}	−0.87	0.004

[a]It is shown that the sound pressure level is often related to the duration of the surgery.

surgeons, anaesthetists, and nursing personnel answered in the affirmative. The differences in the opinions of the three groups were significant ($p<0.050$). On the basis of these figures it appears that anaesthetists are most sensitive to noise and that surgeons are least sensitive.

We found that in response to the question "Does noise in the operating room disturb you?," a relatively low proportion (53% to 54%) of all operating room personnel answered in the affirmative. There was no statistical difference in the answers of the three groups (surgeons, anaesthetists, and nursing personnel) to this question ($p=0.100$).

In response to the question "Which are the main sources of noise in the operating room?" equally high proportions (up to 81%) of all personnel cited "conversation" ($p>0.050$). In lower proportions and with statistically significant differences ($p<0.050$), the three groups also cited: (a) louder conversation (such as arguments), (b) machines being operated, (c) external noise, and (d) air-conditioning systems. Anaesthetists more frequently cited louder conversation as a source of noise, nursing personnel more frequently cited machines being operated, and air-conditioning systems, while surgeons more frequently cited external noise. In contrast, surgeons less frequently cited louder conversation, machines being operated, and air-conditioning systems as sources of noise, while nursing personnel less frequently cited external noise.

Table II shows that the level of responsibility of the individual also influences the perception of noise in the operating room. Specifically, a higher proportion of head surgeons reported that they were disturbed by noise in the operating room than did the rest of the surgical team ($p=0.052$). We also found that both head surgeons and the rest of the surgical team cited machines being operated and external noise as sources of noise, but with statistically significant differences ($p<0.050$). Specifically, a higher proportion of head surgeons cited external noise and a higher proportion of the rest of the surgical team cited machines being operated. Also in Table II, conversation was identified as a source of noise by a higher proportion of head surgeons ($p=0.057$), and air-conditioning systems by a higher proportion of the rest of the surgical team ($p=0.054$). Table II also shows that at three points of the questionnaire the differences were of borderline statistical significance ($p=0.052-0.057$). It is almost certain that with a larger number of responses there would have been clear statistical significance ($p<0.050$).

IV. COMMENTS

A. General

Noise is a very serious problem for patients and hospital staff.[20,21] Instantaneous, sudden, high-level sounds can have an impact on the reflex system of the autonomous nervous system.[14,22] These types of noise are believed to have an impact on human behavior and consequently make cooperation within the surgical team more difficult.[16] Previous authors regard high-energy instantaneous sounds as capable of diverting the route of the surgical lancet during an operation.[23] The effect of sudden noise on personnel is a function of a person's response to imminent noise, the sensitivity of the person to noise, the intensity and the origin of the noise, the information conveyed by the noise (via equipment or conversation), the position of every person in the room, the distance from the source of noise, the resonance of the building, the duration and the frequency of exposure to the noise, etc.[11,14,22,24]

B. Assumptions, predictions, and major findings

The factors that account for the level of noise during every operation are the type of operation (Table VI, Figs. 3 and 4), its duration (Table VII), the stage of the operation (Table IV), the type of equipment used (Table V), the number of people who take part in the operation, and the constructional parameters of the operating room.[23]

As earlier studies have shown,[12,21,24,25] the orthopedic surgical time lasts a long time and is very noisy. These findings (Table VI) confirm and support the view that the hearing of the personnel in orthopedic operating rooms is in danger from the cumulative effects of noise.[12,21,24,25] It is suggested that personnel should make frequent use of audiometry[10,11] and use ear-plugs, at least in the ear that is nearest to an air-drilling machine, or any other noisy tool.

Furthermore, the results of sound measurements in the operating room in L_{eq}, L_1, L_{10}, L_{50}, L_{90}, L_{99}, and MAXL show that measured levels of noise are high (Table IV), which can interfere with the operating staff's work concentration.[26–28]

Moreover, the correct machinery and tools and their maintenance are of significance in the formation of the acoustic environment of the operating room. For example, equipment (stretcher, small tables, benches, or chairs) is less noisy when it is rolled along on wheels than when it is dragged. The materials used in the manufacture of surgical equipment contribute to higher levels of noise in the operating room. Observations in the operating room reveal that wooden furniture, including the daises on which surgeons stand, is very noisy (Table V and as such should be eliminated from the operating room. The lack of central steriliza-

tion departments in most hospitals is responsible for excess noise during the presurgical time from metal sterilization kits and unsuitably packed materials and tools.[16,29]

It is also clear from the sound measurements that the noise indicators in most cases are in excess of 60 dB(A) (Table IV). This high level of noise indicates that in operating rooms conversation is hindered, normal hearing is affected and consequently there is an increased risk of accidents. Staff are forced to shout in order to be heard, which increases noise pollution.[14]

Table I shows that anaesthetists are generally sensitive to noise from conversation and that a higher proportion of nursing personnel report machinery and air conditioning as sources of noise in the operating room. In contrast, surgeons are less sensitive to these noises. The sensitivity of anaesthetists and nursing personnel to noise is to be expected because they work longer hours on a daily basis in operating rooms, while surgeons in Greece are not obliged to be in surgery every day.

The existing literature suggests that the position of every person in the room and the duration of their exposure to the noise of the room are of particular significance in how they perceive noise.[30–32] The continuous presence of anaesthetists in the operating room and the need for them to be in close proximity to the anaesthesiological equipment is likely to influence their opinion.

At another point in Table I it can be seen that a higher proportion of anaesthetists reported louder conversation such as arguments as a source of noise. This finding probably indicates a lack of coordination and problems of communication within the surgical team. In general this is a point that brings a sociological dimension to noise and requires further examination.

We can also see from Table I that a higher proportion of surgeons cited external noise as a source of noise. From Sec. II and specifically the paragraph "The study and the place where the study was conducted" it is clear that external noise is associated with poor building infrastructure and lack of organization in operating rooms. We can conclude from the findings in Table I that surgeons are more sensitive to the general working environment and conditions in the operating room than to noise. Alternatively, it might be said that for the surgeon, noise is part of the problem in the operating room and not an isolated problem.

The head surgeon in Greece has overall legal responsibility for the safety of the operation. In response to the main question of whether they are disturbed by noise in the operating room, head surgeons were more likely to answer in the affirmative than the rest of the surgical team (Table II). From the answers of the head surgeons we can surmise that the problem of noise pollution in the operating room probably makes surgery more difficult. In addition, in Table II, it can be seen that the head surgeon was affected more by external noise and conversation. It also reveals that the head surgeon is more sensitive to short periods of noise containing information. This type of noise is indicated by L_{10}, and we hypothesize that the extent of its impact is closely linked with professional role. On the other hand, the rest of the surgical team was affected more by machines and air-conditioning systems. This finding reveals first that the rest of the surgical team was more sensitive to continuous low-level noise. This type of noise is indicated by L_{90}, and we hypothesize that its impact is closely linked with a typical work environment.

In Tables I and II, it can be seen that a high proportion (up to 85%) of personnel replied in the affirmative to the questions "Is there any noise in the operating room?" and 'Do you feel that noise has a negative impact on your job?' However, a much lower proportion of personnel (48%–58%) replied "yes" to the question "Does noise in the operating room disturb you?" This inconsistency indicates that personnel recognize noise as a problem in the workplace rather than as something to which they are personally sensitive. They are effectively saying that they do not have a problem but that the operating room has a problem or that the question regarding whether noise disturbs them gives them an opportunity to demonstrate their annoyance. The discontent of personnel with the working environment in operating rooms in Greece is clear from a number of previous studies[6,33] and we may perhaps conclude that the problem of noise is experienced more strongly when this discontent exists.

C. Future work

The danger regarding high intensity, instantaneous noise, which is evident during all types of operations, must be further examined. Every high-level noise in excess of 90 dB(A) is dangerous to hearing.[29] Suction apparatus is especially dangerous as it is close to 106 dB(A). Therefore, it is necessary to study the possibility of installing automatic systems that would increase the safety of air pressure before the removal of the air-drill or other machinery. Furthermore, machinery that functions with pressurized air must be developed further, so that it can function as quietly as possible.

In this particular study, the type of noise in the operating room and the anticipated negative influence on staff require further investigation. The hypothesis that the professional role works as a factor of differentiation in the opinions of the operating room personnel is also based on differences among the opinions of head surgeons and nonsurgeons (Table II). It appears that the degree to which personnel are affected by noise is influenced by the type of work they do. For example, the observer who has the task of observing noise levels on the sound level meter for many hours is likely to perceive noise differently than surgeons who take an active part in an operation.[16] On the other hand, staff get used to the noise from machinery alarms, even when they are very loud, and therefore are not affected by them.[34,35] Nevertheless, operating room personnel should be trained in noise reduction. In terms of machine alarms, these should be adjusted to the correct sound level. In the same way, machines should be switched off after use, since their unnecessary operation adds to excess noise.

Finally, the lack of systems for interpersonal communications in all outdated operating rooms increases the noise problem in relation to conversation. However, we must insist on restrictions being laid down regarding unnecessary con-

versation in the room. The frequency of conversation and its intensity could be limited with further training and better coordination between the surgical team.[4]

D. Conclusion

In conclusion, it appears that the evaluation of this research is best achieved via a multidisciplinary approach. It is clear that improvements in the acoustic environment of operating rooms depend on many parameters such as the design and construction of buildings, the physical acoustics, organization and administration, as well as other socio-psychological factors.[2,16,36–40]

ACKNOWLEDGMENTS

The author would like to thank Dr. Vassilis Cutsuridis (Department of Computing Science and Mathematics, University of Stirling, Stirling, UK) for proofreading and editing this manuscript during its second revision.

[1]B. Hodge and J. Thompson, "Noise pollution in the operating theatre," Lancet **335**, 891–894 (1990).
[2]C. Tsiou, D. Efthymiatos, E. Theodossopoulou, P. Notis, and K. Kiriakou, "Noise sources and levels in the Evgenidion Hospital, intensive care unit," Intensive Care Med. **24**, 845–847 (1998).
[3]N. Shankar, K. Malhotra, S. Ahuja, and O. Tandon, "Noise pollution: A study of noise levels in the operation theatres of a general hospital during various surgical procedures," J. Indian Med. Assoc. **99**, 244–247 (2001).
[4]T. Katostaras, C. Tsiou, K. Katsenis, E. Kotrotsiou, E. Konstantinou, and M. Amari, "Nursing administration and the surgical team. Views on and evaluation of the operation of Greek O.R.s," ICUs Nurs. Web J., pp. 1–11 (2002) as cited in http://www.nursing.gr/issue9.html, (last viewed 12/18/2007).
[5]C. Tsiou, K. Yiannakopoulos, T. Papapolychroniou, and E. Michelinakis, "Noise pollution in the operating theatre—Comparison between orthopaedic and non-orthopaedic operation," J. Bone Joint Surg. Br. **83-B**, 184 (2001).
[6]C. Tsiou, T. Katostaras, F. Kiritsi, E. Evagelou, and E. Theodossopoulou, "Apopsis gia tin ipodomi ton xirourgion. Mia ikona gia to perivallon tou xirourgiou stin Hellada" ("Current features of operating room infrastructure: A picture of the operating room environment in Greece") (in Greek), To Vima tou Asklipiou **5**, 249–254 (2006); English translation: Vema of Asklipios **5**, 249–254 (2006) (TEI of Athens with cooperation, Ion publishing, Paper ISSN 1109–4486).
[7]D. W. Kim, H. Y. Kil, and P. F. White, "The effect of noise on the bispectral index during propofol sedation," Anesth. Analg. (Baltimore) **93**, 1170–1173 (2001).
[8]B. Griefahn, *Schlafverhalten und Gerausche* (in German) (Enke, Stuttgart, 1985).
[9]M. Topf and E. Dillon, "Noise-induced stress as a predictor of burnout in critical care nurses," Heart Lung **17**, 567–574 (1988).
[10]H. Love, "Noise exposure in the orthopaedic operating theatre: A significant health hazard," Aust. N. Z. J. Surg. **73**, 836–838 (2003).
[11]C. D. Ray and R. Levinson, "Noise pollution in the operating room: A hazard to surgeons personnel and patients," J. Spinal Disord. **5**, 485–488 (1992).
[12]K. Willett, "Noise-induced hearing loss in orthopaedic staff," J. Bone Joint Surg. Br. **73B**, 113–115 (1991).
[13]R. K. Harrison, "Hearing conservation: Implementing and evaluating a program," AAOHN J. **37**, 107–111 (1989).
[14]E. L. Kinsler, R. A. Frey, B. A. Coppens, and V. J. Sanders, *Fundamentals of Acoustics*, 3rd ed, (Wiley, Chichester, 1982).
[15]R. A. Shapiro and T. Berland, "Noise in the operating room," N. Engl. J. Med. **287**, 1236–1237 (1972).
[16]C. Tsiou, "Dierevnisi tou provlimatos tis ixoripansis sta xirourgia ton Hellinikon nosokomion" ("An investigation into the problem of noise pollution in operating theatres in Greek hospitals") (in Greek), Ph.D. dissertation, University of Athens, Athens, Greece, 1999, as cited in http://thesis.ekt.gr/12311 (verified 12 November 2007).
[17]M. Nott and P. West, "Orthopaedic theatre noise: A potential hazard to patients," Anaesthesia **58**, 784–787 (2003).
[18]G. Liang Jia, "Intraoperative monitoring of the human auditory nerve: Effects of surgical noise and contralateral sound," Ph.D. dissertation, University of Northwestern, Illinois, 1992, available by OCLC:29642856.
[19]R. Swansburg and R. Swansburg, *Introductory Management and Leadership for Nurses: An Interactive Text*, 2nd ed. (Jones and Bartlett, London, 1999), pp. 10–500.
[20]I. Busch-Vishniac, J. West, C. Barnhill, R. Hunter, D. Orellana, and R. Chivukula, "Noise levels in Johns Hopkins Hospital," J. Acoust. Soc. Am. **118**, 3629–3645 (2005).
[21]J. M. Kracht, I. J. Busch-Vishniac, and J. E. West, "Noise in the operating rooms of Johns Hopkins Hospital," J. Acoust. Soc. Am. **121**, 2673–2680 (2007).
[22]D. Efthymiatos, "O ktipogenis thorivos kai i antimetopisi tou" ("The stroke-generated noise and its confronting") (in Greek), Texnika, pp. 38–45 (January 1993), English translation: Technicals, pp. 38–45 (1993) available from Tekdotiki Publishing, Athens, Greece.
[23]D. Efthymiatos, C. Tsiou, and G. Efthymiatos, "Noise pollution in operating rooms: Levels and causes," Proceedings of the Eighth International Symposium on Theoretical Electrical Engineering, Thessaloniki, 22, 23 September 1995, pp. 494–497.
[24]S. M. Mirbod, H. Yoshida, R. Inaba, and H. Iwata, "Exposure to segmental vibration and noise in orthopaedists," Ind. Health **31**, 155–164 (1993).
[25]P. Lewis, J. Staniland, A. Cuppage, and J. M. Davies, "Operating room noise," Can. J. Anaesth. **37**, S79 (1990).
[26]"Guidelines for community noise," edited by B. Berglund, T. Lindvall, D. H. Schwela, and K. T. Goh, Technical Report, World Health Organization, Geneva, 1999.
[27]E. K. McLean and A. Tarnopolsky, "Noise, discomfort and mental health. A review of the socio-medical implications of disturbance by noise," Psychol. Med. **7**, 19–62 (1977).
[28]S. A. Stansfeld, "Noise, noise sensitivity and psychiatric disorder: Epidemiological and psychophysiological studies," Psychol. Med. Monogr. Suppl. **22**, 1–4 (1992).
[29]A. Papadaki, *Enxiridion Xirourgiou Asiptos Texniki (Manual on Sterile Techniques in the Operating Room*), (in Greek), (Argiriou, Athens, 1977), Chaps. 1–12, pp. 5–445.
[30]M. Wallace, M. Ashman, and M. Matjasko, "Hearing acuity of anesthesiologists and alarm detection," Anesthesiology **81**, 13–28 (1994).
[31]G. R. Loeb, R. B. Jones, and K. Behrman, "Recognition accuracy of current operating room alarms," Anesth. Analg. (Baltimore) **75**, 499–505 (1992).
[32]G. A. Finley and A. J. Cohen, "Perceived urgency and anaesthetist: Responses to common operating room monitor alarms," Can. J. Anaesth. **38**, 958–964 (1991).
[33]C. Tsiou, B. Balta, and T. Katostaras, "Dierevnisi diikitikon kai litourgikon adinamion ton xirourgion" ("Examination of management and functional weaknesses of operating rooms") (in Greek), 23rd Meeting of the Panhellenic Association of Nurses, Kavala, 20–23 May 1996 pp. 357–371, available from Hellenic National Association of Nurses of Greece.
[34]V. Murthy, S. Malhotra, I. Bala, and M. Raghunathan, "Detrimental effects of noise on anesthetists," Can. J. Anaesth. **42**, 608–611 (1995).
[35]P. C. Kam, A. C. Kam, and J. F. Thompson, "Noise pollution in anaesthetic and intensive care environment," Anaesthesia **49**, 982–986 (1994).
[36]R. S. Del Campo, O. C. Gutierrez, and M. Jaramillo, "Noise. Acoustic pollution in emergency rooms," Rev. Enferm **28**, 20–24 (2005).
[37]A. Schneider and J. Beibuyck, "Music in the operating room," Lancet **335**, 1407 (1990).
[38]M. A. Tijunelis, E. Fitzullivan, and S. O. Henderson, "Noise in the ED," Am. J. Emerg. Med. **23**, 332–335 (2005).
[39]T. Katostaras, C. Tsiou, E. Nula, M. Amari, K. Katsenis, and E. Konstantinou, "O.R. Nurses and nursing education: A study of operating rooms in Greece," ICUs Nurs. Web J., pp. 1–14 (2002), as cited in http://www.nursing.gr/issue10.html.
[40]A. Joseph and R. Ulrich, "Issue paper #4: Sound control for improved outcomes in healthcare settings," Technical Report, The Center for Health Design, California, as cited in http://www.healthdesign.org/research/reports/index/publicationdate/2007.php, (last viewed 12/18/2007).

Effect of background noise levels on community annoyance from aircraft noise

Changwoo Lim
School for Creative Engineering Design of Next Generation Mechanical and Aerospace Systems, School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, 151-744, Republic of Korea

Jaehwan Kim, Jiyoung Hong, and Soogab Lee[a)]
School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, 151-744, Republic of Korea

(Received 27 December 2006; accepted 15 November 2007)

A study of community annoyance caused by exposures to civil aircraft noise was carried out in 20 sites around Gimpo and Gimhae international airports to investigate the effect of background noise in terms of dose-effect relationships between aircraft noise levels and annoyance responses under real conditions. Aircraft noise levels were mainly measured using airport noise monitoring systems, B&K type 3597. Social surveys were administered to people living within 100 m of noise measurement sites. The question relating to the annoyance of aircraft noise was answered on an 11-point numerical scale. The randomly selected respondents, who were aged between 18 and 70 years, completed the questionnaire independently. In total, 753 respondents participated in social surveys. The result shows that annoyance responses in low background noise regions are much higher than those in high background noise regions, even though aircraft noise levels are the same. It can be concluded that the background noise level is one of the important factors on the estimation of community annoyance from aircraft noise exposure.
 [DOI: 10.1121/1.2821985]

PACS number(s): 43.50.Qp, 43.50.Rq, 43.50.Sr [BSF] Pages: 766–771

I. INTRODUCTION

Background noise level is an important factor on community annoyance response to aircraft noise exposure because subjective responses to aircraft noise were found to decrease with increasing background noise levels. The influence of background noise levels for annoyance responses to aircraft noise has been studied.[1–6]

Pearsons carried out the first study of background noise in 1966.[1] He studied the effects of background noise on perceived noisiness, and reported that the addition of background noise decreased the perceived noisiness of aircraft noise. Wells found that "noise complaint potential" was a function of the difference between aircraft noise and background noise as well as the background noise levels.[2] Powell and Rice found that average annoyance decreased with increasing background noise levels when the background level was held continuous over a test session.[3] Increasing the background noise from 32 to 46 dB reduced annoyance responses by an amount equivalent to 5 dB reduction in aircraft noise level. They concluded that increases in background noise levels caused decreases in subjective responses to aircraft noise. Bottom obtained annoyance scores at nine sites under the conditions of combining three different aircraft noises and three different traffic flows.[4] Traffic flows show strong relationship to traffic noise, with heavier traffic flow causing higher traffic noise. He found that mean annoyance scores were regressed against traffic flows. He concluded that the lower background noise levels were, the greater annoyance responses were at a given aircraft noise level. Johnston and Haasz analyzed annoyance judgments made by 35 subjects under different conditions of background noise levels and signal durations.[5] They also obtained that the increase of background noise levels was associated with reduced annoyance scores. Several authors also reported that background noise levels were relevant to the subjective responses to aircraft noise.[6]

Although several studies on the effects of background noise have been performed, they have mainly been carried out in laboratory situations to determine the subjective responses to aircraft noise in different background noise levels. However, laboratory situations are not real conditions. According to the result reported by Taylor *et al.*, background noise levels do not significantly affect either individual or aggregate responses to aircraft noise in real world conditions, because laboratory studies based on judgments of single flyover events may not be generalizable to the real world conditions of long term exposure to multiple events.[7] Fields also reported that ambient noise levels have no effect in community annoyance because intrusive noise levels which are high enough to be annoying are usually high enough so that they are not usually masked, even by high ambient noise levels.[8] However, in the Swiss study of response to aircraft noise, it was found that annoyance response to a given level of aircraft noise was less in neighborhoods with heavy road traffic

[a)]Author to whom correspondence should be addressed. Electronic mail: solee@snu.ac.kr

than where the road traffic was light.[9] A similar result was reported by Waters and Bottom.[10] Accordingly, a final conclusion about the effects of background noise on the assessment of community noise remains still uncertain under real situations. Therefore, the objective of this paper is to investigate the effect of background noise in terms of community reactions to residents exposed to real noise conditions like long term aircraft noise exposure, and to make sure whether these results are consistent with those reported in laboratory studies as already discussed.

II. METHODOLOGY

In assessing human responses to the aircraft noise under different background noise conditions, the authors employed a field survey that consists of physical measurements and social surveys using a questionnaire.

A. Noise measurement

1. Site selection

Field surveys were performed in 20 sites around two international airports in Korea to investigate the effect of background noise level on community reactions. Eleven sites were selected around Gimpo airport and the others were chosen around Gimhae airport. These areas were chosen, because while both airports mainly service civil aircraft operations that are fixed routes and the regular volumes of flights, their surrounding areas are exposed to similar aircraft noise levels in the whole year.

Areas near Gimpo airport are mostly urban located near by Seoul, while those near Gimhae airport are mostly rural areas with rice fields. Therefore, the two areas are clearly divided into two different background noise levels. The difference of background noise levels between these areas is about 10 dB(A). Measurement sites were mostly under the paths of the aircrafts during landing and take off, and they were also flat and free of obstacles.

Figure 1 shows field survey sites around Gimpo and Gimhae international airports. Most of the houses in the field survey sites around Gimpo airport are apartment buildings built out of ferroconcretes and the majority of houses in those around Gimhae airport are detached houses of bricks.

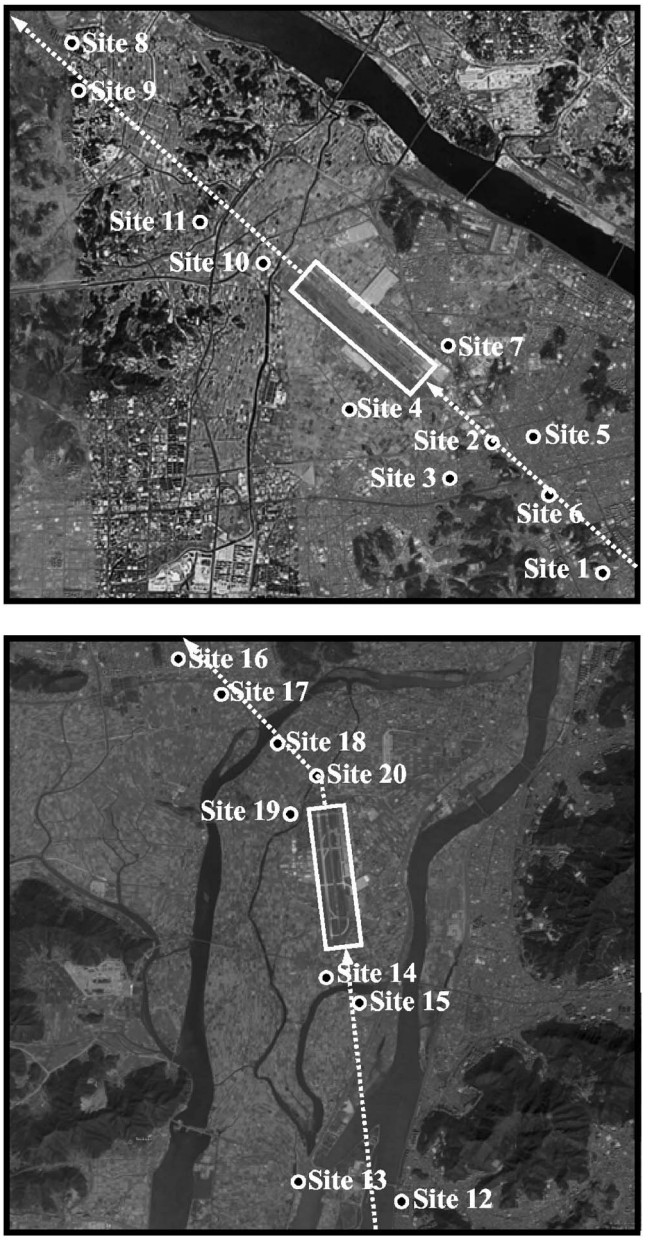


FIG. 1. Field survey sites around Gimpo (above) and Gimhae (below) international airports.

2. Noise measurement

Noise levels were measured at the two international airports with different volumes of aircraft operations. The average number of flights in Gimpo and Gimhae airports is 160 and 80 a day, respectively.

The measurements of aircraft noise were carried out not only with airport noise monitoring systems, but also with portable precision sound level meters at the field survey sites. The aircraft noise levels of 16 sites were measured automatically using airport noise monitoring systems, B&K type 3597. The equipment, managed by the Ministry of Environment in Korea, were mounted on the rooftops of houses to avoid any obtrusions caused by obstacles between the aircraft and the receiver. The others were measured with portable precision sound level meters, B&K type 2238. They were also mounted on the rooftops with a tripod. Microphones were positioned at a height of 1.5 m above the flat, and at least 1 m from any other reflecting surfaces.

It was necessary to carry out extensive measurements to obtain more precisely calculated results of aircraft noise levels. The measurements of airport noise monitoring systems were performed around the clock every day, from January to June of 2004.

WECPNL (weighted equivalent continuous perceived noise level) was used to assess the relationship between aircraft noise levels and annoyance responses, and to investigate the effect of background noise on annoyance reactions. The international civil aviation organization (ICAO) recommended the use of WECPNL to measure and evaluate the aircraft noise at first.[11] In Korea, however, WECPNL was modified from the WECPNL recommended by ICAO to sim-

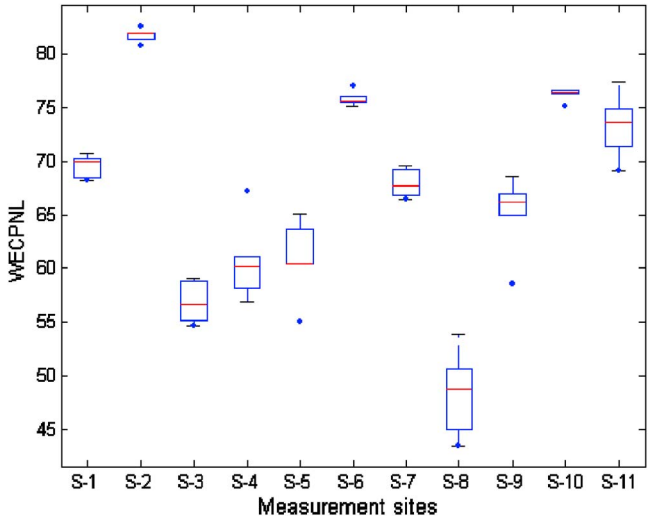


FIG. 2. (Color online) Box plots showing the distribution of aircraft noise levels in field survey sites around Gimpo airport (see Ref. 11).

plify the measurement and evaluation of the aircraft noise. The WECPNL used in Korea (abbr. WECPNL in the following discussion) is defined as follows:[12]

$$\mathrm{WECPNL} = \bar{L}_A + 10\log(N_2 + 3N_3 + 10(N_1 + N_4)) - 27, \quad (1)$$

where, $\bar{L}_A$ denotes the energy mean of all maximum aircraft noise level during a day. N_2 and N_3 are the number of events during the daytime from 07:00 to 19:00 and the nighttime from 19:00 to 22:00. N_1 and N_4 are the number of events during midnight from 00:00 to 07:00 and late nighttime from 22:00 to 24:00, respectively.

Figure 2 indicates the distributions of aircraft noise levels in field survey sites around Gimpo airport.[13] The box plots in Fig. 2 show median values (horizontal lines), interquartile ranges (boxes) that consist of lower quartile (Q1) and upper quartile (Q3), the largest and smallest observations (whiskers) and outliers. Outliers are defined as lying exterior data of 1.5 box lengths from the edge of the box. As shown in this figure, there are almost no outliers, and the interquartile ranges are small. In a word, the exposure levels of aircraft noise in each field survey area are nearly similar during the whole year.

B. Social survey

Subjective responses to aircraft noise were measured by means of a social survey using a questionnaire. The survey was performed in order to investigate the individual's attitude and opinion in regards to different aspects of aircraft noise, and it was administered to residents living within about 100 m of the noise measurement sites. Therefore, it can be assumed residents living in a given area were exposed to similar noise levels with negligible differences.

Questionnaires were comprised of questions relating to the assessments of aircraft noise, as well as some general questions of residents, even if they do not relate to noise. Questions were arranged in three basic sections. The first section sought to obtain demographic data, the second asked questions about nuisance perception from aircraft noise, and the third dealt with health-related symptom questions. Therefore, the questionnaire contained demographic questions, degree of noise annoyance, interferences with daily activities, psychological and physiological health-related symptoms, and reaction to aircraft noise. In order to assess the annoyance responses to aircraft noise, specifically, people were asked questions like "how much were you bothered or annoyed by the aircraft noise, while staying at home, in the last 12 months.[14] The respondents then used an 11 category response ranging from 0 (not annoyed at all) to 10 (extremely annoyed). The 11-point numerical scale is shown in Table I. The 11-point numerical scale was chosen over the shorter 7 or 9-point numeric scales with the assumption that respondents are more cognitively familiar with the 0 to 10 scaling.[15]

To avoid any bias in opinion, the surveys were not introduced to the interviewees in advance and the respondents were randomly selected from residents near the measurement sites based on simple random sampling method. Questionnaires were distributed in person and the questionnaires were completed independently while researchers waited. Each questionnaire took about 20 min to complete and the social surveys were carried out within a given period of the noise measurement at each site. 63.5% of the randomly selected respondents participated in this survey, resulting in a total of 753 respondents for the analysis of exposure-effect relationships between aircraft noise levels and annoyance responses.

III. RESULTS

Thirty three percent of the respondents were male and 67% were female. The ages of the respondents exhibit a wide range: younger than 20 years (6%), 20–40 (36%), 40–60 (38%) and older than 60 years (20%). Most of respondents were female (67%) and were married (80%). These results were due to the nature of the Korean culture where most women become housewives after marriage. The duration of residency of the respondents was as follows: less than 1 year (7%), 1–3 years (17%), 3–10 years (35%), 10–30 years (33%) and more than 30 years (8%). Regarding community response to aircraft noise 70% of the respondents reported that they were worried by aircraft noise. About 68% of the respondents have been surprised at very loud and unexpected aircraft noise. About 35% of them consider that the aircraft

TABLE I. Numerical scale.

0	1	2	3	4	5	6	7	8	9	10
Not annoyed at all										Extremely annoyed

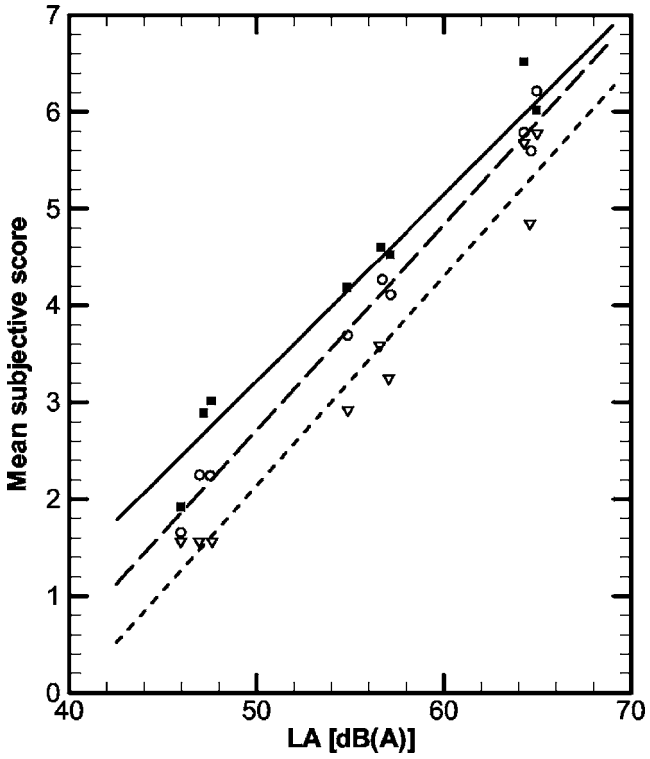


FIG. 3. Regressions of mean subjective scores on L_A for continuous background noise. (Mean background noise level in dB(A): ■—, 32.3;○---, 37.1;▽……., 46.4) (see Ref. 3).

noise is not good for their own health. They also raised complaints that the aircraft noise causes insomnia, nervousness and indigestion.

Annoyance responses to civil aircraft noise were elicited by means of an 11-point numerical scale. Under the definition of the annoyance scale, the term "highly annoyed" is defined as the upper 27–28% of the annoyance scale. Given the goals of the study, the relationships between the civil aircraft noise and the percentage of respondents who were "highly annoyed" according to different background noise levels were compared, and also made sure whether the results of this study were consistent with those reported in laboratory studies like Fig. 3 in showing that subject response to aircraft noise was significantly affected by variations in background conditions.[3]

To assess the effects of noise on health, the percentage of respondents who felt highly annoyed (*%HA*) was selected as the indicator of noise annoyance in many countries.[16–18] The World Health Organization (WHO) also has recommended *%HA* as one of the environmental health indicators

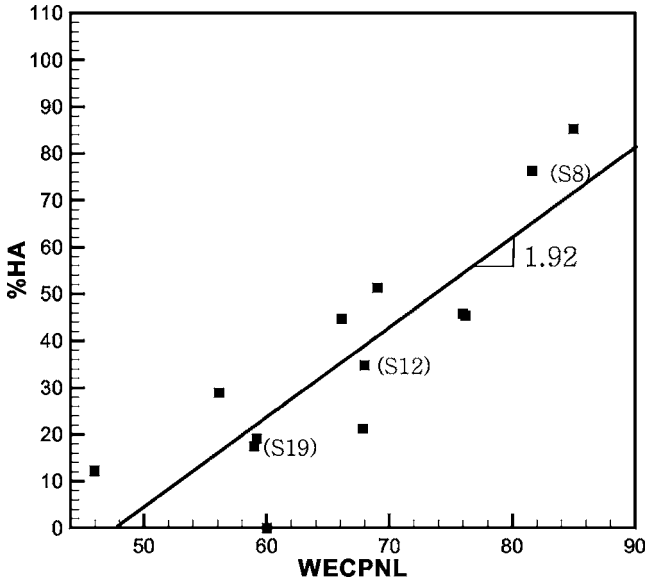


FIG. 4. *%HA* with respect to WECPNL in relatively high background noise levels.

to support the effects of environmental noise on health.[19] Therefore, *%HA* was used to assess the effects of background noise in terms of dose-effect relationships between aircraft noise levels and annoyance responses, and WECPNL was used also as the physical descriptor of aircraft noise in this study, because that is used to evaluate the aircraft noise in Korea.

In order to investigate the influence of background noise, field survey data were divided into two groups based on background noise levels. Groups 1 and 2 are field survey data in relatively low and high background noise levels, respectively. Background noise levels, excluding aircraft noises, were measured. The descriptor for background noise is A-weighted Equivalent Sound Level which is cumulated 1 h exposure, $L_{Aeq,1\ h}$. The information on background noise for each group is shown in Table II. As shown in this table, the mean background noise levels of group 1 (low background noise group) and group 2 (high background noise group) are 42 and 55.5 dB(A), and the difference in the mean values between the two groups is about 13 dB(A).

Figures 4 and figure 5 show *%HA* with respect to WECPNL in relatively high and low background noise levels, respectively. These figures show the rising tendency of *%HA* according to WECPNL, but the rising tendency of *%HA* in Fig. 5 is steeper than that in Fig. 4. Namely, the subjective responses to aircraft noise are different at the

TABLE II. Information about background noise levels.

	L_{Aeq}(max)[a]	L_{Aeq}(min)[b]	L_{Aeq}(mean)[c]	Standard deviation
Group 1 (Low background noise)	45	39	42	2
Group 2 (High background noise)	60	51	55.5	2.9

[a]The greatest noise levels which were calculated except aircraft noise in the group.
[b]The lowest noise levels which were calculated except aircraft noise in the group.
[c]The mean background noise level in the group.

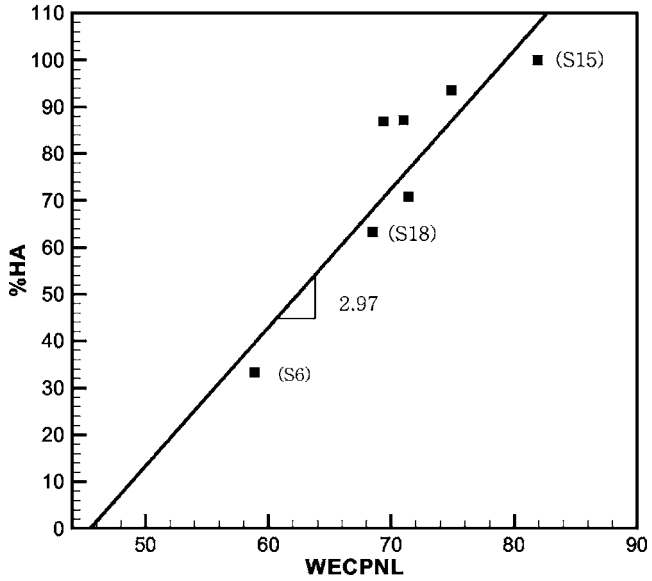


FIG. 5. *%HA* with respect to WECPNL in relatively low background noise levels.

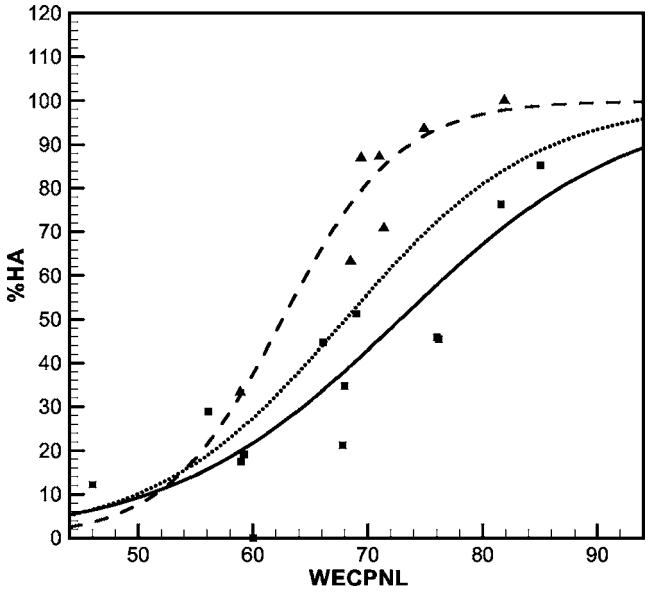


FIG. 6. Comparison between *%HA* prediction curve of civil aircraft noise according to background noise levels. (▲ and ■, field survey data in low and high background noise levels, respectively; ---, *%HA* prediction curve based on field survey data in low background noise group, N=487; —, *%HA* prediction curve based on field survey data in high background noise group, N=212).

same values of WECPNL. The difference increases with the noise level. Table III shows the comparison of measurement data in some field survey sites. As shown in this table, the values of WECPNL are similar at each category C1, C2 and C3, but *%HA* values are different. In other words, there is almost no difference in WECPNL between the site 18 and site 12 in category C2, while there is a difference in *%HA* of about 28% between the two sites. Categories C1 and C3 are just the same. Because the sites 6, 15, 18 are a rural area with rice fields while the sites 8, 12, 19 are an urban area nearby Seoul, the background noise levels of sites 8, 12, 19 are higher than those of sites 6, 15, 18. In view of the results of this table, background noise levels have influence on annoyance responses.

Figure 6 shows the annoyance responses of residents exposed to civil aircraft noise with respect to different background noise levels in WECPNL. The level of aircraft noise exposure ranges from 46 to 85. Triangle and square spots are field survey data showing *%HA* as a function of WECPNL in group 1 and group 2, respectively. The dashed and solid lines are *%HA* prediction curves that are based on field survey data in group 1 and group 2, respectively. As shown in this figure, the annoyance responses of group 1 are much higher than those of group 2 at the same noise levels. The average difference in annoyance responses between group 1 and group 2 is about 35% even though noise levels are the same. The difference increases with the increase in the level. Before the results are accepted, however, they must be verified whether or not the difference of annoyance responses between group 1 and group 2 is valid result, or a simple standard error. A t test is used to assess whether the means of two groups are statistically significant or not. The t test was performed by SAS ver. 8.2[20] to compare the means of two groups. The results of the t test are shown in Table IV. As shown in this table, the variances of the two groups are significantly different. When significant difference is observed in the variances of the two groups, Satterthwaite's estimate is performed.[21] Therefore, the t value is −8.23 and the significance level of the t value is less than 0.0001 where the variances of the two groups are statistically different ($P<0.05$). Based on the results, therefore, the null hypothesis (H_0), the means of the two groups are the same, is rejected. It is shown statistically that annoyance responses between group 1 and group 2 are significantly different. This means that the difference of annoyance responses between group 1 and group 2 is not from standard errors, but from the properties of the

TABLE III. Comparison of measurement data in some field survey sites.

Category	Measurement site	*WECPNL*	*%HA*	Group
C1	Site 3	58.9	33.3	1[a]
	Site 5	59	17.5	2[b]
C2	Site 7	68.5	63.3	1[a]
	Site 10	68	34.8	2[b]
C3	Site 18	81.9	100	1[a]
	Site 19	81.6	76.3	2[b]

[a]Relatively low background noise level group.
[b]Relatively high background noise level group.

TABLE IV. Summary of t test about annoyance response to aircraft noise exposure.

	Method	Variances	$d.f.$	t value	$\Pr>\|t\|$
	Pooled	Equal	466	−7.99	<0.0001
t test	Satterthwaite	Unequal	458	−8.23	<0.0001
Equality of variance	Method	Num $d.f.$	Den $d.f.$	F value	$\Pr>F$
	Folded F	255	211	1.91	0.0001

groups. The results of the statistical analysis show that there is significant difference in the annoyance responses to the same levels of aircraft noise between areas of high and low background noise levels. In other words, it seems that the annoyance response to a given level of aircraft noise is less in neighborhoods with high background noise than those with low background noise.

Consequently, the increase of background noise levels causes a decrease of subjective responses to aircraft noise under real conditions, which is also reported in many laboratory studies. Therefore, it can be concluded that background noise level is an important factor on community annoyance from aircraft noise exposure not only in the laboratory, but also in the field.

IV. CONCLUSIONS

Several studies have been conducted to investigate the difference of subjective responses to aircraft noise in relations to background noise levels. However, they have been mainly carried out in laboratory situations. Since laboratory situations are not real conditions and a final conclusion about the effects of background noise on the assessment of community noise is still premature under real situations, the results of these effects remain uncertain under real conditions. Therefore, the study of community annoyance caused by civil aircraft noise exposures was carried out in 20 sites around Gimpo and Gimhae international airports to investigate the effects of background noise in terms of dose-effect relationships between aircraft noise levels and annoyance responses under real conditions. To assess the dose responses to aircraft noise, the percentage of respondents who felt highly annoyed ($\%HA$) and WECPNL, the aircraft noise index in Korea, were used in this study.

The result shows that annoyance responses in low background noise regions are much higher than those in high background noise regions, even though aircraft noise levels are the same. The difference in annoyance responses between the two regions occurred up to about 45% due to background noise levels. Therefore, it can be concluded that annoyance responses to intrusive noise, such as aircraft noise, are not independent of background noise levels, and background noise level plays an important role in the estimation of community annoyance from aircraft noise exposure.

ACKNOWLEDGMENTS

This work was supported by the Core Environmental Technology Development Project for Next Generation in Korea Institute of Environmental Science and Technology, and Brain Korea 21 Project in 2006

[1] K. S. Pearsons, "The effects of duration and background noise level on perceived noisiness," Federal Aviation Administration Technical Report No. ADS-78 (1966).
[2] R. J. Wells, "Noise complaint potential, ambient noise versus intrusive noise," *Seventh International Congress on Acoustics*, Budapest (1971).
[3] C. A. Powell and C. G. Rice, "Judgments of aircraft noise in a traffic noise background," J. Sound Vib. **38**, 39–50 (1975).
[4] C. G. Bottom, "A social survey into annoyance caused by the interaction of aircraft noise and traffic noise," J. Sound Vib. **19**, 473–476 (1971).
[5] G. W. Johnston and A. A. Haasz, "The influence of background noise level and signal duration on the judged annoyance of aircraft noise," University of Toronto Institute for Aerospace Studies Report No. 228 (1976).
[6] D. G. Nagel, J. E. Parnell, and G. J. Parry, "The effects of background noise upon perceived noisiness," Federal Aviation Administration Technical Report No. DS-67-22 (1967).
[7] S. M. Taylor, F. L. Hall, and S. E. Birnie, "Effect of background levels on community responses to aircraft noise," J. Sound Vib. **71**, 261–270 (1980).
[8] J. M. Fields, "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763 (1993).
[9] E. Grandjean, P. Graf, A. Lauber, H. P. Meier, and R. Muller, "A survey of aircraft noise in Switzerland," *Proceeding of the International Congress on Noise as a Public Health Problem*, Dubrovnik, Yugoslavia, 13–18 May (1973).
[10] D. M. Waters and C. G. Bottom, "The influence of background noise on disturbance due to aircraft," *Proceedings of the VIIth International Congress on Acoustics*, Vol. **4**, pp. 521–524, Budapest (1971).
[11] ICAO, International Standards and Recommended Practices Environmental Protection Annex 16, International Civil Aviation Organization (1971).
[12] Ministry of Environment in Korea, Noise and Vibration control act, Notice No. 2000-31.
[13] C. Lim, J. Kim, J. Hong, S. Lee, and S. Lee, "The relationship between civil aircraft noise and community annoyance in Korea," J. Sound Vib. **299**, 575–586 (2007).
[14] C. Lim and S. Lee, "Questionnaire on environmental noise: The core set," Center for Environmental Noise and Vibration Research (2003).
[15] International Standard Organization (ISO), "Acoustics—Assessment of noise annoyance by means of social and socio-acoustic surveys," ISO/TS 15666 (2003).
[16] T. H. J. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405 (1978).
[17] H. M. E. Miedema and H. Vos, "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[18] L. S. Finegold, C. S. Harris, and H. E. von Gierke, "Community annoyance and sleep disturbance: Updated criteria for assessing the impacts of general transportation noise on people," Noise Control Eng. J. **42**, 25–30 (1994).
[19] WHO, Environmental Health Indicators: Development of a methodology for WHO European Region, World Health Organization (2000), Copenhagen, Denmark.
[20] SAS Institute Incorporated, *SAS User's Guide: Statistics. Version 8.2* (SAS Institute Inc., Cary, NC, 2003).
[21] F. Satterthwaite, "An approximate distribution of estimates of variance components," Biometrics 110–114 (1946).

Effects of social, demographical and behavioral factors on the sound level evaluation in urban open spaces

Lei Yu and Jian Kang[a)]
School of Architecture, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom

(Received 29 May 2007; accepted 10 November 2007)

The aim of this study is to analyze the effects of social, demographical and behavioral factors as well as long-term sound experience on the subjective evaluation of sound level in urban open public spaces. This is based on a series of large scale surveys in 19 urban open spaces in Europe and China. The results suggest that the effects of social/demographical factors, including age, gender, occupation, education and residential status, on the sound level evaluation are generally insignificant, although occupation and education are two related factors and both correlate to the sound level evaluation more than other factors. The effects of some behavioral factors, including wearing earphones, reading/writing and moving activities, are also insignificant on the sound level evaluation, but the watching behavior is highly related to the sound level evaluation. Compared to the social, demographical and behavioral factors, the long-term sound experience, i.e. the acoustic environment at home, significantly affect the sound level evaluation in urban open spaces. It is important to note that between the social/demographical factors, there are generally significant correlations, although the correlation coefficients may not be high. It is also noted that there are considerable variations between different urban open spaces.
 [DOI: 10.1121/1.2821955]

PACS number(s): 43.50.Qp, 43.50.Rq, 43.50.Sr [BSF] Pages: 772–783

I. INTRODUCTION

The evaluation of sound is a complex system and it is related to a number of disciplines including acoustics, physiology, sociology, psychology and statistics. The sound level is an important factor for the subjective evaluation of an acoustic environment. Relationships between annoyance and noise exposure have been intensively studied.[1–6] However, it has been demonstrated in many studies that correlations between noise annoyance and the acoustical/physical factors are often not high.[7–10]

The effects of various social/demographical/economical factors on the sound evaluation have been broadly studied. There are varied results regarding the age effect, whereas most studies seem to suggest that the effect of gender is not important.[11–15] In terms of the education level, some studies show no significant effect on annoyance,[12,13] whereas other studies seem to suggest that people are slightly more annoyed with a higher education level.[14] The effects of a number of other factors on the sound evaluation have also been studied, including income and the economical status,[12,13,16,17] general state of health,[17] marital status,[18] house size/type and the family size,[14,17] length of residence,[12,13] time spent at home,[19] and the type of occupancy.[12–14,16] Noise experience, including the exposure to noise at the place of work and over time, could affect residential noise annoyance[17] as well as sleep process.[20] Behavior and habit is another important aspect which could affect annoyance. This includes, for example, opening and closing windows,[10,17] using sleeping pills, using balconies or gardens, having their home sound insulated, and frequently leaving for weekends.[21]

While the effects of a number of factors on the sound evaluation have been explored, previous studies have been focused on residential areas. The sound evaluation in urban open public spaces, which are important components of a city, has been rarely studied. Since reducing sound level does not always lead to a better soundscape quality in urban open public spaces, especially when Leq_{90} is below a certain value, say 70 dBA,[22] it is important to consider the effects of various other factors including social, demographical and behavioral factors as well as long-term sound experience.[23]

The aim of the study is therefore to analyze the effects of those factors on the subjective evaluation of sound level in urban open public spaces, based on a series of large scale field surveys in Europe and China.[22] Following a brief introduction of the survey methodology, this paper systematically explores the relationships between various social, demographical and behavioral factors and their effects on the sound level evaluation.

II. METHODOLOGY

In Europe, 14 urban open spaces were selected, located in seven cities of five countries, Germany, Greece, Italy, Switzerland and the UK. Two case study sites were selected in each city. In China the surveys were carried out in Beijing and Shanghai, with two and three case study sites, respectively. The 19 case study sites were: 1: Germany Kassel, Bahnhofsplatz; 2: Germany Kassel, Florentiner; 3: Greece Athens, Karaiskaki; 4: Greece Athens, Seashore; 5: Greece Thessaloniki, Kritis; 6: Greece Thessaloniki, Makedonomahon; 7: Italy Milan, IV Novembre; 8: Italy Milan, Piazza

a)Author to whom correspondence should be addressed. Electronic mail: j.kang@sheffield.ac.uk

TABLE I. Overview of the case study sites.

		Case study sites			
Country	City	Code	Site name	Location	Number of interviewees
Germany	Kassel	1	Bahnhofsplatz	Railway station	418
		2	Florentiner	Tourist location	406
Greece	Athens	3	Karaiskaki	City center	655
		4	Seashore	Tourist location	848
	Thessaloniki	5	Kritis	Residential	777
		6	Makedonomahon	City center	1037
Italy	Milan	7	IV Novembre	City center	574
		8	Piazza Petazzi	Residential	599
Switzerland	Fribourg	9	Jardin de Perolles	Residential	888
		10	Place de la Gare	Railway station	1041
UK	Cambridge	11	All Saint's Garden	Tourist location	459
		12	Silver Street	Tourist location	489
	Sheffield	13	Barkers Pool	City center	499
		14	Peace Gardens	City center	510
China	Beijing	15	Chang Chun Yuan Square	Residential	307
		16	Xi Dang Square	City center	304
	Shanghai	17	Century Square	Tourist location	62
		18	Nanjing Road Square	City center	79
		19	Xu Jia Hui Park	Residential	79

Petazzi; 9: Switzerland Fribourg, Jardin de Perolles; 10: Switzerland Fribourg, Place de la Gare; 11: UK Cambridge, All Saint's Garden; 12: UK Cambridge, Silver Street; 13: UK Sheffield, Barkers Pool; 14: UK Sheffield, Peace Gardens; 15: China Beijing, Chang Chun Yuan Square; 16: China Beijing, Xi Dan Square; 17: China Shanghai, Century Square; 18: China Shanghai, Nanjing Road Square; and 19: China Shanghai, Xu Jia Hui Park. The case study sites are summarized in Table I.

Among the 19 case study sites, there was a wide range of variation in terms of their physical conditions and urban morphology, as well as the social, demographical and behavioral factors and long-term sound experience of the users. It can be seen in Table I that five sites were in residential zones, including sites 5, 8, 9, 15 and 19. The other 14 sites were located in various types of general public areas, including city centers (sites 3, 6, 7, 13, 14, 16 and 18), tourist locations (sites 2, 4, 11, 12 and 17), and railway stations (sites 1 and 10).

In terms of soundscape, traffic noise existed at nearly all the case study sites, although less so at some sites, such as Beijing Chang Chun Yuan Square (site 15) and Shanghai Century Square (site 17). Six case study sites were featured by water sound, including Kassel Bahnhofsplatz (site 1), Milan IV Novembre (site 7), Cambridge Silver Street (site 12), Sheffield Peace Gardens (site 14), Shanghai Nanjing Road Square (site 18), and Shanghai Xu Jia Hui Park (site 19). A number of other unique sound elements contributed to the soundscapes of several case study sites, such as music in the Sheffield Barkers Pool (site 13), church bells in the Milan Piazza Petazzi (site 8), construction/demolition sounds in the Thessaloniki Kritis (site 5), Thessaloniki Makedonomahon (site 6), and Sheffield Peace Gardens (site 14).[15] Other sounds included people's chatting, children's shouting, and sounds from sport activities and foot steps.

The questionnaire was initially developed in English, and then translated into other languages. Interviewees were asked to evaluate the sound environment of the sites and their homes in terms of sound level (not necessarily acoustic comfort), where five linear scales were used: −2, very quiet; −1, quiet; 0, neither quiet nor noisy; 1, noisy; and 2, very noisy. In the surveys in China and Sheffield, more detailed questions were added into the questionnaire. Moreover, the questionnaire was not only designed for a soundscape survey, but also as an enquiry relating to general environmental issues including thermal, lighting and visual aspects, which was important to avoid any possible bias towards the acoustic aspect. The sound pressure level measurement was carried out, either when an interviewee filled in the questionnaire, or immediately after the interview.

In total, more than 10,000 interviews were made. The interviewees, who were mostly the users of the public space rather than passersby, were selected randomly. Based on the survey results, a database was compiled using statistical software SPSS,[24] where the interviewees were assigned into eight categories in terms of age: <12, 13–17, 18–24, 25–34, 35–44, 45–54, 55–64, and 65; three categories in terms of occupation: 1- students, 2- working persons, and 3- others (including pensioners, unemployed and housekeepers); three categories in terms of education level: 1- primary, 2- secondary, and 3- higher; and two categories in terms of residential status: local and non-local. Interviewees' behaviors were also observed, including wearing earphones, reading/writing, watching, and moving-related activities. In this paper, the influence of moving-related activities on the sound level evaluation is studied in two ways: one is using five categories according to energy consumption, namely 1- sitting, 2- standing, 3- walking, 4- playing with children, 5-sport; and the other is simply categorizing various activities as moving

TABLE II. Pearson/Spearman correlation coefficients between age and occupation and between age and education, with significance levels (2 tailed); as well as mean differences in terms of age between males and females and between local and non-local residents, with significance levels (t test, 2 tailed).

	Correlation/Significance				Mean difference/Significance (male - female; local - nonlocal)	
	Age/Occupation		Age/Education			
Site	Pearson	Spearman	Pearson	Spearman	Age/Gender	Age/Residence
1	0.74/0.00(**)	0.75/0.00(**)	0.13/0.01(**)	0.12/0.01 (*)	0.09/0.17	−0.02/0.88
2	0.69/0.00(**)	0.68/0.00(**)	0.26/0.00(**)	0.27/0.00(**)	0.25/0.10	−0.19/0.21
3	0.67/0.00(**)	0.61/0.00(**)	−0.29/0.00(**)	−0.26/0.00(**)	0.56/0.00 (**)	−0.45/0.00(**)
4	0.65/0.00(**)	0.61/0.00(**)	−0.30/0.00(**)	−0.27/0.00(**)	0.35/0.00 (**)	−0.40/0.00(**)
5	0.56/0.00(**)	0.55/0.00(**)	−0.42/0.00(**)	−0.44/0.00(**)	0.49/0.00 (**)	−0.90/0.00(**)
6	0.66/0.00(**)	0.68/0.00(**)	−0.40/0.00(**)	−0.38/0.00(**)	−0.03/0.82	−0.80/0.00(**)
7	0.74/0.00(**)	0.75/0.00(**)	−0.33/0.00(**)	−0.30/0.00(**)	0.73/0.00 (**)	−0.45/0.00(**)
8	0.74/0.00(**)	0.75/0.00(**)	−0.40/0.00(**)	−0.41/0.00(**)	0.26/0.10	−0.61/0.01(**)
9	0.72/0.00(**)	0.72/0.00(**)	0.01/0.80	0.02/0.59	−0.33/0.01(**)	−0.46/0.00(**)
10	0.67/0.00(**)	0.69/0.00(**)	−0.05/0.13	−0.01/0.88	0.09/0.42	−0.02/0.82
11	0.50/0.00(**)	0.48/0.00(**)	0.11/0.02 (*)	0.14/0.00(**)	0.32/0.02 (*)	0.53/0.00 (**)
12	0.65/0.00(**)	0.63/0.00(**)	−0.13/0.00(**)	−0.10/0.04 (*)	0.20/0.17	0.47/0.01 (**)
13	0.65/0.00(**)	0.60/0.00(**)	−0.01/0.87	0.09/0.04 (*)	−0.21/0.19	−0.71/0.00(**)
14	0.71/0.00(**)	0.70/0.00(**)	−0.12/0.01(**)	−0.07/0.10	−0.04/0.80	−0.10/0.59
15	−0.15/0.01(**)	−0.10/0.10	−0.05/0.37	0.02/0.69	−0.29/0.08 (*)	−1.36/0.00(**)
16	0.08/0.17	0.12/0.04 (*)	0.31/0.00(**)	0.34/0.00(**)	0.16/0.10	0.35/0.00 (**)
17	0.72/0.00(**)	0.75/0.00(**)	0.57/0.00(**)	0.56/0.00(**)	−0.74/0.04 (*)	0.18/0.68
18	0.62/0.00(**)	0.53/0.00(**)	0.22/0.06	0.34/0.00(**)	0.41/0.26	1.04/0.01 (**)
19	0.56/0.00(**)	0.56/0.00(**)	0.31/0.01(**)	0.35/0.00(**)	0.05/0.90	0.89/0.04 (*)

1: Germany Kassel, Bahnhofsplatz; 2: Germany Kassel, Florentiner; 3: Greece Athens, Karaiskaki; 4: Greece Athens, Seashore; 5: Greece, Thessaloniki, Kritis; 6: Greece Thessaloniki, Makedonomahon; 7: Italy Milan, IV Novembre; 8: Italy Milan, Piazza Petazzi; 9: Switzerland Fribourg, Jardin de Perolles; 10: Switzerland Fribourg, Place de la Gare; 11: UK Cambridge, All Saint's Garden; 12: UK Cambridge, Silver Street; 13: UK Sheffield, Barkers Pool; 14: UK Sheffield, Peace Gardens; 15: China Beijing, Chang Chun Yuan Square; 16: China Beijing, Xi Dan Square; 17: China Shanghai, Century Square; 18: China Shanghai, Nanjing Road Square; and 19: China Shanghai, Xu Jia Hui Park.

(walking, playing with children, and sport) and non-moving (sitting and standing).

It is noted that in this paper, the analysis is more holistic, aiming at examining general sound level evaluation based on a wide range of topographic conditions in urban open spaces and various users' background, rather than focusing on individual case study sites.

III. RELATIONSHIPS AMONG SOCIAL/ DEMOGRAPHICAL FACTORS

As the subjective evaluation of sound level is a complex system, relating to a number of social and demographical factors, it is essential to study the correlations among those factors before examining their effects on the sound level evaluation.

The relationships between age and occupation as well as between age and education at the 19 case study sites are shown in Table II, in terms of the correlation coefficient r and the significance level, where both Pearson and Spearman correlations are shown. It is noted that in this paper, marks * and ** indicate significant difference or correlation, with * representing $p \leqslant 0.05$ and ** representing $p \leqslant 0.01$. From Table II it can be seen that at most case study sites there is generally a significant correlation between age and occupation as well as between age and education. The correlation coefficients between age and occupation are mostly rather high, around 0.5–0.7, and are predominately positive. Between age and education the correlation coefficients are relatively low, typically around 0.3–0.4, and there are both positive and negative correlations.

In Table II both Pearson and Spearman correlations are shown, for parametric and non-parametric measures of correlations, respectively. It can be seen that the results are very similar. This is also revealed in other analyses in this paper and thus, in most tables only Pearson correlations are shown.

Table II also shows the difference between males and females as well as between local and non-local residents in terms of age, based on independent samples t test. Between males and females, the difference reaches a significant level at eight case study sites, and between local and non-local residents significant differences exist at 14 case study sites.

The differences between males and females in terms of occupation, education and residential status are shown in Table III, again based on independent samples t test. Significant differences between males and females in terms of occupation are found at seven case study sites. At the Century Square (site 17), the mean difference reaches 0.56. In terms of education, only five case study sites (sites 5, 7, 10, 11, 17) show significant differences between males and females and the mean differences are rather low, suggesting that gender and education are generally unrelated factors in these urban open spaces. Similarly, in terms of residential status, only four case sites (sites 9, 12, 15, 16) show significant differences between males and females while the mean difference

TABLE III. Mean differences in terms of gender between different occupations, between different education levels, and between local and non-local residents, with significance levels (t test, 2 tailed).

	Mean difference/Significance (male - female)		
Site	Gender/Occupation	Gender/Education	Gender/Residence
1	0.07/0.51	0.02/0.76	−0.04/0.41
2	0.03/0.09	0.07/0.29	0.09/0.08
3	−0.13/0.01(**)	0.09/0.09	0.00/0.10
4	−0.11/0.01(**)	0.03/0.57	−0.03/0.31
5	−0.08/0.16	−0.14/0.01 (*)	−0.06/0.09
6	−0.13/0.01(**)	−0.04/0.37	−0.01/0.66
7	0.16/0.01 (**)	−0.20/0.00(**)	0.03/0.51
8	−0.06/0.36	−0.04/0.48	−0.03/0.30
9	−0.17/0.00(**)	0.04/0.43	−0.08/0.01(**)
10	0.05/0.25	0.08/0.04 (*)	−0.02/0.50
11	0.05/0.48	0.13/0.01 (**)	0.01/0.84
12	−0.00/0.98	0.08/0.10	0.14/0.00 (**)
13	−0.01/0.86	−0.04/0.48	−0.02/0.50
14	0.11/0.12	0.05/0.30	0.02/0.59
15	0.28/0.00 (**)	−0.05/0.43	−0.15/0.01(**)
16	0.04/0.40	0.04/0.51	−0.13/0.03 (*)
17	−0.56/0.00(**)	−0.38/0.01(**)	−0.09/0.47
18	0.03/0.77	0.17/0.21	−0.02/0.86
19	0.08/0.61	−0.09/0.53	−0.20/0.09

is usually less than 0.15, which indicates that gender and residential status can be generally regarded as unrelated variables in the study.

Table IV shows the relationships between occupation and education, as well as the differences between local and non-local residents in terms of occupation and education. It can be seen that between occupation and education, the relationships are statistically significant (Pearson) at 14 case study sites (sites 2–11, 13, 14, 16, 17), with correlation coefficients generally around 0.3, either positive or negative. In terms of occupation, significant differences exist between local and non-local residents at nine case study sites (sites 3–9,

TABLE IV. Pearson/Spearman correlation coefficients between occupation and education, with significance levels (2 tailed); as well as mean differences in terms of residential status between different occupations and between different education levels, with significance levels (t test, 2 tailed).

	Correlation/Significance		Mean difference/Significance (local - nonlocal)	
	Occupation/Education			
Site	Pearson	Spearman	Residence/Occupation	Residence/Education
1	0.07/0.14	0.07/0.17	0.00/0.10	0.04/0.56
2	0.16/0.00 (**)	0.17/0.00 (**)	0.09/0.19	−0.12/0.08
3	−0.32/0.00(**)	−0.33/0.00(**)	−0.23/0.00(**)	0.22/0.00 (**)
4	−0.31/0.00(**)	−0.31/0.00(**)	−0.10/0.03 (*)	−0.01/0.86
5	−0.43/0.00(**)	−0.44/0.00(**)	−0.61/0.00(**)	0.46/0.00 (**)
6	−0.41/0.00(**)	−0.41/0.00(**)	−0.32/0.00(**)	0.37/0.00 (**)
7	−0.35/0.00(**)	−0.35/0.00(**)	−0.21/0.00(**)	0.40/0.00 (**)
8	−0.45/0.00(**)	−0.47/0.00(**)	−0.21/0.04 (*)	0.40/0.00 (**)
9	−0.08/0.02 (*)	−0.09/0.01(**)	−0.22/0.00(**)	0.08/0.13
10	−0.20/0.00(**)	−0.18/0.00(**)	0.08/0.10	−0.11/0.01(**)
11	−0.15/0.00(**)	−0.14/0.00 (**)	−0.00/0.96	0.02/0.72
12	−0.08/0.11	−0.08/0.10	−0.07/0.32	−0.07/0.24
13	−0.23/0.00(**)	−0.23/0.00(**)	−0.17/0.06	−0.08/0.23
14	−0.26/0.00(**)	−0.26/0.00(**)	0.04/0.66	−0.02/0.76
15	−0.07/0.27	−0.12/0.05	0.26/0.00 (**)	−0.24/0.00(**)
16	−0.28/0.00(**)	−0.29/0.00(**)	0.11/0.03 (*)	−0.00/0.95
17	0.40/0.00 (**)	0.41/0.00 (**)	0.08/0.69	−0.11/0.51
18	0.24/0.06	0.26/0.04 (*)	0.21/0.08	0.21/0.14
19	−0.05/0.68	0.02/0.87	0.20/0.20	0.07/0.65

TABLE V. Percentage (number) of the case study sites where significant correlations or differences exist between pairs of social/demographical factors.

	Age	Gender	Occupation	Education	Residence
Age	···	42%(8)	95%(18)	74%(14)	74%(14)
Gender		···	37% (7)	26% (5)	21% (4)
Occupation			···	74%(14)	53%(10)
Education				···	37% (7)
Residence					···

15, 16), and in terms of education, seven case study sites (sites 2, 5–8, 10, 15) present significant differences between local and non-local residents.

In summary, from the above analysis it can be seen that between the social/demographical factors studied, including age, gender, occupation, education and residential status, there are generally considerable correlations, although the correlation coefficients may not be high. Table V shows the percentage and number of the case study sites where such correlations exist. It is also noted that such correlations can be found in a range of cities and countries, although the aim of this paper is not to examine the differences between individual case study sites. It is therefore important to consider the relationships between various social/demographical factors when studying their influence on the sound level evaluation.

IV. SOCIAL/DEMOGRAPHICAL FACTORS AND SOUND LEVEL EVALUATION

The correlations between the sound level evaluation and interviewees' age, occupation and education are shown in Table VI. It is interesting to note that different from sound preference, where it was found that with increasing age, people tend to prefer more natural or cultural-related sounds,[15] age is less important for the sound level evaluation. Among the 19 case study sites only four show significant correlations between age and the sound level evaluation (sites 4, 6, 9, 14), and the coefficient values are rather low, around −0.1. The negative correlations at these four case study sites suggest that with increasing age, people tend to be slightly more tolerant. A possible reason is that these sites are featured by children's shouting and the main function of the sites is recreation and relaxation.

In terms of the effect of occupation on the sound level evaluation, significant correlations are shown at seven case study sites, including sites 3–6, 9, 11 and 14. It is interesting to note that all those sites are located in Europe and the correlation coefficients are negative at six sites. In terms of the effect of education on the sound level evaluation, it is also found that seven case study sites (sites 5, 6, 9, 10, 12, 14, 16) show significant correlations, where four of them are the same sites (sites 5, 6, 9, 14) as those for occupation, which might be because occupation and education are correlated factors at these sites, as can be seen in Table IV. Unlike occupation, the correlation coefficients between the sound level evaluation and education are all positive at the seven sites, although the coefficient values are less than 0.2. This seems to indicate that people with a higher education level show less noise tolerance in urban open spaces.

Previous studies suggested that the effect of gender on the sound annoyance evaluation is generally insignificant,[12–15] although it was reported that males might

TABLE VI. Pearson/Spearman correlation coefficients between sound level evaluation and age, Pearson/Spearman correlation coefficients between sound level evaluation and occupation, and between sound level evaluation and education, with significant levels (2 tailed); as well as mean differences in sound level evaluation between males and females and between local and non-local residents, with significance levels (t test, 2 tailed).

	Correlation/Significance					Mean difference/Significance (male - female; local - nonlocal)	
	Age	Occupation		Education			
Site		Pearson	Spearman	Pearson	Spearman	Gender	Residence
1	−0.04/0.42	0.02/0.70	0.01/0.82	0.09/0.08	0.08/0.11	−0.15/0.02 (*)	−0.12/0.06
2	−0.09/0.07	−0.04/0.38	−0.05/0.37	0.02/0.74	0.02/0.77	0.04/0.57	−0.05/0.40
3	0.07/0.07	0.11/0.01 (**)	0.09/0.02 (*)	−0.04/0.32	−0.02/0.58	−0.03/0.66	−0.11/0.16
4	−0.12/0.00(**)	−0.10/0.01 (**)	−0.10/0.00(**)	0.06/0.11	0.06/0.09	0.00/0.97	0.06/0.90
5	−0.06/0.12	−0.11/0.00 (**)	−0.11/0.00(**)	0.07/0.04 (*)	0.08/0.03 (*)	−0.01/0.92	0.08/0.30
6	−0.12/0.00(**)	−0.12/0.00 (*)	−0.13/0.00(**)	0.12/0.00(**)	0.12/0.00(**)	0.17/0.00 (**)	0.02/0.70
7	0.07/0.11	0.06/0.19	0.05/0.22	0.00/1.00	0.00/0.99	0.05/0.59	−0.05/0.60
8	−0.01/0.90	0.00/0.97	−0.00/0.96	−0.04/0.37	−0.02/0.55	−0.08/0.27	−0.20/0.08
9	−0.09/0.01(**)	−0.12/0.00 (**)	−0.13/0.00(**)	0.12/0.00(**)	0.13/0.00(**)	−0.03/0.57	0.13/0.00 (**)
10	0.04/0.24	0.03/0.323	0.04/0.22	0.08/0.01(**)	0.07/0.04 (*)	−0.17/0.00(**)	−0.09/0.08
11	−0.04/0.35	−0.19/0.00 (**)	−0.21/0.00(**)	0.06/0.18	0.07/0.13	−0.04/0.65	0.08/0.31
12	−0.06/0.21	−0.01/0.88	0.01/0.85	0.10/0.03 (*)	0.11/0.02 (*)	−0.15/0.06	0.34/0.00 (**)
13	−0.00/0.93	−0.03/0.52	−0.04/0.39	−0.03/0.52	−0.03/0.48	−0.01/0.84	−0.08/0.39
14	−0.12/0.01(**)	−0.12/0.01 (**)	−0.12/0.01(**)	0.10/0.03 (*)	0.12/0.01(**)	0.05/0.58	−0.16/0.09
15	0.00/0.99	0.02/0.73	0.05/0.44	0.03/0.61	0.02/0.69	−0.09/0.26	0.02/0.85
16	0.08/0.15	−0.11/0.08	−0.09/0.14	0.17/0.00(**)	0.18/0.00(**)	−0.01/0.95	−0.05/0.56
17	−0.06/0.62	0.01/0.96	−0.03/0.85	0.09/0.53	0.08/0.57	−0.22/0.21	−0.33/0.10
18	0.07/0.57	−0.11/0.40	−0.09/0.47	0.03/0.79	0.06/0.61	−0.05/0.71	0.03/0.85
19	0.09/0.09	0.07/0.58	0.05/0.68	0.23/0.05	0.20/0.09	0.09/0.60	0.30/0.06

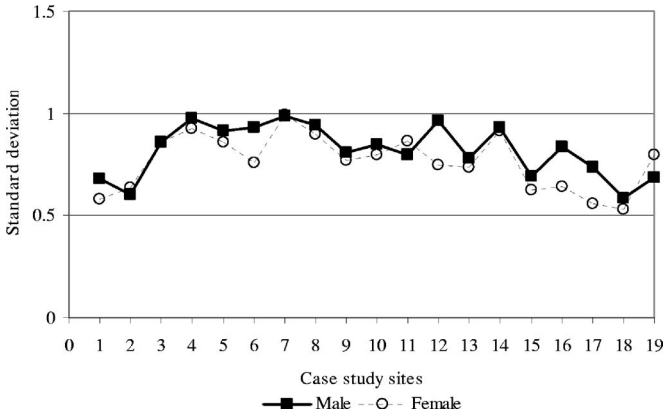


FIG. 1. Standard deviation of males and females in the sound level evaluation.

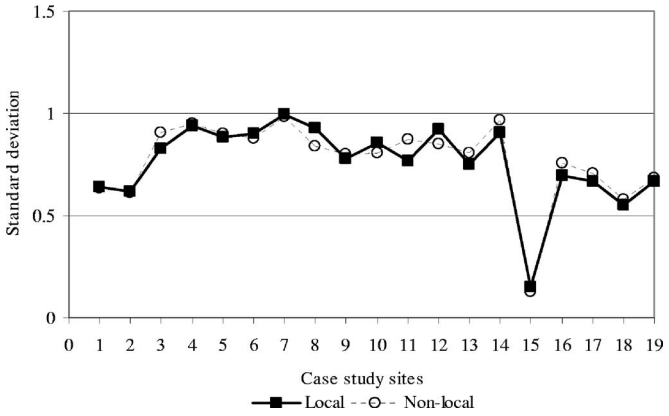


FIG. 2. Standard deviation of locals and non-locals in the sound level evaluation.

be less tolerant than females to low-frequency noise.[20] From Table VI it can be seen that there is generally no significant difference between males and females in terms of the sound level evaluation except at three case study sites (sites 1, 6, 10). Nevertheless, it is also shown that the mean differences are negative at 13 case study sites; namely, the evaluation score of females is slightly higher than that of males, or in other words, males might be more tolerant than females, although the differences are not at a statistically significant level. The standard deviations of the sound level evaluation of males and females are shown in Fig. 1. It is interesting to note that at the majority of the case study sites the standard deviation of males is higher than that of females. This suggests that males tend to have a wider range of evaluation scores, although the differences between males and females are not considerable compared with the standard deviations among males or females, which are both around 0.6–1.

The effect of residential status is generally insignificant, as can be seen in Table VI, where only two case study sites (sites 9, 12) have significant differences between local and non-local residents. At site 12, namely Cambridge Silver Street, the difference is significant, at 0.34 on average, which is probably because at this site there are two distinguished groups, namely local students and overseas tourists. While the standard deviations among locals or non-locals are rather high at the case study sites, at around 0.6–1, as shown in Fig. 2, between the two groups there is generally no significant difference.

Generally speaking, there are some influences from social/demographical factors on the sound level evaluation but they are not very strong. Occupation and education are two related factors and both correlate to the sound level evaluation more than other factors including age, gender and residential status. However, as age is related to occupation and education at most case study sites, their interactions should also be considered. The study also shows notable variations at different case study sites in terms of the effects of social/demographical factors on the sound level evaluation.

V. BEHAVIORS/ACTIVITIES AND SOUND LEVEL EVALUATION

The differences in the sound level evaluation between people wearing and not-wearing earphones during the survey are shown in Table VII, where only the case study sites with over ten samples of wearing earphones are considered due to the limited sample size for this behavioral factor. Consequently, the overall situation in the EU and China is also considered, taking all sites into account. It can be seen that there is no significant difference in any case.

Table VII also shows the difference in the sound level evaluation between people who were reading/writing and not reading/writing during the survey. Again there is no significant difference except at two case study sites, namely the Makedonomahon Square in Thessaloniki (site 6) and the Peace Gardens in Sheffield (site 14). In Table VIII the differences between people who were reading/writing and not reading/writing in terms of social/demographical factors are shown, including age, gender, occupation, education and residential status. It can be seen that while in the Peace Gardens (site 14), reading/writing is an independent factor and people who were reading/writing have a lower evaluation score, by 0.46 in average, in the Makedonomahon Square (site 6), social/demographical factors including age, occupation, education and residential status also significantly affect the differences between those who were reading/writing or not reading/writing, so that the effects of reading/writing on the sound level evaluation may be masked by social/demographical factors. The interactions between reading/writing and social/demographical factors at the other 17 case study sites are also shown in Table VIII. It can be seen that reading/writing relates to age, gender and education at six to seven case study sites, whereas in terms of occupation and residential status, significant differences can only found at three to four sites.

A number of previous studies have suggested that aural and visual aspects are closely related, both contributing to the identification and interpretation of the surrounding spaces.[25–29] In Table IX, the differences in the sound level evaluation between people who were watching and not watching somewhere are shown. It can be seen that there are significant differences at seven case study sites (sites 3–5, 7, 11, 12, 14), which means that watching behavior is more related to the sound level evaluation compared to wearing earphones and reading/writing behaviors. It is interesting to note that all the seven sites are in Europe, two each in Athens

TABLE VII. Mean differences in sound level evaluation between people wearing and not-wearing earphones (only those sites with over 10 samples of wearing earphones are included), as well as between people who were reading/writing and not reading/writing, with significance levels (t test, 2 tailed).

	Mean difference/Significance (not-wearing - wearing; not-reading/writing - reading/writing)	
Site	Earphone	Reading/Writing
1: Germany Kassel, Bahnhofsplatz	0.02/0.95	−0.05/0.68
2: Germany Kassel, Florentiner		−0.02/0.90
3: Greece Athens, Karaiskaki		0.25/0.06
4: Greece Athens, Seashore		0.12/0.48
5: Greece Thessaloniki, Kritis		0.23/0.08
6: Greece Thessaloniki, Makedonomahon	0.26/0.27	0.26/0.03 (*)
7: Italy Milan, IV Novembre		−0.16/0.19
8: Italy Milan, Piazza Petazzi		0.30/0.07
9: Switzerland Fribourg, Jardin de Perolles	−0.07/0.79	−0.21/0.15
10: Switzerland Fribourg, Place de la Gare	0.38/0.06	0.07/0.65
11: UK Cambridge, All Saint's Garden		0.15/0.12
12: UK Cambridge, Silver Street		0.22/0.10
13: UK Sheffield, Barkers Pool		−0.20/0.18
14: UK Sheffield, Peace Gardens		0.46/0.01 (**)
15: China Beijing, Chang Chun Yuan Square		0.15/0.18
16: China Beijing, Xi Dan Square		−0.35/0.08
17: China Shanghai, Century Square		−0.33/0.35
18: China Shanghai, Nanjing Road Square		0.16/0.78
19: China Shanghai, Xu Jia Hui Park		0.20/0.41
EU	−0.05/0.67	
China	0.34/0.16	
Overall	−0.00/0.99	

and Cambridge, and one each in Thessaloniki, Milan and Sheffield. At the case study sites in Athens and Cambridge, the watching people gave lower values of evaluation than non-watching people, namely the watching people felt quieter. A possible reason is that these four sites are located in historic and tourist cities and visually attractive, so that the

TABLE VIII. Mean differences between people who were reading/writing and not reading/writing in terms of age, gender, occupation, education and residential status, with significance levels (t test, 2 tailed). The bold numbers correspond to the case study sites with significant levels in reading/writing activities in Table VII.

	Mean difference/Significance (not-reading/writing - reading/writing)				
Site	Age	Gender	Occupation	Education	Residence
1	0.26/0.37	−0.09/0.31	0.13/0.27	−0.06/0.59	0.18/0.03
2	−0.26/0.38	0.00/1.00	−0.14/0.27	−0.18/0.19	−0.02/0.90
3	0.46/0.10	0.08/0.33	0.20/0.03 (*)	−0.33/0.00(**)	0.01/0.06
4	0.89/0.00 (**)	−0.11/0.22	0.26/0.02 (*)	−0.24/0.04 (*)	0.18/0.03 (*)
5	1.90/0.00 (**)	−0.12/0.10	0.12/0.30	−0.16/0.17	0.14/0.04 (*)
6	**0.56/0.02** (*)	**−0.22/0.00**(**)	**0.10/0.36**	**−0.33/0.00**(**)	**0.30/0.00**(**)
7	−0.27/0.17	0.05/0.37	−0.12/0.17	−0.13/0.11	0.28/0.00 (**)
8	−0.08/0.82	0.22/0.01 (**)	0.02/0.89	−0.26/0.02 (*)	0.10/0.09
9	0.75/0.03 (*)	−0.25/0.01(**)	0.24/0.08	−0.14/0.24	0.10/0.22
10	0.22/0.49	−0.02/0.85	0.23/0.08	−0.26/0.02 (*)	0.08/0.38
11	−0.58/0.00(**)	0.14/0.01 (**)	−0.13/0.10	−0.13/0.03 (*)	−0.01/0.93
12	1.05/0.00 (**)	−0.14/0.07	0.44/0.00 (**)	0.05/0.56	0.08/0.19
13	−0.13/0.71	−0.15/0.13	0.08/0.61	−0.18/0.10	−0.08/0.28
14	**−0.30/0.42**	**−0.09/0.35**	**−0.13/0.40**	**−0.19/0.08**	**−0.03/0.77**
15	−0.07/0.77	0.17/0.05 (*)	0.18/0.10	−0.19/0.05 (*)	0.03/0.69
16	−0.50/0.03 (*)	0.08/0.56	0.23/0.08	−0.10/0.46	0.18/0.21
17	−0.21/0.78	0.20/0.45	0.15/0.68	−0.34/0.32	0.02/0.95
18	−0.99/0.54	0.44/0.39	0.09/0.84	−0.49/0.41	0.32/0.51
19	0.49/0.40	0.43/0.01 (**)	0.35/0.10	0.21/0.27	−0.03/0.88

TABLE IX. Mean differences in sound level evaluation between people who were watching and not watching, moving and not moving, with significance levels (t test, 2 tailed), as well as Spearman correlation coefficients between moving activities and the sound level evaluation, with significance levels (2 tailed).

	Mean difference/Significance (not-watching - watching; not-moving - moving)		Correlation/Significance
Site	Watching	Moving	Moving activities
1: Germany Kassel, Bahnhofsplatz	0.07/0.38	0.11/0.23	−0.01/0.81
2: Germany Kassel, Florentiner	0.08/0.23	0.19/0.01 (**)	−0.10/0.04 (*)
3: Greece Athens, Karaiskaki	0.13/0.05 (*)	−0.09/0.34	0.05/0.24
4: Greece Athens, Seashore	0.30/0.00 (**)	−0.08/0.37	0.06/0.06
5: Greece Thessaloniki, Kritis	−0.31/0.00(**)	−0.04/0.53	0.03/0.41
6: Greece Thessaloniki, Makedonomahon	0.12/0.17	0.10/0.09	−0.03/0.29
7: Italy Milan, IV November	−0.18/0.03 (*)	0.16/0.11	0.02/0.57
8: Italy Milan, Piazza Petazzi	0.05/0.51	0.26/0.01 (**)	−0.12/0.00(**)
9: Switzerland Fribourg, Jardin de Perolles	0.01/0.85	0.06/0.30	−0.04/0.24
10: Switzerland Fribourg, Place de la Gare	0.00/0.95	−0.02/0.79	0.00/0.90
11: UK Cambridge, All Saint's Garden	0.20/0.03 (*)	−0.33/0.13	0.05/0.31
12: UK Cambridge, Silver Street	0.48/0.00 (**)	−0.21/0.06	0.06/0.18
13: UK Sheffield, Barkers Pool	0.21/0.18	0.07/0.32	−0.03/0.50
14: UK Sheffield, Peace Gardens	−0.49/0.01(**)	−0.01/0.95	0.02/0.70
15: China Beijing, Chang Chun Yuan Square	−0.08/0.42	0.07/0.57	−0.04/0.53
16: China Beijing, Xi Dan Square	0.54/0.07	0.18/0.44	−0.06/0.35
17: China Shanghai, Century Square	0.23/0.24	0.12/0.54	−0.05/0.66
18: China Shanghai, Nanjing Road Square	0.14/0.28	−0.05/0.85	−0.05/0.66
19: China Shanghai, Xu Jia Hui Park	−0.09/0.65	0.25/0.51	0.00/0.98

tranquility perception is enhanced. At the other three case study sites (sites 5, 7, 14), conversely, watching people felt nosier than non-watching people, for which a possible reason is that the views from the nearby traffic could bring certain acoustic nuisance. Overall, among the 19 case study sites, only five (sites 5, 7, 14, 15, 19) have negative mean difference, generally indicating that watching people tend to feel quieter.

In Table X the differences between people who were watching and not watching in terms of social/demographical

TABLE X. Mean differences between people who were watching and not watching in terms of age, gender, occupation, education and residential status, with significance levels (t test, 2 tailed). The bold numbers correspond to the case study sites with significant levels in watching activities in Table IX.

	Mean difference/Significance (not-watching - watching)				
Site	Age	Gender	Occupation	Education	Residence
1	−0.55/0.01(**)	0.16/0.01 (**)	−0.20/0.02 (*)	0.02/0.78	−0.05/0.37
2	0.05/0.76	−0.01/0.84	0.11/0.12	0.08/0.30	0.08/0.23
3	**0.46/0.00**(**)	**0.01/0.85**	**0.13/0.01**(**)	**−0.02/0.68**	**−0.06/0.12**
4	**−0.26/0.04** (*)	**0.03/0.38**	**−0.06/0.20**	**0.14/0.01**(**)	**0.01/0.86**
5	**−0.66/0.00**(**)	**−0.01/0.86**	**−0.17/0.06**	**−0.00/0.98**	**0.01/0.82**
6	−0.62/0.00(**)	−0.02/0.70	−0.26/0.00(**)	0.30/0.00 (**)	−0.09/0.07
7	**−0.25/0.08**	**0.00/0.96**	**−0.07/0.23**	**−0.02/0.75**	**−0.00/0.97**
8	−0.63/0.00(**)	−0.00/0.94	−0.22/0.00(**)	0.11/0.03 (*)	−0.02/0.42
9	−0.04/0.75	0.05/0.13	−0.03/0.54	0.07/0.13	−0.01/0.65
10	−0.07/0.54	0.00/0.93	−0.01/0.78	0.03/0.44	0.08/0.02 (*)
11	**−0.17/0.29**	**−0.02/0.66**	**0.07/0.35**	**0.23/0.00**(**)	**0.16/0.00**(**)
12	**−0.84/0.00**(**)	**0.01/0.75**	**−0.24/0.00**(**)	**−0.05/0.29**	**0.00/0.99**
13	0.15/0.67	0.04/0.73	0.06/0.69	0.07/0.55	0.07/0.35
14	**0.16/0.65**	**0.04/0.71**	**0.07/0.63**	**0.22/0.03** (*)	**−0.06/0.48**
15	0.33/0.12	0.10/0.19	0.12/0.21	0.03/0.70	0.12/0.10
16	0.07/0.84	0.07/0.74	0.12/0.62	0.13/0.53	0.24/0.24
17	−0.35/0.39	−0.13/0.37	−0.36/0.06	−0.35/0.04 (*)	−0.03/0.82
18	−0.82/0.02 (*)	0.05/0.67	0.01/0.96	−0.05/0.71	0.11/0.31
19	−0.37/0.46	0.02/0.89	0.22/0.21	−0.02/0.93	−0.09/0.51

TABLE XI. Pearson correlation coefficients between moving activities and age, occupation and education, with significance levels (2 tailed), as well as mean differences between people with and without moving activities in terms of gender and residential status, with significance levels (*t* test, 2 tailed). The bold numbers correspond to the case study sites with significant levels in moving activities in Table IX.

	Correlation/Significance			Mean difference/Significance (male-female; local-nonlocal)	
Site	Age	Occupation	Education	Gender	Residence
1	0.19/0.00 (**)	0.13/0.10	0.10/0.04 (*)	−0.01/0.84	−0.07/0.30
2	**0.20/0.00**(**)	**0.09/0.07**	**0.15/0.00**(**)	**0.27/0.00**(**)	**−0.07/0.48**
3	0.07/0.08	0.08/0.04 (*)	0.06/0.15	−0.09/0.30	−0.21/0.02 (*)
4	−0.08/0.03 (*)	−0.04/0.26	0.14/0.00 (**)	−0.20/0.00(**)	0.03/0.63
5	−0.09/0.02 (*)	−0.08/0.03 (*)	−0.02/0.56	−0.05/0.49	−0.01/0.92
6	−0.06/0.05 (*)	−0.06/0.06	−0.01/0.81	0.05/0.45	−0.04/0.58
7	−0.06/0.16	−0.08/0.07	0.04/0.33	0.15/0.03 (*)	−0.04/0.60
8	**−0.04/0.32**	**−0.08/0.05** (*)	**0.05/0.26**	**0.05/0.43**	**0.16/0.13**
9	0.08/0.02 (*)	0.00/0.90	−0.09/0.01(**)	0.07/0.31	−0.38/0.00(**)
10	0.03/0.32	0.01/0.66	−0.03/0.40	−0.04/0.32	−0.02/0.65
11	−0.01/0.85	−0.02/0.76	0.12/0.01 (**)	0.03/0.49	−0.08/0.09
12	0.11/0.02 (*)	0.14/0.00 (**)	0.13/0.00 (**)	0.15/0.03 (*)	−0.02/0.83
13	−0.07/0.10	−0.17/0.00(**)	−0.03/0.56	−0.04/0.69	−0.02/0.89
14	−0.03/0.50	−0.04/0.38	0.07/0.13	−0.19/0.02 (*)	0.21/0.03 (*)
15	0.18/0.00 (**)	−0.05/0.43	0.09/0.14	−0.14/0.02 (*)	−0.17/0.01(**)
16	−0.06/0.29	−0.08/0.17	0.02/0.72	−0.07/0.01(**)	−0.02/0.44
17	−0.44/0.00(**)	−0.46/0.00(**)	−0.49/0.00(**)	1.66/0.00 (**)	−0.11/0.83
18	−0.21/0.06	−0.04/0.77	−0.18/0.12	−0.01/0.92	−0.40/0.00(**)
19	0.05/0.75	−0.16/0.17	−0.05/0.69	0.02/0.88	−0.13/0.39

factors are shown, including age, gender, occupation, education and residential status. It can be seen that watching behavior is related, at a significant level, to age, education and occupation at eight, six and five case study sites, respectively, whereas the effect of gender and residential status is only significant at one and two sites, respectively. By comparing Tables IX and X, it can be seen at the seven case study sites with significant differences in the sound level evaluation between watching and non-watching people, the effects of social/demographical factors are generally not significant compared with other sites, suggesting that the effects of watching behavior are not from social/demographical factors.

Table IX also shows the differences in the sound level evaluation between people who were moving and not moving, as well as Spearman correlation coefficients between moving activities and the sound level evaluation. It can be seen that significant differences and corrections are only found at two case study sites, Florentiner in Kassel (site 2) and Piazza Petazzi (site 8) in Milan, and the mean differences are around 0.2 and the correlation coefficients are around −0.1, which are both not high. In Table XI the correlations between moving activities and various social/demographical factors are shown, including age, gender, occupation, education and residential status. It can be seen that at the Florentiner Square (site 2) the moving activities are also significantly correlated to age, education and gender, although at the Piazza Petazzi (site 8) only occupation is a significant factor. At other case study sites there are some correlations between moving activities and social/demographical factors, with a significant level at eight sites for gender, six sites for occupation, seven sites for education, five sites for residential status and nine sites for age, but the correlation coefficients are generally rather low, typically below 0.1–0.2, except for residential status, around 0.2–0.4.

In summary, the above results suggest that the effects of wearing earphones, reading/writing, and moving activities on the sound level evaluation are generally insignificant. Conversely, the effects of watching behavior are much more related to the sound level evaluation, again indicating visual/aural interactions. The effects of social/demographical factors on various behaviors and activities are summarized in Fig. 3. It can be seen that only at less than 30–40% of the

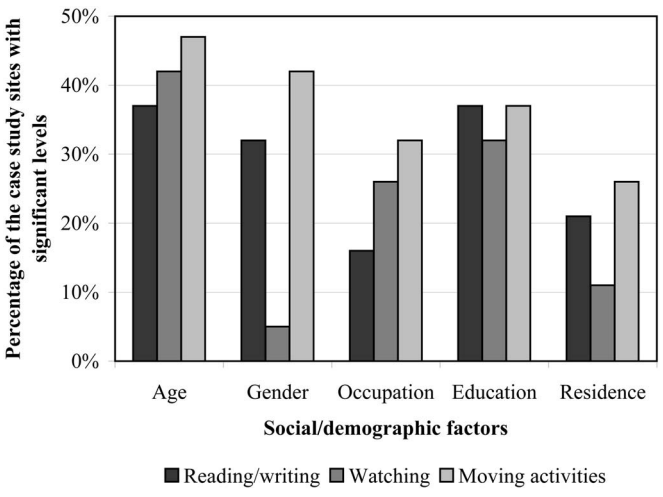


FIG. 3. Percentage of the case study sites where significant correlations or differences exist between social/demographical factors and various behaviors and activities.

TABLE XII. Pearson correlation coefficients between the sound level evaluation at case study sites and at home (2 tailed).

Site	Correlation/Significance
1: Germany Kassel, Bahnhofsplatz	−0.12/0.02 (*)
2: Germany Kassel, Florentiner	−0.06/0.22
3: Greece Athens, Karaiskaki	0.05/0.20
4: Greece Athens, Seashore	0.09/0.01 (*)
5: Greece Thessaloniki, Kritis	0.07/0.04 (*)
6: Greece Thessaloniki, Makedonomahon	−0.01/0.81
7: Italy Milan, IV November	0.02/0.65
8: Italy Milan, Piazza Petazzi	0.18/0.00 (**)
9: Switzerland Fribourg, Jardin de Perolles	0.04/0.22
10: Switzerland Fribourg, Place de la Gare	0.03/0.31
11: UK Cambridge, All Saint's Garden	−0.02/0.68
12: UK Cambridge, Silver Street	−0.03/0.47
13: UK Sheffield, Barkers Pool	0.33/0.00 (**)
14: UK Sheffield, Peace Gardens	0.49/0.00 (**)
15: China Beijing, Chang Chun Yuan Square	0.05/0.39
16: China Beijing, Xi Dan Square	0.03/0.63
17: China Shanghai, Century Square	−0.12/0.38
18: China Shanghai, Nanjing Road Square	−0.15/0.20
19: China Shanghai, Xu Jia Hui Park	0.21/0.07

case study sites the effects are significant, and the mean differences and correlation coefficients are generally not high.

VI. LONG-TERM SOUND EXPERIENCE AND SOUND LEVEL EVALUATION

Long-term sound experience is another important factor for the evaluation of sound quality in urban areas.[30,31] The Pearson correlation coefficients between the sound level evaluation at case study sites and at home are shown in Table XII. It can be seen that significant correlations are reached at six case study sites and among them five correlation coefficients are positive, implying that people living in noisy homes might be less tolerable in an urban open space, although the correlation coefficient values are all below 0.49. Correspondingly, Figs. 4 and 5 show the differences in the mean evaluation score and in the standard deviation of evaluation scores between home and the case study sites, respectively. It is interesting to note that the mean evaluation score at home is generally lower than that at the case study sites, by over 0.5 on average, except at four case study sites, but where the differences are all less than 0.2. This indicates that usually the home environment is quieter than that in urban open spaces, at least in terms of aural perception. Moreover, the standard deviation for the home sound level evaluation is mostly greater than that at the case study sites, by 0.18 in average, except the Kritis Square in Thessaloniki (site 5) and the Peace Gardens in Sheffield (site 14), suggesting that people have a great variety when evaluating the home environment.

Table XIII shows the correlation coefficients between the home sound level evaluation and age, occupation and education, as well as the mean differences in terms of the home sound level evaluation between males and females and between local and non-local residents. These are also compared with the significance levels at the case study sites. It is interesting to note that for those case study sites with significant effects from social/demographical factors, the home sound level evaluation is generally not significantly affected

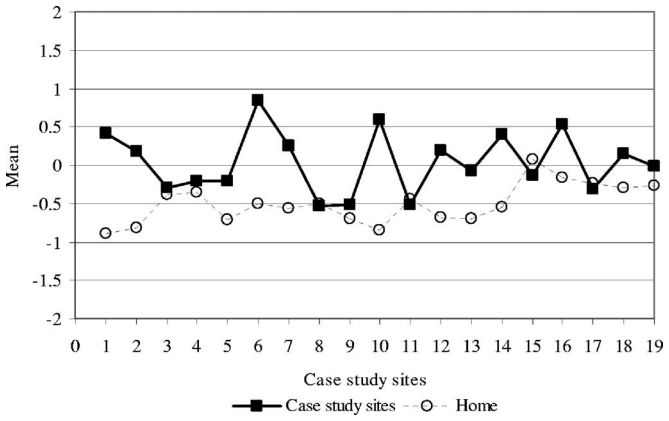


FIG. 4. Comparison in the sound level evaluation between home and case study sites.

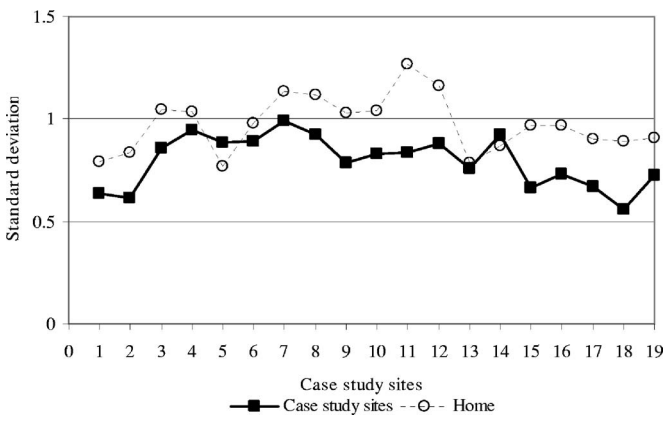


FIG. 5. Comparison in the standard deviation of sound level evaluation between home and case study sites.

TABLE XIII. Pearson correlation coefficients between the home sound level evaluation and age, occupation and education, with significance levels (2 tailed); as well as the mean differences in terms of the home sound level evaluation between males and females and between local and non-local residents, with significance levels (t test, 2 tailed). The corresponding significance levels at the case study sites are also shown.

	Correlation/Significance						Mean difference/Significance (male-female; local - nonlocal)			
	Age		Occupation		Education		Gender		Residence	
Site	Home	Site	Home	Site	Home	Site	Home	Site	Home	Site
1	0.04/0.45		0.05/0.32		0.04/0.37		0.04/0.60	(*)	−0.21/0.01(**)	
2	0.03/0.61		0.00/0.99		0.11/0.03 (*)		0.04/0.68		−0.22/0.01 (*)	
3	−0.05/0.21		0.02/0.67	(**)	0.03/0.45		−0.18/0.03 (*)		−0.05/0.61	
4	−0.05/0.13	(**)	−0.00/0.95	(**)	−0.06/0.11		−0.06/0.44		0.41/0.00(**)	
5	0.09/0.01 (*)		0.04/0.33	(**)	0.00/0.95	(*)	0.06/0.25		0.01/0.90	
6	0.09/0.00 (**)	(**)	0.08/0.01 (**)	(*)	−0.09/0.00(**)	(**)	−0.07/0.23	(**)	−0.36/0.00(**)	
7	0.06/0.17		0.08/0.05 (*)		−0.11/0.01(**)		0.36/0.00 (**)		−0.25/0.02 (*)	
8	0.05/0.26		0.05/0.23		−0.06/0.13		0.09/0.33		−0.30/0.04 (*)	
9	−0.02/0.58	(**)	0.04/0.19	(**)	−0.10/0.00(**)	(**)	−0.10/0.15		−0.51/0.00(**)	(*)
10	−0.09/0.01(**)		−0.05/0.09		−0.03/0.33	(**)	0.00/0.98	(**)	−0.35/0.00(**)	
11	−0.10/0.04 (*)		−0.03/0.51	(**)	−0.10/0.03 (*)		−0.25/0.04 (*)		−0.28/0.02 (*)	
12	0.19/0.00 (**)		0.10/0.03 (*)		−0.03/0.50	(*)	0.01/0.91		0.16/0.21	(**)
13	−0.06/0.18		−0.11/0.02 (*)		−0.00/0.95		0.07/0.34		0.02/0.79	
14	−0.06/0.17	(**)	−0.13/0.00(**)	(**)	0.15/0.00 (**)	(*)	0.13/0.10		−0.06/0.53	
15	−0.17/0.00(**)		0.03/0.61		−0.04/0.45		0.16/0.16		0.12/0.31	
16	−0.07/0.26		0.02/0.79		−0.10/0.10	(**)	0.05/0.66		−0.08/0.51	
17	−0.18/0.19		−0.02/0.90		−0.28/0.04 (*)		0.28/0.25		0.40/0.14	
18	−0.26/0.03 (*)		−0.33/0.01(**)		−0.10/0.39		−0.24/0.25		0.43/0.06	
19	−0.02/0.88		−0.05/0.71		0.15/0.21		−0.04/0.84		0.18/0.39	

by the social/demographical factors, and *vice versa*, except at three sites, namely, the Makedonomahon Square in Thessaloniki (site 6), the Jardin de Perolles in Fribourg (site 9) and the Peace Gardens in Sheffield (site 14), where two of them, sites 6 and 9, are located in residential areas. This suggests that the difference between the sound level evaluation at home and at the case study sites is generally not from those social/demographical factors.

In summary, it seems that the long-term sound experience at home could be an important factor to influence the sound level evaluation in urban open spaces, and this influence is unlikely from social/demographical factors. People from nosier homes could show less sound tolerance in urban open spaces, and the variation is greater when people evaluate the home sound level.

VII. CONCLUSIONS AND DISCUSSIONS

This study explores the effects of social, demographical and behavioral factors on the sound level evaluation, based on intensive field surveys across EU and China. The results suggest that the effects of social/demographical factors including age, gender, occupation, education and residential status, on the sound level evaluation are generally insignificant, although occupation and education are two related factors and both correlate to the sound level evaluations more than other factors including age, gender and residential status. The effects on the sound level evaluation by some behavioral factors including wearing earphones, reading/writing, and moving activities, are also insignificant, but the watching behavior is highly related to the sound level evaluation, indicating visual/aural interactions. Compared to the social, demographical and behavioral factors, the long-term sound experience, i.e., the acoustic environment at home, significantly affects the sound level evaluation in urban open spaces.

It is important to note that between the social/demographical factors studied, including age, gender, occupation, education and residential status, there are generally considerable correlations, although the correlation coefficients may not be high. It is therefore important to consider these correlations when studying the sound level evaluation.

Given the complicated relationships between various factors, it is of great significance to develop prediction models to integrate the effects of multiple factors, for example, using the artificial neural network technique.[32,33] Such models would be useful for aiding urban/architectural designs. Since there are considerable variations between various urban open spaces, it would be more appropriate to establish models for different categories of urban open spaces, instead of a universal model.

ACKNOWLEDGMENTS

The data in this paper are mainly from two research projects funded by the European Commission and the British Academy. The authors are indebted to the project partners for their permission to use the data, and to Dr. Mei Zhang, Dr. Wei Yang, and Dr. Kostas Triantafyllopoulos for useful discussion.

[1]T. J. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405 (1978).
[2]K. D. Kryter, "Community annoyance from aircraft and ground vehicle

noise," J. Acoust. Soc. Am. **72**, 1222–1242 (1982).
[3]H. M. E. Miedema and H. Vos, "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[4]M. Arana and A. Garcia, "A social survey on the effects of environmental noise on the residents of Pamplona, Spain," Appl. Acoust. **53**, 245–253 (1998).
[5]S. A. Ali and A. Tamura, "Road traffic noise levels, restrictions and annoyance in Greater Cairo, Egypt," Appl. Acoust. **64**, 815–823 (2003).
[6]R. Klaeboe, A. H. Amundsen, A. Fyhri, and G. S. Solber, "Road traffic noise—the relationship between noise exposure and noise annoyance in Norway," Appl. Acoust. **65**, 893–912 (2004).
[7]R. Guski, "Psychological determinants of train noise annoyance," in *Proceedings of the Euro-Noise*, Munich, Germany (1998).
[8]B. Berglund, "Community noise in a public health perspective," in *Proceedings of Inter-Noise*, Christchurch, New Zealand (1998).
[9]R. F. S. Job, "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001 (1988).
[10]P. Lercher, "Deviant dose-response curves for traffic noise in 'sensitive areas'," in *Proceedings of Inter-Noise*, Christchurch, New Zealand (1998).
[11]R. Rylander, S. Sorensen, and A. Kajland, "Annoyance reaction from aircraft noise exposure," J. Sound Vib. **24**, 419–444 (1972).
[12]J. M. Fields, "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763 (1993).
[13]R. Tonin, "A method of strategic traffic noise impact analysis," in *Proceedings of Inter-Noise*, Liverpool, UK (1996).
[14]H. M. E. Miedema and H. Vos, "Demographic and attitudinal factors that modify annoyance from transportation noise," J. Acoust. Soc. Am. **105**, 3336–3344 (1999).
[15]W. Yang and J. Kang, "Acoustic comfort evaluation in urban open public spaces," Appl. Acoust. **66**, 211–229 (2005).
[16]M. Maurin and J. Lambert, "Exposure of the French population to transport noise," Noise Control Eng. J. **35**, 5–18 (1990).
[17]D. Bertoni, A. Franchini, M. Magnoni, P. Tartoni, and M. Vallet, "Reaction of people to urban traffic noise in Modena, Italy," in *Proceedings of the 6th Congress on Noise as a Public Health Problem, Noise, and Man*, Nice, France (1993).
[18]J. M. Fields and J. G. Walker, "The response to railway noise in residential areas in Great Britain," J. Sound Vib. **85**, 177–255 (1982).
[19]B. Schulte-Fortkamp, "Combined methods to investigate effects of noise exposure and subjective noise assessment," in *Proceedings of Inter-Noise*, Liverpool, UK (1996).
[20]A. Verzini, C. Frassoni, and A. H. Oritiz, "A field study about effects of low frequency noise on man," in *Proceedings of the 137th Meeting of the Acoustical Society of America and Forum Acusticum*, Berlin, Germany, Abstract published in J. Acoust. Soc. Am. **105**, 942 (1999).
[21]J. Lambert, F. Simonnet, and M. Vallet, "Patterns of behavior in dwellings exposed to road traffic noise," J. Sound Vib. **92**, 159–172 (1984).
[22]W. Yang and J. Kang, "A cross-cultural study of the soundscape in urban open public spaces," in *Proceedings of the 10th International Congress on Sound and Vibration*, Stockholm, Sweden (2003).
[23]J. Kang, *Urban Sound Environment* (Taylor & Francis incorporating Spon, London, 2006).
[24]J. Pallant, *SPSS Survival Manual*, 2nd ed. (Open University Press, United Kingdom, 2005).
[25]J. Lang, *Symbolic Aesthetics in Architecture: Toward a Research Agenda*, in J. L. Nasar (ed.), *Environmental Aesthetics* (Cambridge University Press, United Kingdom, 1988).
[26]J. L. Carles, I. L. Barrio, and J. V. Delucio, "Sound influence on landscape values," Landsc. Urban Plann. **43**, 191–200 (1999).
[27]M. Southworth, "The sonic environment of cities," Environ. Behav. **1**, 49–70 (1969).
[28]J. L. Carles, F. G. Bernaldez, and J. V. Delucio, "Audiovisual interactions and soundscape preferences," Landscape Res. **17**, 52–56 (1992).
[29]R. Pheasant, B. Barrett, K. Horoshenkov, and G. Watts, "Visual and acoustic factors affecting the assessment of tranquility," in *Proceedings of Inter-Noise*, Honolulu, Hawaii (2005).
[30]D. Bertoni, A. Franchini, M. Magnoni, P. Tartoni, and M. Vallet, "Reaction of people to urban traffic noise in Modena, Italy," in *Proceedings of the 6th Congress on Noise as a Public Health Problem, Noise, and Man*, Nice, France (1993).
[31]B. Schulte-Fortkamp and W. Nitsch, "On soundscapes and their meaning regarding noise annoyance measurements," in *Proceedings of Inter-Noise*, Fort Lauderdale, FL (1999).
[32]L. Yu and J. Kang, "Integration of social/demographic factors into the soundscape evaluation of urban open spaces using artificial neural networks," in *Proceedings of Inter-Noise*, Honolulu, Hawaii (2006).
[33]L. Yu and J. Kang, "Soundscape evaluation in city open spaces using artificial neural network," *UIA 2005 – XXII World Congress of Architecture*, Istanbul, Turkey (2005).

Annoyance and disturbance of daily activities from road traffic noise in Canada

David S. Michaud[a] and Stephen E. Keith
Consumer and Clinical Radiation Protection Bureau, Health Canada, Ottawa, Ontario, Canada, K1A 1C1

Dale McMurchy
Dale McMurchy Consulting Inc., Box 252, Norland, Ontario, Canada

(Received 16 May 2007; accepted 15 November 2007)

This study evaluated road traffic noise annoyance in Canada in relation to activity interference, subject concerns about noise and self-reported distance to a major road. Random digit dialing was employed to survey a representative sample of 2565 Canadians 15 years of age and older. Respondents highly annoyed by traffic noise were significantly more likely to perceive annoyance to negatively impact health, live closer to a heavily traveled road and report that traffic noise often interfered with daily activities. Sex, age, education level, community size and province had statistically significant associations with traffic noise annoyance. High noise annoyance consistently correlated with frequent interference of activities. Reducing noise at night (10 pm–7 am) was more important than during the rest of the day. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2821984]

PACS number(s): 43.50.Rq, 43.50.Sr, 43.50.Lj, 43.50.Ba [BSF] Pages: 784–792

I. INTRODUCTION

Noise, defined most simply as unwanted sound, is commonly associated with annoyance reactions. Environmental noise is ubiquitous and annoyance is one of the most widely studied adverse reactions to noise. The World Health Organization (WHO) defines "health" as "*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*" (WHO, 1999). In keeping with this definition, the WHO considers noise-induced annoyance, sleep disturbance and interference with communication as adverse health impacts. More recently, the WHO published a study that showed how strong annoyance towards traffic noise and other neighborhood noise was associated with a statistically significant increase in the adjusted odds ratio for the prevalence of a variety of illnesses, as diagnosed by a physician (Niemann and Maschke, 2004; Maschke and Niemann, 2005).

On an individual scale, the relationship between community noise levels and high noise annoyance is weak because of the influence that multiple personal and situational factors (including misclassification of noise exposure) can have (Fields, 1993; Schultz, 1978; Broadbent, 1972). On a community scale, however, high noise annoyance is more uniform so that estimating community annoyance is possible through the use of established dose-response functions.

In 1978, Schultz first published the relationship between day-night sound level (DNL) and the percentage of an exposed population highly annoyed by any transportation noise source (Schultz, 1978). Schultz defined the "highly annoyed" descriptor as a response to a social survey question on noise annoyance in the top 27%–29% on an anchored numerical scale or in the top two categories on an adjectival, five-point scale. Socio-acoustic research over the last 30 years has produced enough evidence to support separate dose-response functions for predicting high annoyance from aircraft, road traffic and electric rail (Miedema and Oudshoorn, 2001). The International Organization for Standardization (ISO) has published a standard (ISO, 2003) for assessment procedures for environmental noise, which can be done in terms of the percentage highly annoyed. The relationship between the rating level (RL) and percentage highly annoyed is given by

$$\%\,\text{highly annoyed} = 100/[1 + \exp(10.4 - 0.132^{*}\text{RL})], \quad (1)$$

where RL is typically an adjusted DNL, with adjustments made depending on the type of noise source, source characteristics (e.g., tonality) and time of day. The ISO standard specifies that the relationship for road traffic noise is obtained when RL equals DNL. The resulting curve nearly coincides with Schultz's original curve.

Internationally, estimates of exposure to road traffic noise have been made for Europe, Australia and the U.S. In 1996, it was estimated that, for Europe, 40% of the population was exposed to A-weighted traffic sound levels between 55 and 65 dB (DNL) and 20% (nearly 80 million people) was exposed to levels over 65 dB (Commission of the European Communities, 1996). Approximately 8% of the Australian population was exposed to outdoor road traffic noise levels greater than 65 dB during daytime hours (OECD, 1991). The Organization for Economic Co-operation and Development (OECD) estimated that 30% of the U.S. population was exposed to a 24 h time-averaged (Leq24) traffic noise level between 55 and 65 dB and 7% was exposed to traffic levels above 65 dB (Leq24) (OECD, 1986). It has been estimated that approximately 138 million Americans

[a]Author to whom correspondence should be addressed. Electronic mail: dmichaud@hc-sc.gc.ca.

were exposed to outdoor DNL levels above 55 dB, with more than 25 million U.S. citizens exposed to levels above 65 dB (Eldred, 1990).

In Europe, adverse reactions to road traffic noise have been reported to range from 20% to 25% for Austria and France (annoyed), Germany (severely affected) and the Netherlands (highly annoyed) (Lambert and Vallet, 1994). A recent survey conducted in the United Kingdom (U.K.) showed that somewhere between 7% and 9% of the respondents indicated they were either very or extremely bothered, annoyed or disturbed by traffic noise (BRE Environment, 2002). This is close to the recent Canadian survey (Michaud *et al.*, 2005) where traffic noise was identified as the most annoying source of noise with 6.7% of Canadians indicating that they were very or extremely annoyed by traffic noise (i.e., 1.8 million ± 350,000 Canadians 15 and older). Unfortunately, that study was unable to estimate how the annoyance scores correlated with noise levels. Therefore, one of the objectives of the present study was to determine if annoyance reported by Canadians had changed in the three years since the preceding survey.

Social survey research has demonstrated that the magnitude of annoyance reactions towards intruding noise is at least partially related to the extent to which the intruding noise interferes with activities (Fidell *et al.*, 1988; Fields and Hall, 1987). Such activities can include oral communication between people, reading, writing, watching television or listening to the radio. Annoyance reactions may be especially strong when noise interferes with sleep. To date, the extent to which traffic noise interferes with such activities has not been assessed on a national level in Canada. Therefore, a second objective of this study was to correlate annoyance with activity interference. To provide a more detailed picture, this study also asked respondents (1) what was most important with respect to road traffic noise and sleep and (2) what time of day traffic noise levels should be the lowest.

Noise metrics such as Leq24 do not provide any adjustment to noise levels for time of day, suggesting that people are affected-equally by noise, day, evening or night. Clues as to whether this is the case can be obtained from questions about an individual's time preferences for traffic noise to be lowest. In particular, responses can shed light on whether there are time periods that noise restrictions should be more stringent. Increased stringency could be achieved via several ways, e.g., (1) using a separate night time noise metric such as Lnight, (2) using a metric with a time of day adjustment such as DNL, but not Leq 24 alone.

As noted above, noise annoyance is accepted as an adverse health impact unto itself (WHO, 1999). A recent study suggested "strong annoyance" toward community noise was associated with increased risk of some adverse stress-related conditions (Niemann and Maschke, 2004; Maschke and Niemann, 2005). Therefore, one of the objectives of this study was to assess whether high annoyance to traffic noise was perceived to have a negative impact on one's health.

II. METHODS

A. Sampling

The road traffic noise questions were part of a larger public opinion survey that covered a range of health-related issues. The survey commenced with the following introduction: "*Hello, my name is* [***] *we are conducting an important Canada-wide study about health-related issues affecting Canadians. We are not selling or soliciting anything. All your answers will be anonymous and-confidential*." This telephone survey, conducted during October 2005, included a probability sample of 2565 Canadians 15 years and older. Most provinces were allocated a sample size reflecting a 5% margin of error with a 95% confidence interval; the Atlantic provinces had smaller sample sizes and were grouped together for the purposes of analyses. Within each province, the sample was distributed among community strata according to their relative contributions to the overall provincial population. The six community strata used were as follows: (i) less than 5000; (ii) 5000–9999; (iii) 10,000–29,999; (iv) 30,000–99,999; (v) 100,000–999,999; and (vi) one million plus. The number of respondents/completed surveys allocated to each region/province was based on achieving sufficient regional representation—a margin of error of 5%—to analyze results by province. The data were weighted thereafter for the purposes of achieving a national distribution/representation.

The Waksberg-Mitofsky technique is a two-stage random-digit dialing design. First, it identifies three numbers (the prefix) of a telephone number within an area code and then randomly generates a pair of digits to be added to the prefix to produce potential Primary Sampling Units (PSUs). Then, within each region, a random sample of PSUs are selected and a pair of digits is randomly added to produce eligible telephone numbers that are called until the desired sample is reached (Waksberg, 1978). Using the Waksberg–Mitofsky technique, random digit dialing was used to generate potential telephone numbers and, within each household, one subject over 15 years of age was randomly selected using a modified Troldahl–Carter technique. The Troldahl–Carter method chooses a person in the randomly selected household based on their demographic composition (age and sex). This method randomly selects a household member from a grid that includes information on the number of adult household members and the number of adult men or women in the household. It then generates a request to speak to an individual based on their sex and their age relative to other household members (Troldahl and Carter, 1964). Once a potential respondent was chosen using this technique, no other person in the household could be substituted as a respondent. To adjust for any differences in response rates among demographic groups, data were weighted within provinces by age, sex and community size upon completion of the survey. They were also weighted nationally to reflect each province's relative contribution to the overall Canadian population. The overall national margin of error for this study is plus or minus 1.9 percentage points in 19 samples out of 20. The report on telephone interviews is shown in Table I. The response rate was 20%; down from 33% reported in our previous sur-

TABLE I. Report on telephone interviewing.

Ineligible numbers		6371
Non-residential/duplicate	920	
Not in service/fax	5451	
Interview not possible		6910
No answer/busy	2015	
Answering machine	2180	
Screened call	302	
Call-backs	904	
Language barrier	625	
Mental/physical disabilities/age	337	
Respondent not available/quota filled	547	
Refused interviews		10,287
Refusal (screening/introduction)	9061	
Refusal (incomplete interview)	1226	
Completed interviews		2565
Total telephone numbers dialed		26,133
Completion rate (Completed/(Completed+Refused))		19.96%

vey (Michaud *et al.*, 2005). Respondents made their decision to participate without having any knowledge that they would be asked questions on road traffic noise. The eight questions on road traffic noise (see the Appendix) came at the end of a questionnaire designed to probe attitudes towards health care in Canada. The other 91 questions dealt with a wide range of issues such as perceived threats to health, things people would do to change their health, use of health care professionals, details of health status as determined by a health professional and self-medication. The entire questionnaire took, on average, 20 minutes to complete.

B. Statistics

The analytical work commenced with constructing the dataset. First, the dataset was examined for missing information and assessed for potentially spurious data.

Regarding checking and cleaning the data, initial frequencies are run on each response to see whether there are any miscoded responses or outliers. Also dates, ages, etc. are checked to identify glaring errors. The data were reviewed for validity and outliers, and obvious erroneous data were removed. Univariate and bivariate (frequencies, cross-tabulations and analysis of variance) analyses were employed using statistical data management software, SPSS® version 11.5. Results reported were statistically significant at the 0.05 level. Where multiple variables were significant and deemed relevant, logistic regression was employed to identify those factors most predictive of the various outcomes.

It should be noted that data quality verification showed that there were some obvious incongruent responses in the adjectival and numerical items. Since the adjectival item was always presented to the respondent first, responses to this were taken as "valid" reflections of annoyance. When there were extreme inconsistencies on the follow-up numerical question (e.g., a "not at all annoyed" on the adjectival scale was followed by a numerical value of 7 and above or when a "very" or "extremely" annoyed response was followed by either a 0 or 1 on the numerical scale) the responses on the numerical scale were removed from the analysis ($n=46, 1.9\%$). Including these data in the numerical item yielded a lower correlation coefficient between the adjectival and numerical scales (i.e., 0.415, $p<0.01$). It also resulted in a statistically significant increase in high annoyance (i.e., 7 and above) compared to the results obtained using the same numerical scale in 2002 ($p=0.02$). As noted above, these differences were taken as invalid as a result of incongruent responses on the adjectival and numerical items.

III. RESULTS

The eight items related to noise in this survey are provided in the Appendix. Respondents were asked to indicate how annoyed, bothered or disturbed they were by traffic noise on both a five-point adjectival and 11-point numerical

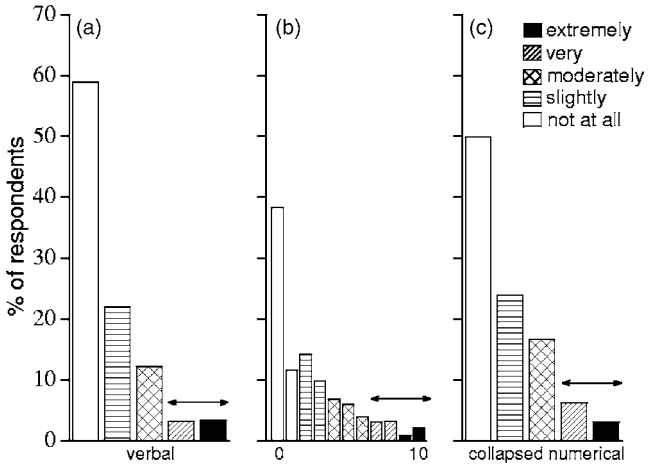


FIG. 1. The distribution of self-reported annoyance towards road traffic noise among respondents interviewed using the ISO/TS 15666 recommended questions for assessing community annoyance. Panel a shows the response on the five-point adjectival scale, panel b shows the range of annoyance on the 11-point numerical scale and panel c presents the results from the numerical scale collapsed according to the following breakpoints (0+1=not at all; 2+3=slightly; 4+5+6=moderately; 7+8=very and 9+10=extremely). Collapsing the numerical scale in this manner yielded a correlation coefficient of 0.726 ($p<0.001$) between panel a and panel c. Bars with arrowheads on each panel demarcate the range of respondents considered "highly annoyed."

TABLE II. Percentage of Canadians highly annoyed by road traffic noise by proximity to a heavily traveled road.

Distance from heavily traveled road	Number of respondents (% of total)	Percentage highly annoyed	
		Adjectival scale	Numeric scale
Less than 30 m	711 (28%)	14.7[a]	18.2[a]
Greater than 30 m, less than, 500 m	844 (33%)	5.1	8.0
Greater then 500 m	1008 (39%)	2.8	4.3

[a]Significant difference in percentage highly annoyed with declared distance, $p<0.0001$.

scale, as recommended by ISO/TS 15666 (ISO, 2002). Figure 1 shows that on an adjectival scale, 6.7% of respondents indicated that they were either very or extremely annoyed (i.e., highly annoyed), and when the numerical scale was collapsed according to the following breakpoints (0+1=not at all; 2+3=slightly; 4+5+6=moderately; 7+8=very and 9+10=extremely), high annoyance was expressed by 9.4% of the respondents. Collapsing the numerical scale as described above yielded a correlation coefficient of 0.726 ($p<0.0001$) between the two questions. No significant differences were observed between these results and those obtained in 2002, where the corresponding percentages highly annoyed on the adjectival and numerical scales were 6.7% and 9.1%, respectively.

Table II shows that respondents who indicated that they lived within 30 m of a heavily traveled road were more likely to indicate that they were either very or extremely annoyed by traffic noise than those who lived further from the road.

Using odds ratio as a measure of likelihood, respondents living within 30 meters of the road were 6.0 times (95% CI 3.88–9.22) more likely to be very or extremely bothered by road traffic noise than those living more than half a kilometer away; and those living half a kilometer or less from the road were 1.8 times (95% CI 1.13–3.01) more likely than those living farther away to be as bothered.

Respondents were asked to rate on an 11-point numerical scale, where 0 was equivalent to "no effect" and ten was equivalent to "very strong effect," the extent to which their annoyance towards traffic noise was perceived to have a negative impact on their health. Table III shows that 94% of the respondents who indicated they were not at all, slightly or moderately annoyed by road traffic noise, perceived the impact of this annoyance on their health to be equivalent to 6 or less, compared to 6% who rated it as 7 and above. Among respondents who were either very or extremely annoyed, 61% perceived the impact of this annoyance on their health to be equivalent to 6 or less, whereas 39% rated it as 7 and above.

Table IV shows the variation in the extent to which respondents were annoyed by road traffic across several demographic variables, including sex, age, education, community size, gross salary, marital status and self-reported health status. Statistically significant differences ($p<0.05$) were found in the percentage highly annoyed among respondents residing in communities with 100,000 or more compared to those in smaller communities. Females were also more likely to be either very or extremely bothered than males. Likewise, individuals between the ages of 25 and 44 years of age, those with a partner and those with an annual income of $20,000–$49,999 were more likely to indicate that they were highly annoyed by traffic noise. Finally, there was also some statistically significant provincial variation, with individuals in Ontario most likely to be either very or extremely bothered by road traffic noise and those in British Columbia being the least likely.

A. Interference of road traffic noise on Canadians' activities while at home

Table V shows the frequency with which traffic noise caused respondents to raise their voice while outside their homes in order to speak with someone next to them. The majority (58%) never had to adjust vocal effort; however, about 18% had to, at least sometimes, raise their voices for this reason. Among those who reported that they were not at all to moderately annoyed by road traffic noise, only 2.7% had to often or always raise their voice to speak to others outdoors. This compares to 36.8% of those who were very or extremely annoyed ($p=<0.0001$).

As might be expected, those residing within 30 m of a major road were more likely to often or always have to raise

TABLE III. The extent to which annoyance towards road traffic noise was perceived to have a negative impact on health.

Annoyance response (N=2510)	Response range 1–6 *n* (%)	Response range 7–10 *n* (%)
Not at all, slightly or moderately	2201 (94.0)[a]	140 (6.0)
Highly annoyed	103 (60.9)[b]	66 (39.1)[a]
Total	2304 (91.8)	206 (8.2)

[a]Significant difference between adjectival annoyance categories, $p<0.05$.
[b]Significant difference between numerical response ranges for perceived impact on health.

TABLE IV. Percentage of Canadians extremely, very, moderately, slightly or not at all bothered by road traffic outside their home by demographic characteristics.

		Not at all	Slightly	Moderately	Very	Extremely	Highly[b]
	Number of respondents (percentage of total N=2532)	1687 (62.7)[c]	552 (19.4)	302 (11.2)	106 (3.9)	73 (2.7)	179 (6.6)
Sex	Male	816 (65.3)	234 (18.7)	135 (10.8)	35 (2.8)	30 (2.4)	65 (5.2)
	Female[a]	678 (52.9)	322 (25.1)	174 (13.6)	50 (3.9)	58 (4.5)	108 (8.4)
Age	15–24	261 (60.6)	102 (23.7)	43 (10.0)	12 (2.8)	13 (3.0)	25 (5.8)
	25–44[a]	521 (55.1)	206 (21.8)	132 (14.0)	41 (4.3)	45 (4.8)	86 (9.1)
	45–64	463 (61.1)	171 (22.6)	86 (11.3)	17 (2.2)	21 (2.8)	38 (5)
	65+	250 (63.0)	78 (19.6)	47 (11.8)	14 (3.5)	8 (2.0)	22 (5.5)
Education	<secondary	76 (69.1)	12 (10.9)	19 (17.3)	0 (0)	3 (2.7)	3 (2.7)
	Secondary	561 (62.1)	182 (20.2)	96 (10.6)	35 (3.9)	29 (3.2)	64 (7.1)
	>secondary	856 (56.4)	363 (23.9)	194 (12.8)	49 (3.2)	55 (3.6)	104 (6.8)
Community size (est. by respondent)	<5000	193 (67.0)	48 (16.7)	30 (10.4)	6 (2.1)	11 (3.8)	17 (5.9)
	5000–99,999	333 (65.2)	103 (20.2)	56 (11.0)	9 (1.8)	10 (2.0)	19 (3.8)
	>100,000[a]	968 (55.8)	406 (23.4)	224 (12.9)	70 (4.0)	66 (3.8)	136 (7.8)
Gross salary ($1000/yr)	<20	235 (58.2)	87 (21.5)	56 (13.9)	10 (2.5)	16 (4.0)	26 (6.5)
	20–50[a]	410 (54.8)	161 (21.5)	109 (14.6)	39 (5.2)	29 (3.9)	68 (9.1)
	>50	647 (60.7)	250 (23.5)	106 (9.9)	31 (2.9)	32 (3.0)	63 (5.9)
Marital status	Partner[a]	717 (57.8)	281 (22.6)	141 (11.4)	43 (3.5)	59 (4.8)	102 (8.3)
	No partner	768 (60.1)	275 (21.5)	166 (13.0)	41 (3.2)	28 (2.2)	69 (5.4)
Self-report health status	Poor/fair	238 (53.0)	114 (25.4)	69 (15.4)	17 (3.8)	11 (2.4)	28 (6.2)
	Good/ excellent	1253 (60.3)	442 (21.3)	240 (11.5)	67 (3.2)	76 (3.7)	143 (6.9)
Province	B.C.	178 (54.6)	94 (28.8)	43 (13.2)	3 (0.9)	8 (2.5)	11 (3.4)
	Alberta	196 (60.5)	83 (25.6)	28 (8.6)	8 (2.5)	9 (2.8)	17 (5.3)
	Sask.	231 (70.4)	63 (19.2)	18 (5.5)	10 (3.0)	6 (1.8)	16 (4.8)
	Manitoba	223 (67.2)	67 (20.2)	25 (7.5)	10 (3.0)	7 (2.1)	17 (5.1)
	Ontario[a]	195 (59.6)	78 (23.9)	27 (8.3)	12 (3.7)	15 (4.6)	27 (8.3)
	Quebec	178 (56.2)	48 (15.1)	68 (21.5)	14 (4.4)	9 (2.8)	23 (7.2)
	Atlantic	365 (62.0)	124 (21.1)	67 (11.4)	15 (2.5)	18 (3.1)	33 (5.6)

[a]Significantly more likely to report being highly bothered by traffic noise, $p<0.05$.
[b]Highly annoyed is calculated by adding "very" and "extremely" and is shown to faciliate comparisons with other tables.
[c]Cells for each variable may not always add to the corresponding sample size because respondents could choose to not answer questions.

their voice to speak with someone next to them compared to those living between 30 m and half a km or more than a half a km from a heavily traveled road (i.e., 12% versus 3% versus 2%, respectively). Furthermore, increasing vocal effort to communicate outdoors was statistically more likely to occur in communities with populations greater than 100,000, among female respondents and for those residing in either Quebec or Ontario.

TABLE V. Percentage of Canadians who have had to raise their voices because of road traffic noise when talking to someone outside their homes by annoyance with road traffic noise.

How often vocal effort is increased to communicate	Not at all, slightly or moderately annoyed N=2356 (%)	Highly annoyed N=171 (%)	Total N=2529 (%)
Never	61.2	7.6	57.6
Seldom	23.9	17	23.4
Sometimes	12.1	38.6	13.9
Often	2.0	21.6[a]	3.3
Always	0.7	15.2[a]	1.7

[a]Statistically significant difference from those not at all, slightly or moderately annoyed, $p<0.0001$.

Respondents were also asked whether traffic noise *often* interfered with various daily activities over the past 12 months. Fourteen percent of respondents indicated that road traffic noise often interfered with their ability to sleep, 13% said it interfered with them hearing other people or the TV/radio, and 11% said it often affected their ability to concentrate on tasks such as reading and writing. These activity interferences were significantly more likely to be reported by those who were either very or extremely annoyed by traffic noise compared to those not at all, slightly or moderately annoyed (see Table VI).

Table VII shows that self-reported proximity to a major road was also related to the extent to which road traffic noise interfered with the respondents' sleep and leisure time, with those within 30 m more likely to report that road traffic noise *often* interfered with these activities.

The data analyses also showed that those who indicated that their daily activities were *often* interfered with by noise in the last 12 months were also more likely to perceive their level of annoyance with road traffic noise as having a negative impact on their health (seven or higher on the 11-point numerical scale) (Table VIII).

TABLE VI. Interference of road traffic noise on various behaviors among Canadians while at home over the past 12 months.

	Yes (%)	Among those highly bothered by road traffic noise (%)
Road noise often interfered with the ability to sleep (N=2533)	13.8	69.8[a]
Road noise often interfered with the ability to hear people, the TV and radio (N=2532)	13.4	55.0[a]
Road noise often interfered with the ability to read and write (N=2528)	10.5	50.6[a]

[a]Statistically significant from not at all, slightly and moderately annoyed, $p<0.05$.

Respondents were asked what was most important with respect to road traffic noise and sleep. Table IX shows that among those who were not at all, slightly or moderately annoyed by road traffic noise ($n=2334$), 24% indicated that road traffic noise should not make it difficult to fall asleep, and 21% responded that it should not wake them during the night. Fewer (14%) felt that road noise should not wake them up too early in the morning, and 6% placed equal importance on all aspects of sleep quality. For 37% of these respondents, none of these issues were important to them. Among those very or extremely (highly) annoyed by road traffic noise ($n=171$), 16% placed the highest importance on being able to fall asleep, about 35% felt it most important that they were not awakened during the night compared to 26% who did not want to be awakened prematurely in the morning. Another 13% placed equal emphasis on all aspects of sleep and 11% said that none of these issues were important to them.

Finally, there was a clear preference among respondents that it was most important to have lower traffic noise levels between 10 pm and 7 am; 66% of respondents wanted traffic noise levels to be lowest at this time of day, compared to 14% who indicated evening hours (7 pm–10 pm). There was a nearly equal split between those who wanted levels to be lowest during daytime hours (7 am–7 pm) and those who were not concerned with traffic noise levels (8%). Slightly less than 4% indicated that they wanted traffic noise levels to be low at all times of the day. Among those very or extremely annoyed by road traffic noise, 48% wanted levels to be lowest at night compared to 30% who indicated evening and 10% who did not distinguish between times of day.

TABLE VII. Interference of road traffic noise with Canadians' daily activities while at home over the past 12 months by proximity to a major road.

	Yes (%)
Road traffic noise often interfered with sleep N=2511	
Less than 30 meters	26.8[a]
Half a km or less	12.0
Greater than half a km	6.5
Road traffic noise often interfered with hearing people, TV or radio N=2509	
Less than 30 meters	22.1[a]
Half a km or less	13.7
Greater than half a km	7.1
Road traffic noise often interfered with reading and writing N=2507	
Less than 30 meters	18.7[a]
Half a km or less	11.2
Greater than half a km	4.3

[a]Statistically significant difference between self-reported distance, $p<0.00001$.

IV. DISCUSSION

There is sufficient scientific and anecdotal evidence showing that transportation noise can represent a significant source of noise annoyance. Our previous study showed that traffic noise was identified as the most common source of noise annoyance in a representative sample of Canadians. Statistics Canada estimated that in 2003 the Canadian population aged 15 and over was approximately 26 million (Statistics Canada, 2004). Thus, our results suggest that 1.8 million Canadians over the age of 15 (±350,000) are highly annoyed by traffic noise. Using both the adjectival and numerical scales to characterize the magnitude of this annoyance response showed remarkable stability between our previous work and the current report. Thus, on a national level, the expression of high annoyance towards traffic noise has not changed significantly in the three years that elapsed between surveys. Our results are comparable to those obtained

TABLE VIII. Percentage of Canadians who perceived a negative health effect due to their annoyance with road traffic noise by interference of road traffic noise on daily activities while at home over the past 12 months.

	Perceived negative impact on health due to annoyance with road traffic noise (7 or higher on an 11-point numerical scale)
Road traffic noise often interfered with sleep N=2513	
Yes	26.7%[a]
No	5.3%
Road traffic noise often interfered with hearing people, TV or radio N=2512	
Yes	20.6%[a]
No	6.4%
Road traffic noise often interfered with reading and writing N=2508	
Yes	27.9%[a]
No	5.8%

[a]Statistically significant from respondents indicating the following behaviors were not often interfered with, $p<0.0001$.

TABLE IX. Impact on sleep that Canadians indicated as the most important with respect to road traffic noise by how annoyed they were with road traffic noise over the last 12 months.

Aspect of sleep quality indicated to be most important regarding road traffic noise.	Annoyance category: Not at all; slightly; moderately (n=2334)	Annoyance category: highly (n=171)
Road traffic noise should not make it difficult to fall asleep	24%	16%
Road traffic noise should not cause an awakening during the night	21%	35%
Road traffic noise should not cause an awakening too early in the morning	14%	26%
All of the above	6%	13%
None of the above	37%	11%

in the national survey conducted in the U.K. where it was found that 8% of the population was either very or extremely annoyed by traffic noise.

Consistent with our previous work, our results indicated that traffic noise annoyance was greater among women. However, in this study, respondents in the middle income category ($24,000–$49,999) were most likely to be highly annoyed, whereas this annoyance was greatest among the highest income earners in 2002. In either study, these results were not entirely consistent with those of Fields in his review of the personal and situational factors contributing to noise annoyance. He found that education, income and age had no influence on annoyance ratings (Fields, 1993). Indeed, it could be argued that our inconsistent findings with respect to income and annoyance suggest that reported income is not strongly related to annoyance scores.

Since acoustic variables may account for one-third of the variance in annoyance, (Guski, 1999) the present study sought to improve on our previous national surveys (Michaud *et al.*, 2005) by asking respondents to estimate how far their home was from a heavily traveled road. Respondents were told that a heavily traveled road was considered to be one that had a posted speed limit of at least 80 km per hour or one that had four lanes. From the respondent's estimate, the interviewer categorized the response as being in one of three categories: (1) immediately next to the road (within 30 m), (2) between 30 m and half a km or (3) beyond half a km.

The correlation between vocal effort outdoors during conversation and distance from the heavily traveled road supports this parameter as an index of noise exposure. The key words in the vocal effort question (Q.5 in the Appendix) and the responses were the references to duration (i.e., often or always) and distance (i.e., right next to you). The "often" or "always" rule out spurious events that might cause people to raise their voice when speaking outdoors and the "right next to you" suggests a distance somewhere between 0.5 and 2 m between speakers. At these distances, the steady sound levels that would require someone to raise their voice in order to communicate are typically in the range of 66–78 dB (U.S. Environmental Protection Agency, 1974).

As the self-reported distance to a heavily traveled road increased, there was a substantial drop in the percentage of respondents indicating that they *often* found it difficult to sleep, hear people/television/radio, read or write. Similarly, among those respondents who indicated they were highly annoyed by traffic noise, a greater percentage lived within 30 m of a heavily traveled road than 500 m or further. This latter finding places some hesitation in assuming that individuals who live closest to a noise source are self-selected because they are less bothered by the intruding noise source. However, in this regard, this study did not ascertain the presence of reflecting surfaces, window type, position of main living quarters relative to the road, or if the respondent was the individual who made the decision to live at that location. Although, given that daily activity disturbance was greater within 30 m of a heavily traveled road, it is unlikely that such dwellings represent those effectively protected against intruding noise with sound barriers or similar noise-attenuating methods.

One of the objectives to this study was to assess the extent to which traffic noise annoyance was viewed as having a negative impact on one's health. The results showed that when compared to the lower annoyance categories, the respondents indicating they were highly annoyed by road traffic noise were significantly more likely to indicate that this annoyance had a negative impact on their health (seven or greater on an 11-point numerical scale). Only 6% of respondents who claimed lower magnitudes of annoyance (not at all, slightly and moderately) also reported a negative impact on their health. This finding supports self-reported high annoyance as a health effect specifically, rather than lower magnitudes of annoyance. It should be noted, however, that respondents were not asked how their health was adversely impacted by noise or how they knew it was adversely affected. In fact, there was no correlation between a separate question about self-reported health status (i.e., good or poor) and high noise annoyance. Despite this, the subjective impact that traffic noise annoyance was perceived to have on health is consistent with recent findings from a study directed by the WHO on housing and health status. The results of that study suggested that *strong* annoyance, compared to *moder-*

ate annoyance, towards traffic noise had a greater association with the prevalence of a variety of illnesses, such as hypertension and migraines (Niemann and Maschke, 2004; Maschke and Niemann, 2005).

This survey was conducted during a period of the year when residential windows are much more likely to be completely closed and when people tend to spend more time indoors. Fully closed windows can attenuate outdoor sound levels by as much at 35 dB. When this transmission loss is combined with an increase in the time spent indoors, it is likely that the level of intruding noise during the month of October to which respondents were exposed was lower than it would have been in warmer months. For these reasons, respondents were specifically instructed to respond based on their experience *over the last* 12 months *or so*, but such instruction may not fully account for seasonal effects. Seasonal effects on noise annoyance have been shown to account for as much as 10% of the variability in annoyance (Fields *et al.*, 2000). Nevertheless, our results remain comparable to our previous study and the U.K. study since both were conducted in winter months. Furthermore, respondents who were highly annoyed were much more likely than those less bothered to indicate that traffic noise caused them to raise their voice outdoors when talking to someone next to them. Specifically, when asked if they had to do this *never, seldom, sometimes, often*, or *always*, those highly annoyed were more likely to indicate *often* or *always*.

In Canada there is no standard as to what a minimum acceptable response rate for random digit dialed telephone surveys is; and while our 20% response rate was down from around 33% in our previous survey (Michaud *et al.*, 2005), this rate is still well within typical response rates for most commercial telephone surveys, which tend to range between 10% and 20%, or lower, depending upon the nature of the survey (PWGSC, 2007). In the United States, the Council for Marketing and Opinion Research (CMOR) found that the average response rate for 761 surveys that used random digit dialing was 9.17% (CMOR, 2002). Although low response rates are common for public opinion research that utilizes random digit dialling (O'Rourke *et al.*, 1998), a low response rate does not *necessarily* weaken the validity of the data, just as a high response rate does not guarantee it (see references in PWGSC, 2007). This study ensured that respondents were randomly selected and included provincial and national sample selection and weighting according to community size, sex and age to increase the representativeness of the data. Furthermore, all stages of this survey, from questionnaire design to administration and data verification standards, follow the best practices in public opinion research identified by PWGSC (2007) and the Canadian Association of Marketing Research Organizations (CAMRO).

Despite these efforts, one might speculate that chronic exposure to road traffic noise at high levels could discourage potential respondents from communicating on the telephone and/or engender a general state of irritability that makes potential respondents less willing to participate in telephone surveys. Conversely, choosing to participate in a health-related survey might reflect a greater general interest in health among participants.

Biases with respect to noise were minimized in this study because a respondent's initial decision to participate or refuse to participate in the telephone survey was made without any knowledge that the survey would contain questions related to environmental noise. Also, the results for the %HA in this study were similar to that reported previously, despite the drop in response rate. Furthermore, the majority of the respondents were not bothered by road traffic noise and finally, follow-up calls were made to individuals with soft refusals and telephone numbers with no initial response (see Table I above).

V. CONCLUSION

National surveys with representative samples provide important insight not only on how road traffic noise annoyance in Canada compares to other countries, but also on how the dynamics of this response change over time. The consistency between this study and our previous survey (Michaud *et al.*, 2005) suggests that the extent of high annoyance towards road traffic noise is close to 6.7% in Canada.

The current study has also shown that on a national scale, self-reported distance to a major road positively correlated with self-reported adverse impacts from road traffic noise, including the magnitude of annoyance and the disturbance of daily activities. In terms of providing protection to the public against the adverse health effects of traffic noise, this study supports the distinction between high annoyance and lower magnitudes of annoyance, insofar as those highly annoyed were more likely than those less annoyed to report disturbance of daily activities (e.g., sleep, oral or written communication) and perceive their annoyance as having had a negative impact on their health.

These results show that there is promise of providing a dose-response curve for traffic noise annoyance in Canada, but only if future questionnaire items are able to provide more information on variables that would influence immission sound levels at the respondent's home. In addition to self-reported set back distances, these would include, but not be limited to, the presence or absence of a sound barrier, the type of dwelling (e.g., bungalow or multi-level) and if the respondent has a line of sight to the road traffic from the most frequently occupied room in the house.

APPENDIX: TELEPHONE QUESTIONNAIRE

Opening recruitment script: "*Hello, my name is* [***] *we are conducting an important Canada-wide study about health-related issues affecting Canadians. We are not selling or soliciting anything. All your answers will be anonymous and confidential.*"

The following eight noise-related questions follow 91 general health-related questions.

1. How close is your home to a heavily traveled major road? [Prompt—heavily traveled can be considered to be either one that has four or more lanes or has a posted speed limit of 80 km an hour and over]
   Do not read. Mark only one.
   - home is on or next to this road, e.g., less than 30 m

- half a kilometer or less or about five-blocks from the road
- more than half a kilometer
- refused
- don't know

2. Thinking about the last 12 months or so, when you are at home, how much does noise from road traffic bother, disturb, or annoy you: Not at all, Slightly, Moderately, Very or Extremely?
3. I would now like to ask you this question based on a scale of zero to ten where zero is not at all bothered at home and ten is extremely bothered. Thinking about the last 12 months or so, what number from zero to ten best shows how much you are bothered, disturbed or annoyed by road traffic noise?
4. Again on a scale of zero to ten where zero is no effect and ten is a very strong effect to what extent do you think your annoyance with road traffic noise has a negative impact on your health?
5. When you are outside your home and talking to someone who is right next to you, do you never, seldom, sometimes, often or always have to raise your voice because of road traffic noise?
6. Over the past 12 months or so, while you were at home, did road traffic noise *often* interfere with your ability to… [Prompt, to a point where it was perceived as problematic]
   - sleep [Prompt, this could be related to falling asleep, staying asleep or cause early awakenings]
   - hear other people or the TV and radio inside your home
   - concentrate on tasks such as reading and writing
7. Which of the following is MOST important to you with respect to road traffic noise and the quality of your sleep? It does not….*Select one only*
   - make it difficult to fall asleep
   - waken you during the night
   - wake you up too early in the morning
   - Volunteer—all of the above
   - Volunteer—none of the above
8. When do you want the level of road traffic noise at your home to be the *lowest*? During the…
   - Day
   - Evening
   - Night
   - Volunteer—the level of road traffic noise at my home is not important to me
   - Volunteer—all of the above times

BRE Environment (**2002**). "The 1999/2000 national survey of attitudes to environmental noise—Volume 3 United Kingdom results," Report No. 205217f.

Broadbent, D. E. (**1972**). "Individual differences in annoyance by noise," Sound **6**, 56–61.

Commission of the European Communities. (**1996**). "Future Noise Policy," Brussels, 1–31.

Council for Marketing and Opinion Research (CMOR) (**2002**). "Tracking System—Cooperation, Refusal and Response Rates," Available at: http://www.cmor.org/rc/studies.cfm (last viewed on 08/30/2007).

Eldred, K. M. (**1990**). "Noise at the year 2000," *Proceedings of the Fifth International Congress on Noise as a Public Health Problem*, edited by B. Berglund *et al.*, Stockholm, Sweden. Swedish Council for Building Research, 355–381.

Fidell, S., Schultz, T., and Green, D. M. (**1988**). "A theoretical interpretation of the prevalence rate of noise-induced annoyance in residential populations," J. Acoust. Soc. Am. **84**, 2109–2113.

Fields, J. M. (**1993**). "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M., Ehrlich, G. E., and Zador, P. (**2000**). "Theory and design tools for studies of reactions to abrupt changes in noise exposure," NASA/CR-2000-210280.

Fields, J. M., and Hall, F. L. (**1987**). "Community effects of noise," in *The Effects of Transportation Noise on Man*, edited by P. M. Nelson London: (Butterworths, London), pp. 1–27.

Guski, R. (**1999**). "Personal and social variables as co-determinants of noise annoyance," Noise Health **1**, 45–56.

ISO. (**2002**). "Acoustics-assessment of noise annoyance by means of social and socio-acoustic surveys," ISO/TS 15666:2002

ISO. (**2003**). "Acoustics-description, measurement and assessment of environmental noise-Part 1: Basic quantities and assessment procedures," ISO 1996-1:2003(E).

Lambert, J., and Vallet, M. (**1994**). "Study related to the preparation of a communication on a future EC noise policy," Report No. 9420:1-143.

Maschke, C., and Niemann, H. (**2005**). "Health effects of neighborhood noise-induced annoyance," *Proceedings of Inter-Noise*, Rio de Janeiro, Brazil.

Michaud, D. S., Keith, S. E., and McMurchy, D. (**2005**). "Noise annoyance in Canada," Noise Health **7**, 39–47.

Miedema, H. M., and Oudshoorn, C. G. (**2001**). "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416.

Niemann, H., and Maschke, C. (**2004**). "Noise effects and morbidity-WHO LARES Final report," available at http://www.euro.who.int/Document/HOH/LARES_results.pdf (accessed on 08/21/2006).

O'Rourke, D., Chapa-Resendez, G., Hamilton, L., Lind, K., Owens, L., and Parker, V. (**1998**). "An inquiry into declining RDD response rates. Part 1: Telephone survey practices," Survey Res. **29**, 1–14.

OECD. (**1986**). "Fighting noise: Strengthening noise abatement policies," Renouf Pub. Co. Ltd., France.

OECD. (**1991**). "Fighting noise in the nineteen nineties," OECD & Devel, France.

PWGSC. (**2007**). "Best practices in public opinion research: Improving respondent cooperation for telephone surveys," Public Works Government Services Canada, Ottawa, Canada, Catalogue Number: P103-2/2007E-PDF.

Schultz, T. (**1978**). "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.

Statistics Canada (**2004**). "Annual Demographic Statistics (2004)," Cataloge No. 91-213-XIB.

Troldahl, V. C., and Carter, R. E. (**1964**). "Random selection of respondents with households in phone surveys," J. Marketing Surveys **1**, 71–76.

U.S. EPA. (**1974**). "Information on levels of environmental noise requisite to protect public health and welfare with an adequate margin of safety," Report No. 550/9-74-004:1-G15.

Waksberg, J. (**1978**). "Sampling methods for random digit dialing," J. Am. Stat. Assoc. **73**, 40–46.

WHO. (**1999**). "Guidelines for Community Noise," edited by B. Berglund, T. Lindvall, and D. H. Schwela, World Health Organization, Geneva. http://www.who.int/docstore/peh/noise/guidelines2.html (accessed on (02/26/2007).

Prediction of the sound field above a patchwork of absorbing materials

R. Lanoye[a] and G. Vermeir
Afdeling Akoestiek en Thermische Fysica - Afdeling Bouwfysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, BE-3001 Heverlee, Belgium

W. Lauriks[b]
Afdeling Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, BE-3001 Heverlee, Belgium

F. Sgard[c]
Laboratoire des Sciences de l'Habitat, DGCB URA CNRS 1652, Ecole Nationale des Travaux Publics de l'Etat, Rue Maurice Audin, F-69518 Vaulx-en-Velin Cedex, France

W. Desmet
Department of Mechanical Engineering, Division PMA, Katholieke Universiteit Leuven, Celestijnenlaan 300B, BE-3001 Heverlee, Belgium

(Received 11 October 2006; revised 17 November 2007; accepted 20 November 2007)

The aim of this paper is to investigate the acoustic performance of sound absorbing materials through a numerical wave based prediction technique. The final goal of this work is to get insight into the acoustic behavior of a combination of sound absorbing patches. In order to address a wide frequency range, a model based on the Trefftz approach is adopted. In this approach, the dynamic field variables are expressed in terms of global wave function expansions that satisfy the governing dynamic equations exactly. Therefore, approximation errors are associated only with the boundary conditions of the considered problem. This results in a computationally efficient technique. The main advantage of this method is the fact that the sound absorbing patches do not have to be locally reacting. In this article, the wave based method is described and experimentally validated for the case of normal incidence sound absorption identification in a standing wave tube. Afterwards, the method is applied to simulate some interesting setups of absorbing materials.
 [DOI: 10.1121/1.2823781]

PACS number(s): 43.55.Dt, 43.20.Bi, 43.20.Hq, 43.55.Ev [NX] Pages: 793–802

I. INTRODUCTION

Sound absorbing finishings are commonly used as a noise control measure in room acoustics applications, transport vehicles and industrial machinery. Since the acoustic performance of porous materials is poor at low frequencies, layered media or patchwork patterns of different materials can be used to improve this performance. For simple layered combinations, analytical methods, such as the transfer matrix method, may be used to evaluate the acoustical properties,[1] but these methods are limited to one-dimensional wave propagation. These methods cannot be used to investigate more complex three-dimensional configurations, such as patchwork patterns. To know the frequency dependent properties of these complex configurations, the use of numerical techniques is inevitable since solutions in analytical form of the problem are rare due to the complexity of the geometry and/or boundary conditions.

To model the absorbing materials, the full Biot theory can be used.[2] For materials with a heavy or rigid frame and low flow resistivity, simpler models such as the equivalent complex fluid can be used as an alternative to Biot's theory to describe the wave propagation inside the porous material. For locally reacting porous materials, the material can even be replaced by an impedance boundary condition.

Nowadays, different calculation techniques can be used to describe the acoustic behavior of patchworks of absorbing materials. A first technique is the periodic structures homogenization method. This technique was used by Olny and Boutin[3] to study the behavior of double porosity materials. This method also allows for the prediction of the sound absorption of porous materials with slits.

Atalla *et al.*[4] and Sgard *et al.*[5] developed a finite element based numerical model to study acoustic absorption of porous media made up of porous patches with different properties. In particular, their numerical formulation accounted for the assembling of air cavities and multiple porous materials. Today, $\{\mathbf{u},\mathbf{p}\}$ formulations of the Biot–Allard theory,[6] based on the displacement $\mathbf{u}$ in the solid phase and the interstitial fluid phase macroscopic pressure $\mathbf{p}$, are mainly used. They have been proven to be numerically more efficient than

[a]Research Assistant of the Fund for Scientific Research - Flanders (Belgium).
[b]Author to whom correspondence should be addressed. Electronic mail: walter.lauriks@fys.kuleuven.be
[c]Presently at IRSST Service de la recherche, 505 Boulevard de Maisonneuve Quest, Montreal, Quebec H3A 3C2 Canada.

the classical $\{\mathbf{u},\mathbf{U}\}$ formulation,[2] which uses the displacement $\mathbf{u}$ in the solid phase and the displacement $\mathbf{U}$ in the fluid phase as unknowns.

Also the boundary element method is a well-known calculation technique in acoustics. Boutin[7] developed the formalism of boundary integral equations for porous media. Tanneau[8] used the boundary element method to model the acoustic behavior of porous media. The mentioned articles only deal with layered media, patchworks of absorbing materials have not been calculated with the boundary element method so far.

Furthermore, hybrid solution techniques have been explored by Sgard.[9] Work was performed on describing the absorption qualities of single absorbing materials and multilayered media by use of a hybrid solution technique.

To solve the problem of interest in an alternative way, a new deterministic prediction technique, based on the indirect Trefftz method, can be used.[10] In this method, which is referred to as the wave based method (WBM), the steady-state dynamic variables in the acoustic subdomains are expressed in terms of a set of acoustic wave functions which are solutions of the homogeneous parts of the governing dynamic equations, completed with a particular solution function of the inhomogeneous equation. Since the functions are exact solutions of the governing equations, the contribution of a certain function in a set is merely determined by the acoustic boundary conditions. As only a finite number of functions can be considered, the boundary conditions can only be satisfied approximately. The contributions of the basis functions follow from a weighted residual formulation of the boundary conditions. This new prediction method has proven to have a better convergence rate than the finite element method for classical fluid-structure interaction problems.[10–12]

This article validates the WBM for analyzing the acoustic behavior of special setups with porous absorbing materials. The porous materials are assumed to behave as equivalent fluids. The final goal of this work is to get insight into the acoustic behavior of a combination of sound absorbing patches.

The first part of the article deals with the basic principles of the wave based method. A short description of the governing equations and boundary conditions and the set of wave functions in the expansions of the variables are given. Afterwards, a validation example is shown. The validation of the method is performed by means of measurements of the normal incidence sound absorption coefficient of different configurations in a large standing wave tube. In the last part, the method is applied to simulate some interesting setups of absorbing materials.

II. BASIC PRINCIPLES OF WBM

A. General acoustic problem

The geometry of the problem is depicted in Fig. 1. It consists of a three-dimensional patchwork pattern inserted in a rectangular room. The system is excited by an acoustic point source placed in the room. The patches are rectangular

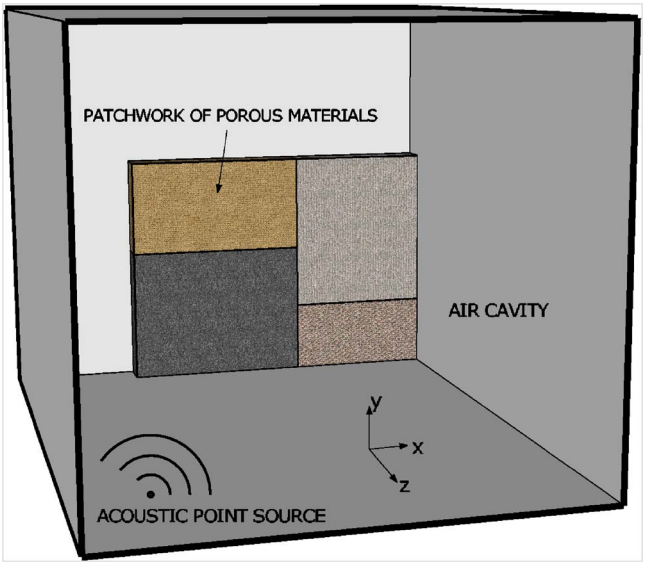


FIG. 1. (Color online). Geometry of the problem.

and each of them is made from a homogeneous porous material. In the following, a time dependence $e^{j\omega t}$ is assumed, with $j=\sqrt{-1}$ the imaginary unit.

1. Problem description

Consider the general problem displayed in Fig. 2. It consists of a cavity of volume $V=V^{(air)}\cup V^{(por)}$ subjected to Dirichlet, Neumann and mixed boundary conditions applied, respectively, on the boundaries Ω_p, Ω_v and Ω_Z. The $V^{(air)}$ cavity is filled with air with an ambient density ρ_o and a sound speed c_o. The cavity $V^{(por)}$ is filled with an equivalent fluid with a porosity h, a flow resistivity σ, a tortuosity α_∞ and characteristic viscous and thermal lengths Λ and Λ'. The fluid in the air cavity is excited by an acoustic volume velocity point source at position $\boldsymbol{r}_q$ with an amplitude q and a circular frequency ω. The steady-state pressure $p^{(air)}(\boldsymbol{r},\omega)$ and $p^{(por)}(\boldsymbol{r},\omega)$ in the air cavity and porous material, respectively, are denoted as $p^{(air)}(\boldsymbol{r})$ and $p^{(por)}(\boldsymbol{r})$, where $\boldsymbol{r}$ is the position vector of the considered point in the cavity V.

The pressure at any position $\boldsymbol{r}$ in the air cavity $V^{(air)}$ is governed by the inhomogeneous Helmholtz equation

$$\nabla^2 p^{(air)}(\boldsymbol{r})+k_{air}^2 p^{(air)}(\boldsymbol{r})=-A\,\delta(\boldsymbol{r},\boldsymbol{r}_q)\quad \boldsymbol{r}\in V^{(air)} \tag{1}$$

with ∇^2 the Laplace operator and $k_{air}=\omega/c_o$ the acoustic wave number in air. On the right-hand side of Eq. (1), δ is the Dirac function and A is equal to $j\omega\rho_o q$. The pressure at

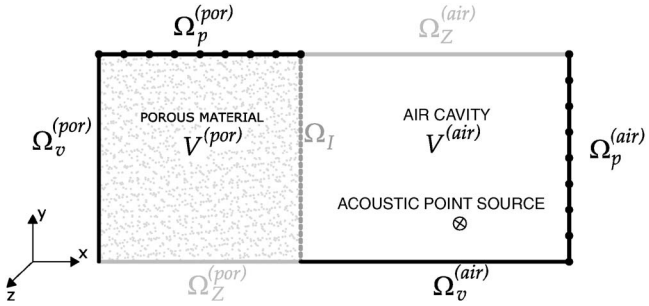


FIG. 2. Setup of the considered problem.

any position $\boldsymbol{r}$ in the porous material $V^{(por)}$ is governed by the homogeneous Helmholtz equation

$$\nabla^2 p^{(por)}(\boldsymbol{r}) + k_{por}^2 p^{(por)}(\boldsymbol{r}) = 0 \quad \boldsymbol{r} \in V^{(por)} \tag{2}$$

with $k_{por} = \omega / c_{por}(\omega) = \omega\sqrt{\tilde{\rho}(\omega) / \tilde{K}(\omega)}$ the acoustic wave number in the porous material. This wave number can be calculated based on the theory of Biot-Johnson-Allard (see the Appendix).

The boundary $\Omega^{(air)}$ of cavity $V^{(air)}$ consists of three parts. On the different parts $\Omega_p^{(air)}$, $\Omega_v^{(air)}$ and $\Omega_Z^{(air)}$, prescribed pressure, normal velocity and normal surface impedance distributions are imposed, respectively.

The boundary conditions can be expressed, respectively, as

$$\boldsymbol{r} \in \Omega_p^{(air)}: p^{(air)}(\boldsymbol{r}) = \bar{p}^{(air)}(\boldsymbol{r}) \tag{3}$$

$$\boldsymbol{r} \in \Omega_v^{(air)}: \frac{j}{\rho_o \omega} \frac{\partial p^{(air)}(\boldsymbol{r})}{\partial n^{(air)}} = \bar{v}_n^{(air)}(\boldsymbol{r}) \tag{4}$$

$$\boldsymbol{r} \in \Omega_Z^{(air)}: \frac{j}{\rho_o \omega} \frac{\partial p^{(air)}(\boldsymbol{r})}{\partial n^{(air)}} = \frac{p^{(air)}(\boldsymbol{r})}{\bar{Z}^{(air)}(\boldsymbol{r})}, \tag{5}$$

where $\bar{p}^{(air)}$, $\bar{v}_n^{(air)}$ and $\bar{Z}^{(air)}$ are the frequency-dependent imposed pressure, normal velocity and normal surface impedance and $\partial / \partial n$ is the derivative in the normal direction. On the boundary $\Omega^{(por)}$, similar conditions are imposed.

2. Coupling conditions

To avoid discontinuities in the dynamic variables at the interface between both subdomains, continuity conditions have to be imposed.

a. pv-coupling. A way of coupling the two cavities is imposing a pressure and normal velocity continuity. The coupling conditions between both regions at surface $\Omega_I^{(air,por)}$ can be written as follows:

$$\boldsymbol{r} \in \Omega_I^{(air,por)}: p^{(air)}(\boldsymbol{r}) = p^{(por)}(\boldsymbol{r}) \tag{6}$$

$$\boldsymbol{r} \in \Omega_I^{(air,por)}: \frac{j}{\rho_o \omega} \frac{\partial p^{(air)}(\boldsymbol{r})}{\partial n^{(air)}} = -\frac{jh}{\tilde{\rho}\omega} \frac{\partial p^{(por)}(\boldsymbol{r})}{\partial n^{(por)}}, \tag{7}$$

where $n^{(air)}$ and $n^{(por)}$ are the outward normals to, respectively, subdomain $V^{(air)}$ and subdomain $V^{(por)}$.

b. 'equivalent normal velocity' coupling. One can also enforce the continuity of a linear combination of both quantities. By dividing the pressure term by an acoustic impedance $\bar{Z}_{int}$, an equivalent normal velocity is obtained. The coupling conditions at the interface $\Omega_I^{(air,por)}$ become

$$\boldsymbol{r} \in \Omega_I^{(air,por)}: \frac{j}{\rho_o \omega} \frac{\partial p^{(air)}(\boldsymbol{r})}{\partial n^{(air)}} - \frac{1}{\bar{Z}_{int}} p^{(air)}(\boldsymbol{r}) = -\frac{jh}{\tilde{\rho}\omega} \frac{\partial p^{(por)}(\boldsymbol{r})}{\partial n^{(por)}} - \frac{1}{\bar{Z}_{int}} p^{(por)}(\boldsymbol{r}). \tag{8}$$

$$\boldsymbol{r} \in \Omega_I^{(air,por)}: \frac{jh}{\tilde{\rho}\omega} \frac{\partial p^{(por)}(\boldsymbol{r})}{\partial n^{(por)}} - \frac{1}{\bar{Z}_{int}} p^{(por)}(\boldsymbol{r}) = -\frac{j}{\rho_o \omega} \frac{\partial p^{(air)}(\boldsymbol{r})}{\partial n^{(air)}} - \frac{1}{\bar{Z}_{int}} p^{(air)}(\boldsymbol{r}). \tag{9}$$

The value of $\bar{Z}_{int}$ determines the relative importance of the velocity term compared to the pressure term in the equivalent velocity. Pluymers[13] showed that these coupling conditions result in a more stable numerical model since singularities which can occur at eigenfrequencies of the associated subdomain are avoided. In the following, these expressions, referred to as Z expressions, will be used.

3. Field variable approximations

The pressure in both domains is governed by Eqs. (1) and (2). The WBM describes the steady-state pressure in both domains as an expansion of wave functions Φ that exactly satisfy the governing Helmholtz equations. The steady-state acoustic pressure fields $p^{(air)}(\boldsymbol{r})$ and $p^{(por)}(\boldsymbol{r})$ are, respectively, approximated as $\hat{p}^{(air)}(\boldsymbol{r})$ and $\hat{p}^{(por)}(\boldsymbol{r})$, by truncating the number of solution functions to limited values of, respectively, $N^{(air)}$ and $N^{(por)}$ for the two subdomains

$$p^{(air)}(\boldsymbol{r}) \approx \hat{p}^{(air)}(\boldsymbol{r}) = \sum_{a=1}^{N^{(air)}} p_{w,a}^{(air)} \Phi_a^{(air)}(\boldsymbol{r}) + \hat{p}_q^{(air)}(\boldsymbol{r}) = \boldsymbol{\Phi}^{(air)}(\boldsymbol{r}) \boldsymbol{p}_{\boldsymbol{w}}^{(air)} + \hat{p}_q^{(air)}(\boldsymbol{r}). \tag{10}$$

$$p^{(por)}(\boldsymbol{r}) \approx \hat{p}^{(por)}(\boldsymbol{r}) = \sum_{b=1}^{N^{(por)}} p_{w,b}^{(por)} \Phi_b^{(por)}(\boldsymbol{r}) = \boldsymbol{\Phi}^{(por)}(\boldsymbol{r}) \boldsymbol{p}_{\boldsymbol{w}}^{(por)}. \tag{11}$$

Each function $\Phi_a^{(air)}(\boldsymbol{r})$ and $\Phi_b^{(por)}(\boldsymbol{r})$ in the respectively $(1 \times N^{(air)})$ and $(1 \times N^{(por)})$ row vectors $\boldsymbol{\Phi}^{(air)}(\boldsymbol{r})$ and $\boldsymbol{\Phi}^{(por)}(\boldsymbol{r})$ is a solution function of the homogeneous parts of Helmholtz Eqs. (1) and (2). $p_{w,a}^{(air)}$ and $p_{w,b}^{(por)}$ are the contribution factors for each of the selected waves $\Phi_a^{(air)}(\boldsymbol{r})$ and $\Phi_b^{(por)}(\boldsymbol{r})$. These weighting factors are collected in $(N^{(air)} \times 1)$ and $(N^{(por)} \times 1)$ vectors of degrees of freedom $\boldsymbol{p}_{\boldsymbol{w}}^{(air)}$ and $\boldsymbol{p}_{\boldsymbol{w}}^{(por)}$.

a. Homogeneous solution. The row vectors $\boldsymbol{\Phi}^{(\alpha)}(\boldsymbol{r})$, with α equal to *air* or *por* can be divided in a r, s and t set with $N^{(\alpha)} = N_r^{(\alpha)} + N_s^{(\alpha)} + N_t^{(\alpha)}$.

$$\Phi_r^{(\alpha)}(\boldsymbol{r}) = e^{-jk_{x,r}^{(\alpha)}(x - f_{x,r}^{(\alpha)} L_x^{(\alpha)})} \cos(k_{y,r}^{(\alpha)} y) \cos(k_{z,r}^{(\alpha)} z). \tag{12}$$

$$\Phi_s^{(\alpha)}(\boldsymbol{r}) = \cos(k_{x,s}^{(\alpha)} x) e^{-jk_{y,s}^{(\alpha)}(y - f_{y,s}^{(\alpha)} L_y^{(\alpha)})} \cos(k_{z,s}^{(\alpha)} z). \tag{13}$$

$$\Phi_t^{(\alpha)}(\boldsymbol{r}) = \cos(k_{x,t}^{(\alpha)} x) \cos(k_{y,t}^{(\alpha)} y) e^{-jk_{z,t}^{(\alpha)}(z - f_{z,t}^{(\alpha)} L_z^{(\alpha)})}, \tag{14}$$

with

$$k_\alpha^2 = (k_{x,\beta}^{(\alpha)})^2 + (k_{y,\beta}^{(\alpha)})^2 + (k_{z,\beta}^{(\alpha)})^2 \quad \text{for } \beta = r,s,t. \tag{15}$$

$L_x^{(\alpha)}$, $L_y^{(\alpha)}$ and $L_z^{(\alpha)}$ are the dimensions of the smallest rectangular bounding box circumscribing subdomain $V^{(\alpha)}$. The values $f_{\gamma,\beta}^{(\alpha)}$ are defined in Eq. (16). These parameters are scaling factors, which force the amplitudes of the solution functions to be smaller or equal to 1.

$$f_{\gamma,\beta}^{(\alpha)} = \begin{cases} 1 & \text{if } \Im[k_{\gamma,\beta}^{(\alpha)}] > 0 \\ 0 & \text{if } \Im[k_{\gamma,\beta}^{(\alpha)}] \le 0 \end{cases} \text{for } (\gamma,\beta) = (x,r),(y,s),(z,t), \tag{16}$$

where $\Im[\ ]$ denotes the imaginary part of the complex number between the brackets.

In order that the pressure expansions converge towards the exact solution, functions with the wave numbers given below are selected from the infinite number of possible wave functions

$$(k_r^{(\alpha)}) = \left(\pm\sqrt{k^2 - (k_{y,r}^{(\alpha)})^2 - (k_{z,r}^{(\alpha)})^2}, \frac{n_{y,r}^{(\alpha)}\pi}{L_y^{(\alpha)}}, \frac{n_{z,r}^{(\alpha)}\pi}{L_z^{(\alpha)}} \right) \tag{17}$$

with $n_{y,r}^{(\alpha)}, n_{z,r}^{(\alpha)} = 0,1,2,\ldots$.

$$(k_s^{(\alpha)}) = \left(\frac{n_{x,s}^{(\alpha)}\pi}{L_x^{(\alpha)}}, \pm\sqrt{k^2 - (k_{x,s}^{(\alpha)})^2 - (k_{z,s}^{(\alpha)})^2}, \frac{n_{z,s}^{(\alpha)}\pi}{L_z^{(\alpha)}} \right) \tag{18}$$

with $n_{x,s}^{(\alpha)}, n_{z,s}^{(\alpha)} = 0,1,2,\ldots$.

$$(k_t^{(\alpha)}) = \left(\frac{n_{x,t}^{(\alpha)}\pi}{L_x^{(\alpha)}}, \frac{n_{y,t}^{(\alpha)}\pi}{L_y^{(\alpha)}}, \pm\sqrt{k^2 - (k_{x,t}^{(\alpha)})^2 - (k_{y,t}^{(\alpha)})^2} \right) \tag{19}$$

with $n_{x,t}^{(\alpha)}, n_{y,t}^{(\alpha)} = 0,1,2,\ldots$.

Desmet[10] has shown that the pressure approximation based on this wave function set will converge.

b. Particular solution. In Eq. (10), $\hat{p}_q^{(air)}(\boldsymbol{r})$ represents a particular solution for the acoustic source term in the right-hand side of the inhomogeneous Helmholtz Eq. (1). It is the free field Green's function, the free field solution of a point source

$$\hat{p}_q^{(air)}(\boldsymbol{r}) = \frac{A}{4\pi} \cdot \frac{e^{-jkr_s}}{r_s} \tag{20}$$

with $r_s = |\boldsymbol{r} - \boldsymbol{r_q}| = \sqrt{(x-x_q)^2 + (y-y_q)^2 + (z-z_q)^2}$.

4. Weighted residual formulation of the boundary conditions

The contribution of each function $\Phi^{(\alpha)}(\boldsymbol{r})$ in the field variable expansion $\hat{p}^{(\alpha)}(\boldsymbol{r})$ is determined by the boundary conditions. Therefore, the coupling and boundary conditions are expressed in N_V weighted residual formulations, one for each subdomain. At the interface between two subdomains, one of the two coupling conditions will be imposed in the weighted residual formulation of each subdomain. In the following, $\Omega_I^{(air,por)}$ will be used if the interface is considered in subdomain $V^{(air)}$ and $\Omega_I^{(por,air)}$ if the interface is considered in subdomain $V^{(por)}$. The error functions following from the coupling conditions are

$$R_{\Omega_I^{(air,por)}} = \frac{j}{\rho_o\omega}\frac{\partial p^{(air)}(\boldsymbol{r})}{\partial n^{(air)}} - \frac{1}{\bar{Z}_{int}} p^{(air)}(\boldsymbol{r}) + \frac{jh}{\tilde{\rho}\omega}\frac{\partial p^{(por)}(\boldsymbol{r})}{\partial n^{(por)}} + \frac{1}{\bar{Z}_{int}} p^{(por)}(\boldsymbol{r}). \tag{21}$$

$$R_{\Omega_I^{(por,air)}} = \frac{jh}{\tilde{\rho}\omega}\frac{\partial p^{(por)}(\boldsymbol{r})}{\partial n^{(por)}} - \frac{1}{\bar{Z}_{int}} p^{(por)}(\boldsymbol{r}) + \frac{j}{\rho_o\omega}\frac{\partial p^{(air)}(\boldsymbol{r})}{\partial n^{(air)}} + \frac{1}{\bar{Z}_{int}} p^{(air)}(\boldsymbol{r}). \tag{22}$$

The error functions following from the boundary conditions are expressed as

$$R_p^{(\alpha)} = p^{(\alpha)}(\boldsymbol{r}) - \bar{p}^{(\alpha)}(\boldsymbol{r}), \quad \boldsymbol{r} \in \Omega_p^{(\alpha)}, \tag{23a}$$

$$R_v^{(\alpha)} = \frac{j}{\rho^{(\alpha)}\omega}\frac{\partial p^{(\alpha)}(\boldsymbol{r})}{\partial n^{(\alpha)}} - \bar{v}_n^{(\alpha)}(\boldsymbol{r}), \quad \boldsymbol{r} \in \Omega_v^{(\alpha)}, \tag{23b}$$

$$R_Z^{(\alpha)} = \frac{j}{\rho^{(\alpha)}\omega}\frac{\partial p^{(\alpha)}(\boldsymbol{r})}{\partial n^{(\alpha)}} - \frac{p^{(\alpha)}(\boldsymbol{r})}{\bar{Z}^{(\alpha)}(\boldsymbol{r})}, \quad \boldsymbol{r} \in \Omega_Z^{(\alpha)}, \tag{23c}$$

where $\rho^{(\alpha)}$ is equal to ρ_o or $\tilde{\rho}$ depending on the assumed subdomain. For each subdomain $V^{(\alpha)}$, the approximation errors are orthogonalized with respect to a weighting function $\tilde{p}^{(\alpha)}$ or its derivative. The weighted residual formulation of subdomain $V^{(\alpha)}$ can be found in Eq. (24). Here, the position dependency of the vectors is omitted for the sake of conciseness

$$0 = \int_{\Omega_I^{(\alpha,\beta)}} \tilde{p}^{(\alpha)} R_{\Omega_I^{(\alpha,\beta)}} d\Omega_I^{(\alpha,\beta)} - \int_{\Omega_p^{(\alpha)}} \frac{j}{\rho^{(\alpha)}\omega}\frac{\partial \tilde{p}^{(\alpha)}}{\partial n^{(\alpha)}} R_p^{(\alpha)} d\Omega_p^{(\alpha)} + \int_{\Omega_v^{(\alpha)}} \tilde{p}^{(\alpha)} R_v^{(\alpha)} d\Omega_v^{(\alpha)} + \int_{\Omega_Z^{(\alpha)}} \tilde{p}^{(\alpha)} R_Z^{(\alpha)} d\Omega_Z^{(\alpha)} \quad \text{for } \boldsymbol{r} \in V^{(\alpha)}. \tag{24}$$

5. Matrix representation

If the same set of acoustic wave functions as used in the pressure expansions [see Eqs. (12)–(14)] is used as weighting functions in the N_V weighted residual formulations of boundary and coupling conditions, a matrix representation can be obtained. The contributions of the different solution functions can be deduced from that matrix equation.

III. VALIDATION MEASUREMENTS

A. Validation measurement setup

The goal of this part is the experimental validation of the wave based method. Measurements were performed in the Ecole Nationale des Travaux Publics de l'Etat in a rectangular concrete standing wave tube. This tube has an inner section of 0.6×0.6 m^2 and a length of 5.5 m. These measurements are compared to simulations made with the WBM.

If the excitation is symmetric, the generation of the asymmetric modes will be limited. Furthermore, the contribution of these asymmetric modes to the pressure field in the center of the tube is theoretically equal to zero and thus, one can shift the cut-off frequency to $f_{02} \approx 570$ Hz [mode (2,0,0)] by measuring the pressure in the center of the tube. Thus, the

FIG. 3. (Color online). Validation measurement setup.

standing wave tube can be used at low frequencies for measurements of the absorption coefficient at normal incidence of samples with large dimensions.

In the standing wave tube, used for these experiments, the air is excited with periodic random noise generated by a system of 4 Audax loudspeakers placed symmetrically compared to the axis of the tube. It was shown experimentally that a plane wave is achieved at less than 0.5 m away from the sound source system. Since the precision of the transfer function measurement in the tube is the best of at a microphone gap equal to $\frac{\lambda}{4}$,[14] a moving 1/2 in. B&K microphone, that measures the transfer function at a few positions, is used. The movement of the microphone is controlled by a step motor and a sensor in the same plane as the microphone which guarantees the precise determination of the location of the material plane. With this configuration, the uncertainty about the location of the measurement point can be minimized. A picture of the loudspeaker setup and the moving microphone can be found in Fig. 3.

These measurements are used to determine the absorption coefficient in a large frequency range. For each frequency, the ideal microphone positions are determined and the absorption coefficient is calculated based on the measurements at the positions that approximate the best of these positions.

B. Material description

The tube is used to measure the normal incidence sound absorption coefficient of different setups with an open-cell melamine foam. Its properties are given in Table I. The material properties of the melamine were measured with the appropriate techniques. Descriptions of the methods to measure the characteristic lengths are given by Leclaire *et al.*[15] The techniques to measure porosity and tortuosity are described by Fellah *et al.*[16] The method to measure flow resistivity is given in ISO-9053.[17] These properties were used to perform the numerical simulations. The theory of Johnson–Allard (see Appendix) can be used to calculate the propagation parameters in the melamine foam.

Open-cell melamine foam exhibits an elastic behavior at low frequencies. Therefore, a more sophisticated model is required to describe the wave propagation inside the material (Biot's model), at these frequencies. In this text, the melamine is modeled as an equivalent fluid for the whole frequency range, since this model describes the behavior of the material thoroughly except around the frequencies where the skeleton resonates.

C. Measurement configurations

Samples with a thickness of 50 mm and 100 mm are used. In Fig. 4, the different setups which are used to validate the numerical simulation technique can be seen.

Cases 1 and 2 are, respectively, a 50-mm- and 100-mm-thick open-cell melamine foam glued directly on the hard termination of the standing wave tube. Case 3 is a 50-mm-thick melamine plate backed by a 50-mm-thick air layer and a hard wall. Case 4 is a combination of cases 2 and 3. Case 5 is a simulation of a porous material placed on a variable backing. The plenum behind the porous material varies from 0 to 50 mm. This case is the combination of cases 1 and 3.

D. Numerical implementation

1. Truncation rule

For numerical implementation, $n_{y,r}^{(\alpha)}$, $n_{z,r}^{(\alpha)}$, $n_{x,s}^{(\alpha)}$, $n_{z,s}^{(\alpha)}$, $n_{x,t}^{(\alpha)}$ and $n_{y,t}^{(\alpha)}$ in Eqs. (17)–(19) have to be truncated. The following truncation is proposed:

$$\frac{n_{y,r_{max}}^{(\alpha)}}{L_y^{(\alpha)}} \approx \frac{n_{z,r_{max}}^{(\alpha)}}{L_z^{(\alpha)}} \approx \frac{n_{x,s_{max}}^{(\alpha)}}{L_x^{(\alpha)}} \approx \frac{n_{z,s_{max}}^{(\alpha)}}{L_z^{(\alpha)}} \approx \frac{n_{x,t_{max}}^{(\alpha)}}{L_x^{(\alpha)}} \approx \frac{n_{y,t_{max}}^{(\alpha)}}{L_y^{(\alpha)}} \geq \max_{\alpha=1}^{N_V}\left(\frac{n_{tr}k^{(\alpha)}}{\pi}\right) \tag{25}$$

with $k^{(\alpha)}=\omega/c^{(\alpha)}$ the wave number in subdomain $V^{(\alpha)}$ and n_{tr} a truncation factor. The truncation rule expresses that the wave number components of the wave functions are smaller than or about equal to the physical wave number in the considered subdomain times the truncation factor n_{tr}. Unless otherwise stated, a truncation factor $n_{tr}=2$ will be used throughout this text, since numerical experiments have

TABLE I. Properties of melamine.

Open-cell melamine foam	
Porosity	$h=0.98$
Tortuosity	$\alpha_\infty=1.1$
Flow resistivity	$\sigma=10{,}000\ \mathrm{Nm^{-4}\,s}$
Viscous characteristic length	$\Lambda=100\ \mu\mathrm{m}$
Thermal characteristic length	$\Lambda'=150\ \mu\mathrm{m}$
Air cavity	
Sound speed	$c_0=340\ \mathrm{ms^{-1}}$
Density	$\rho_0=1.225\ \mathrm{kg\,m^{-3}}$

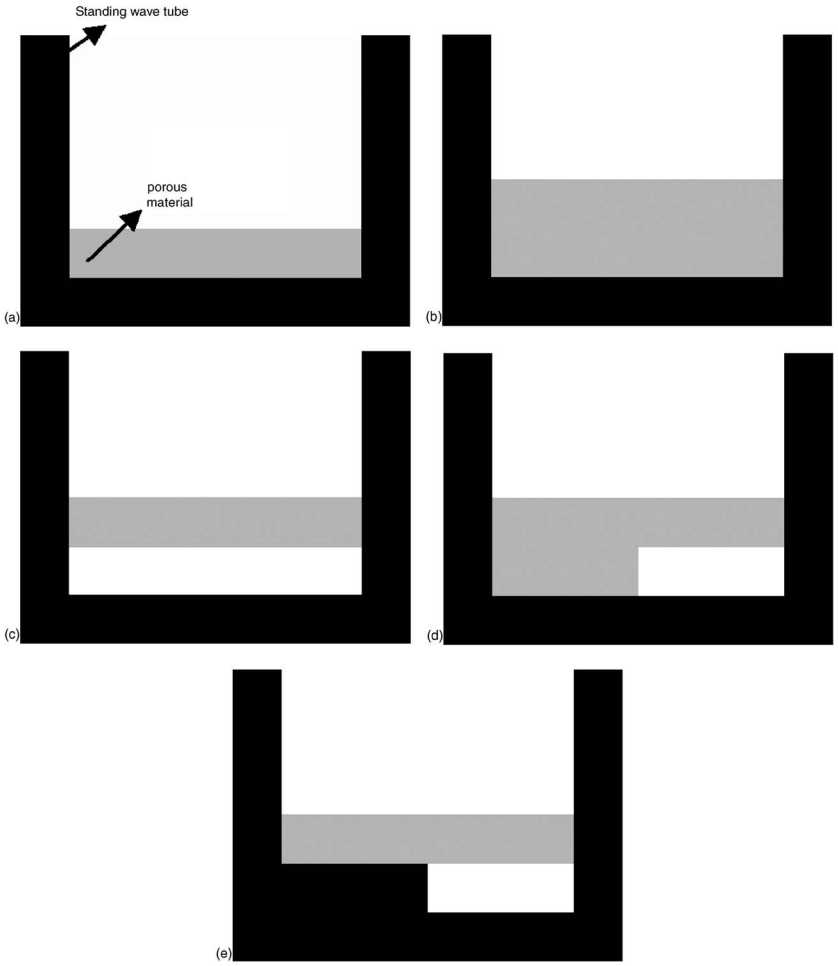


FIG. 4. Measurements configurations: (a) case 1, (b) case 2, (c) case 3, (d) case 4, and (e) case 5.

shown that this factor is sufficient to achieve convergence of the solution.

2. Coupling conditions

For large values of the coupling factor $\bar{Z}_{int}$, the normal velocity component dominates in the continuity conditions (8) and (9). For small values, the pressure component is dominant. Choosing a zero or infinite value for $\bar{Z}_{int}$ results in identical continuity conditions and yields a singular wave based model. Pluymers[13] suggests to choose the coupling factor to be the characteristic impedance factor of air, i.e., $\bar{Z}_{int}=\rho_o c_o$. In the numerical simulations, $\bar{Z}_{int}$ is thus chosen equal to $340\ \frac{\mathrm{m}}{\mathrm{s}} \times 1.225\ \mathrm{kg/m^3} = 416.5\ \frac{\mathrm{Pas}}{\mathrm{m}}$.

3. Calculation of the absorption coefficient

The normal incidence sound absorption coefficient is calculated from the sound field in the center of the tube. Based on the ratio s of the minimum and maximum pressure, the absorption coefficient of the sample can be calculated with Eq. (26) for each frequency of interest.

$$\alpha_n = \frac{4}{s + \frac{1}{s} + 2}. \tag{26}$$

E. Measurement results

Validation measurements for the cases mentioned in Sec. III C can be seen in Figs. 5 and 6.

Figure 5 shows a comparison of WBM simulations with the measurement results of cases 2, 3 and 4. It is shown that there is a quite good agreement between the measurements and the computations. At low frequencies, however, a reso-

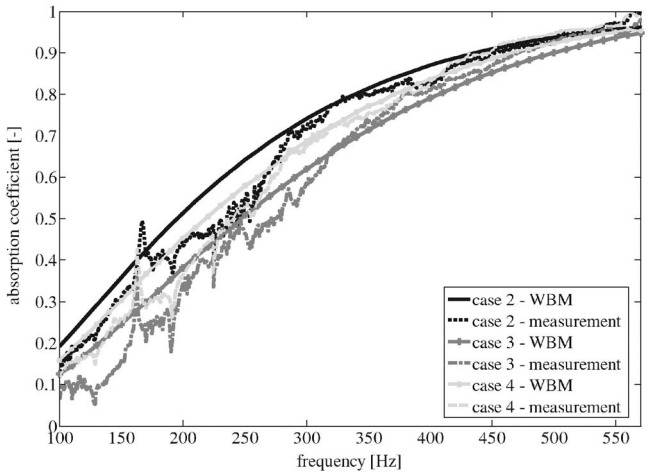


FIG. 5. Comparison of WBM simulation and measurement for cases 2, 3 and 4.

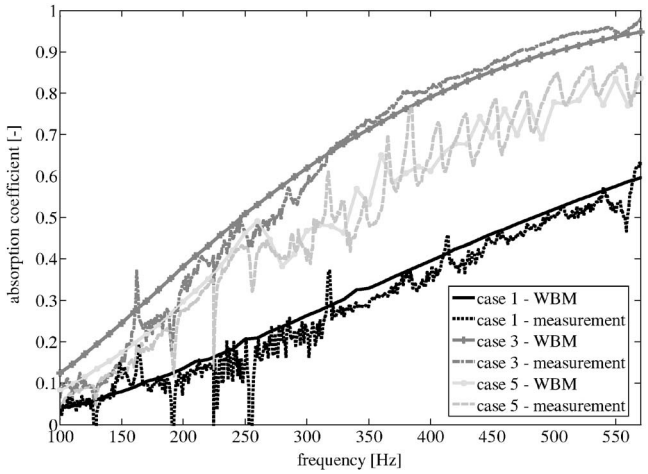


FIG. 6. Comparison of WBM simulation and measurement for cases 1, 3 and 5.

nance of the melamine skeleton is noticeable in the measurement results. This skeleton resonance is not accounted for in the equivalent fluid model which is used in the simulations. The same is observed in Fig. 6 which displays the comparison of the absorption coefficient for cases 1, 3 and 5. An existence of oscillations in the absorption coefficient at higher frequencies is observed for case 5. This behavior will be the subject for further research. A comparison of the WBM solution with a transfer matrix method solution was performed for cases 1, 2 and 3. The curves for the different cases matched very well.

IV. CASE STUDIES

In this section, the developed tool is used to simulate some interesting setups of absorbing materials.

A. Validity of the local reactivity assumption

The effect of a layer of sound absorbing porous material on the sound field in a room is often modeled as a normal impedance boundary condition for the pressure field in the room. By modeling the sound absorbing material in this way, the material is assumed to be locally reacting. Since the porous layer can be modeled as an equivalent fluid in the wave based model, the effect of the porous layer on the sound field in the room can be modeled without making the assumption of the locally reacting behavior of the sound absorbing material. In this section, the validity of the assumption of local reactivity for different sound absorbing materials will be verified.

1. Problem description

Figure 7 depicts the two-dimensional (2D) acoustic problem considered in this study. The bottom of the air-filled cavity is covered with a 50-mm-thick porous layer with a 50-mm-thick air plenum behind it. The other walls are perfectly rigid. The plenum and air cavity are filled with air with a sound velocity $c_o=340\ \frac{\mathrm{m}}{\mathrm{s}}$ and a density $\rho_o=1.225\ \frac{\mathrm{kg}}{\mathrm{m}^3}$. A

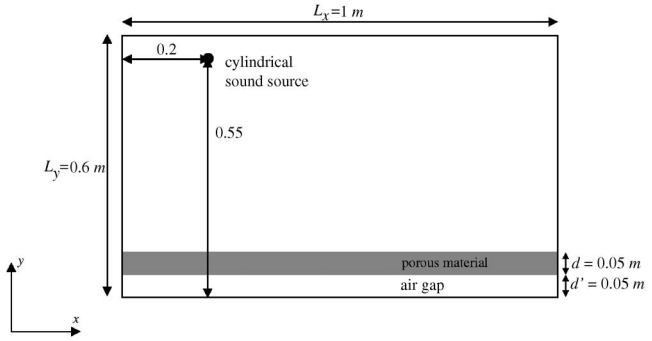


FIG. 7. 2D acoustic problem—comparison of the local reactivity assumption and the complete modeling of an absorbing structure.

cylindrical acoustic volume velocity source excites the air in the cavity. This source has a volume velocity $q=0.01\ \mathrm{s}^{-1}$ and is positioned at location (0.2, 0.55).

2. Surface impedance

If the sound absorbing structure is not fully modeled, a normal impedance boundary condition has to be imposed to take its influence into account. This impedance is taken equal to the impedance at normal sound incidence of the combination of the porous layer and the plenum and can be calculated with the equations given by Allard.[1]

3. Numerical simulations

The dynamic response of the cavity due to the presence of the cylindrical acoustic source is predicted with the wave based prediction technique. To verify the validity of the spatial invariance assumption of the absorbing structure, the amplitude of the normal impedance along the air-porous layer interface above the sound absorbing structure can be plotted. The properties of the first absorbing layer used in the simulations can be found in Table I. The second type of absorbing layer has the same properties except the flow resistivity, which is taken equal to 100,000 $\frac{\mathrm{Ns}}{\mathrm{m}^4}$. This is a hypothetic material, the behavior of a material with this flow resistivity has probably to be modeled with the full Biot model in reality.

The normal impedance calculated at the interface between the air cavity and the foam layer in both situations is plotted in Fig. 8 for arbitrarily chosen frequencies 100, 500, and 900 Hz. The solid lines are the simulations calculated with a wave based model with a full modeling of the foam layer and the air gap behind it. The dashed lines are the spatially invariant normal impedances, calculated with the equations given by Allard.[1]

It is clear from Figs. 8(a) and 8(b) that the spatial variation of the normal impedance at the interface between the air cavity and the absorber cannot be neglected at 100 Hz. On the other hand, a much better approximation of a spatially invariant impedance can be seen at 500 and 900 Hz for a material with a high flow resistivity, here equal to 100,000 $\frac{\mathrm{Ns}}{\mathrm{m}^4}$. For the other material, the spatial variation of the normal impedance at the interface is pronounced and certainly cannot be neglected at these frequencies.

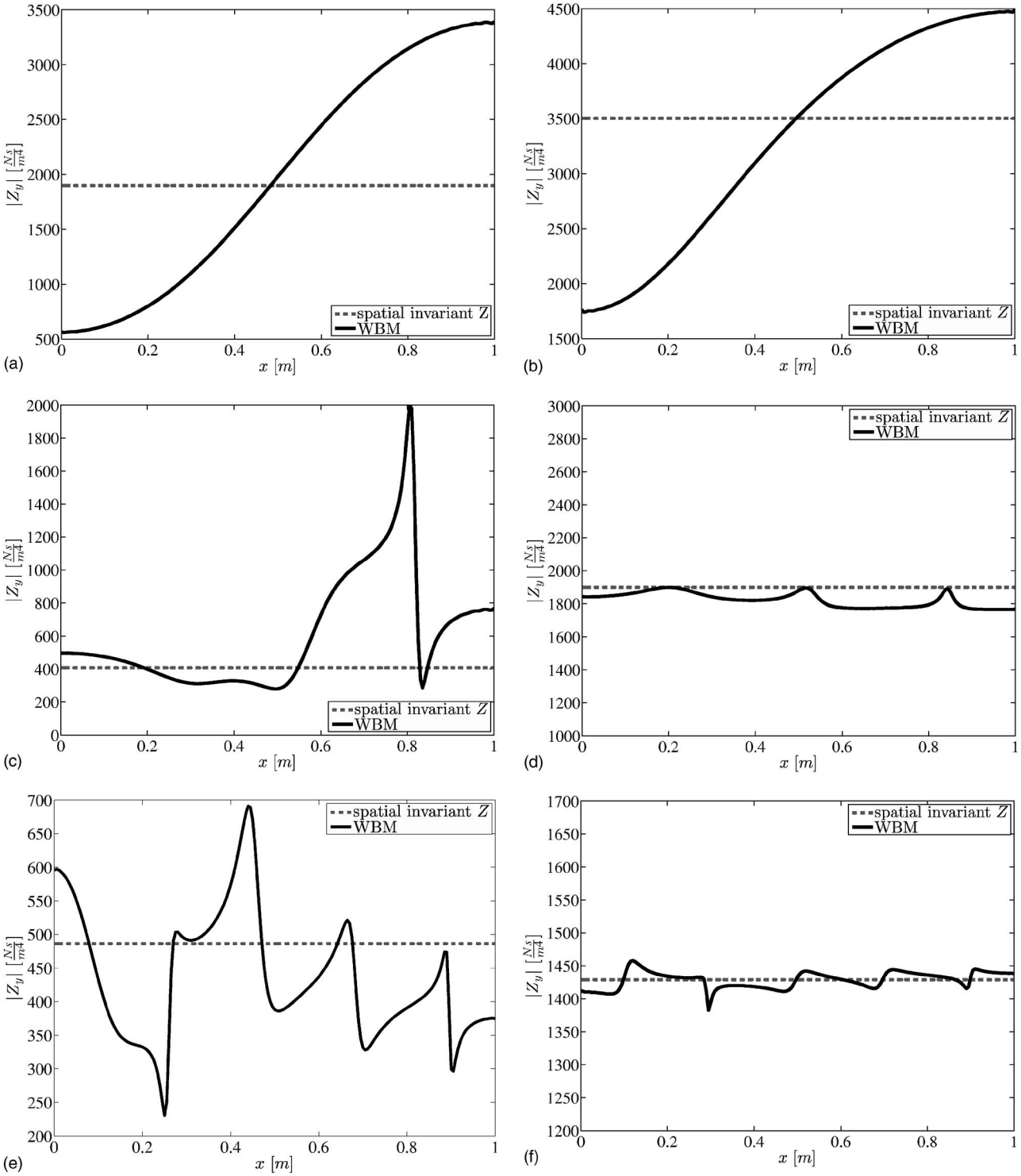


FIG. 8. Amplitudes of the normal impedances at the interface between the air cavity and the porous absorbing layer: (a) 100 Hz$-\sigma=10{,}000\ \frac{\mathrm{Ns}}{\mathrm{m}^4}$, (b) 100 Hz$-\sigma=100{,}000\ \frac{\mathrm{Ns}}{\mathrm{m}^4}$, (c) 500 Hz$-\sigma=10{,}000\ \frac{\mathrm{Ns}}{\mathrm{m}^4}$, (d) 500 Hz$-\sigma=100{,}000\ \frac{\mathrm{Ns}}{\mathrm{m}^4}$, (e) 900 Hz$-\sigma=10{,}000\ \frac{\mathrm{Ns}}{\mathrm{m}^4}$, and (f) 900 Hz$-\sigma=100{,}000\ \frac{\mathrm{Ns}}{\mathrm{m}^4}$.

The assumption of a spatially invariant normal impedance can induce substantial approximation errors for the pressure field in the whole cavity above the absorbing medium. The size of the induced errors depends on the frequency range of interest and the type of sound absorbing material.

B. Positioning of the sound absorbing material

It is a common question in acoustical engineering if an optimal position of sound absorbing material in a room can be indicated. In this section, a short 2D study is performed to show the usefulness of the wave based method to simulate the changes in sound field induced by the modification of the absorber location.

Consider a room, filled with air with a sound velocity $c_o=340\ \frac{\mathrm{m}}{\mathrm{s}}$ and a density $\rho_o=1.225\ \frac{\mathrm{kg}}{\mathrm{m}^3}$. The dimensions of the room and the position of the sound source are indicated in Fig. 9. A certain amount of absorbing material, with a thickness of 100 mm and properties indicated in Table I, is positioned in the room at two different positions. Figures 9(a)

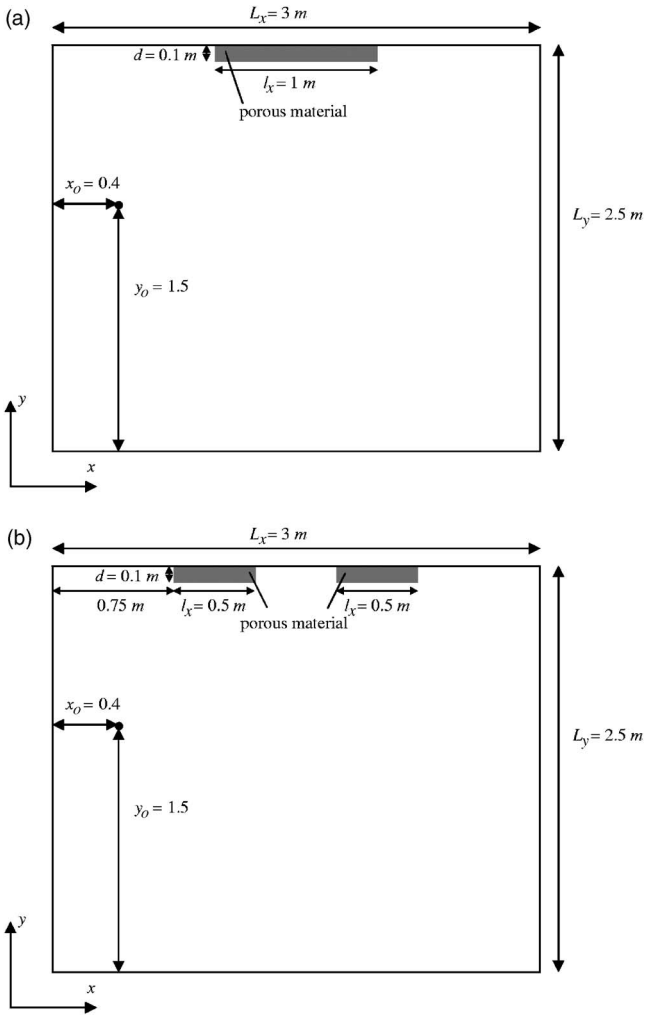


FIG. 9. Configuration of sound absorbing material in a room: (a) configuration 1, (b) configuration 2.

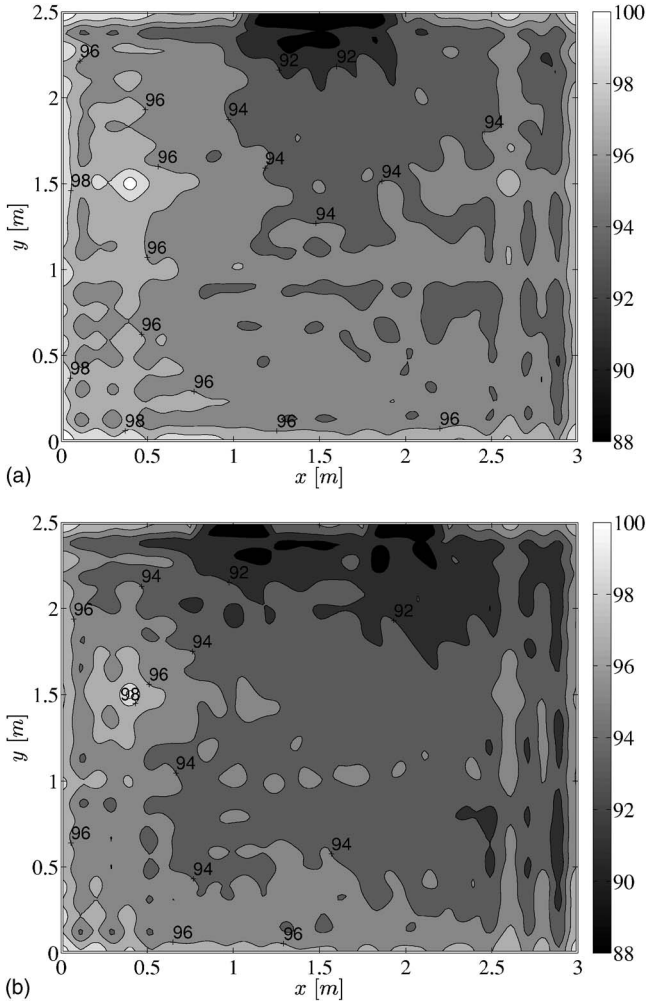


FIG. 10. Mean pressure level $L_{p,m}$ $[dBre.2\cdot10^{-5}$ Pa]. (a) configuration 1, (b) configuration 2.

and 9(b) indicate, respectively, the first and second configuration of the sound absorbing porous material. In the first configuration, the layer is positioned in the center of the ceiling. In the second configuration, the same amount of material is used, but divided in two patches and the patches are placed 0.5 m apart from one another.

1. Mean pressure level contour plots

During this study, use has been made of a frequency dependent number of wave functions, with a truncation factor $n_{tr}=4$. For both configurations, the sound field in the room was simulated for frequencies from 100 Hz up to 1000 Hz in steps of 1 Hz. The sound field in the rooms is given in terms of the mean pressure level $L_{p,m}$, averaged over all frequencies. $L_{p,m}$ is defined as

$$L_{p,m}(x,y) = 10\log\left(\frac{\frac{1}{n_f}\sum_{i=1}^{n_f}|p(x,y,f_i)|^2}{(2\cdot10^{-5})^2}\right). \quad (27)$$

In order to obtain a realistic simulation of the sound field, some damping is introduced in the setup. In the air cavity, the wave number is written as $k=\frac{\omega}{c_o}-j0.0002\frac{\omega}{c_o}$. Figures 10(a) and 10(b) display, respectively, the spatial distribution of the mean pressure level $L_{p,m}$ on a 20 mm grid.

It is clear from these figures that the mean pressure level in the considered frequency range is lower when the material is divided in two patches, especially in the center of the room and close to the material, the effect can be seen. The difference in sound level between both situations varies between 0 and 2 dB.

V. CONCLUSIONS

In this paper, a numerical method based on Trefftz's approach has been proposed to predict the sound field above a patchwork of rigid frame porous materials. In this wave based method, the steady-state dynamic variables in the studied domain are described in terms of a set of wave functions, solutions of the homogeneous part of the governing equations, together with a particular solution function for the inhomogeneous part. A weighted residual formulation of the coupling and boundary conditions allows for the determination of the contributions of each wave function. The sound field in each subdomain can then be obtained.

The method has been validated in several configurations of materials excited by a normal incidence plane wave. The

calculations have been successfully compared to experimental measurements carried out in a standing wave tube. Afterwards, the developed tool was applied to some real-life problems that illustrate the broad applicability of the proposed deterministic prediction technique. The perspective of this work is to get physical insight into the behavior of patchworks of different materials excited in a complex manner and to predict the absorbing qualities of patchworks for normal and oblique incidence.

ACKNOWLEDGMENTS

R.L. is a Research Assistant of the Fund for Scientific Research-Flanders (Belgium). The authors are grateful to the team of the Laboratoire des Sciences de l'Habitat of ENTPE, Lyon, who made it possible to perform the validation measurements.

APPENDIX: JOHNSON–ALLARD THEORY

Several theories can be used to calculate the wave number and complex density of a porous material with a motionless frame at a circular frequency ω. In this paper, the theory of Johnson–Allard[1] is chosen. The complex density is given by Eq. (A1)

$$\tilde{\rho}(\omega) = \alpha_\infty \rho_{air} - \frac{j\sigma h}{\omega}\tilde{G}(\omega) \tag{A1}$$

with

$$\tilde{G}(\omega) = \sqrt{1 + \frac{4j\alpha_\infty^2 \eta \rho_{air}\omega}{\sigma^2 \Lambda^2 h^2}} \tag{A2}$$

and σ the flow resistivity, h the porosity, α_∞ the tortuosity, Λ the viscous characteristic length and η the dynamic viscosity of air.

The bulk modulus is given by

$$\tilde{K}(\omega) = \frac{\gamma P_o}{\gamma - (\gamma - 1)\left[1 + \frac{8\eta}{jPr\omega\rho_{air}\Lambda'^2}\sqrt{1 + \frac{j\rho_{air}\omega Pr\Lambda'^2}{16\eta}}\right]^{-1}} \tag{A3}$$

with Pr the Prandtl number and Λ' the thermal characteristic length.

The wave number $k_{por} = \frac{\omega}{c(\omega)} = \omega\sqrt{\frac{\tilde{\rho}(\omega)}{\tilde{K}(\omega)}}$ in the porous material can then be calculated using Eqs. (A1) and (A3).

[1] J. F. Allard, *Propagation of Sound in Porous Media, Modeling of Sound Absorbing Materials* (Elsevier, Amsterdam, 1999).
[2] M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid," J. Acoust. Soc. Am. **28**, 2495–2510 (1956).
[3] X. Olny and C. Boutin, "Acoustic wave propagation in double porosity media," J. Acoust. Soc. Am. **114**(1), 73–89 (2003).
[4] N. Atalla, F. Sgard, X. Olny, and R. Panneton, "Acoustic absorption of macro-perforated porous materials," J. Sound Vib. **243**(4), 659–678 (2001).
[5] F. Sgard, X. Olny, and N. Atalla, "On the use of perforations to improve the sound absorption of porous materials," Appl. Acoust. **66**(6), 625–651 (2005).
[6] N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**(3), 1444–1452 (1998).
[7] C. Boutin, "Dynamique des milieux poreux saturés déformables - Fonctions de Green—Perméamètre dynamique" (Dynamic behavior of porous saturated deformable media—Greens functions—Dynamic permeameter), Ph.D. thesis, Université Scientifique, Technologique et Médicale de Grenoble (1987).
[8] O. Tanneau, P. Lamary, and Y. Chevalier, "A boundary element method for porous media," J. Acoust. Soc. Am. **120**(3), 1239–1251 (2006).
[9] F. Sgard, "Modélisation par éléments finis des structures multi-couches complexes dans le domaine des basses fréquences" (Finite element modeling of complex multi-layered structures at low frequencies), thesis to demonstrate the capacity of supervising research in science (in French), INSA of Lyon - University Lyon I (2002).
[10] W. Desmet, "A wave based prediction technique for coupled vibro-acoustic analysis," Ph.D. thesis, K. U. Leuven, retrieved July 17, 2007 from http://www.mech.kuleuven.be/dept/resources/docs/desmet.pdf (1998).
[11] W. Desmet, B. van Hal, P. Sas, and D. Vandepitte, "A computationally efficient prediction technique for the steady-state dynamic analysis of coupled vibro-acoustic systems," Adv. Eng. Software **33**, 527–540 (2002).
[12] B. Pluymers, W. Desmet, and D. Vandepitte, "Application of an efficient wave-based prediction technique for the analysis of vibro-acoustic radiation problems," J. Comput. Appl. Math. **168**(1-2), 353–364 (2004).
[13] B. Pluymers, "Wave based modeling methods for steady-state vibro-acoustics," Ph.D. thesis, K. U. Leuven, retrieved July 17, 2007 from http://people.mech.kuleuven.be/~bpluymer/docs/thesis_bpluymers.pdf (2006).
[14] X. Olny, "Absorption acoustique des milieux à simple et double porosité. Modélisation et validation expérimentale" (Acoustic absorption of single and double porosity media. Modeling and experimental validation), Ph.D. thesis, INSA of Lyon - ENTPE (1999).
[15] P. Leclaire, L. Kelders, W. Lauriks, C. Glorieux, and J. Thoen, "Determination of the viscous characteristic length in air-filled porous materials by ultrasonic attenuation measurements," J. Acoust. Soc. Am. **99**(4), 1944–1948 (1996).
[16] Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristégui, and J.-Y. Chapelon, "Measuring the porosity and the tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **113**(5), 2424–2433 (2003).
[17] ISO-9053:1991, "Acoustics materials for acoustical applications-determination of air-flow resistance" International Organization for Standardization, Geneva (1999).

Effect of room absorption on human vocal output in multitalker situations

Lau Nijs,[a] Konca Saher,[b] and Daniël den Ouden[c]
Faculty of Architecture, Delft University of Technology, Berlageweg 1, 2628 CR Delft, The Netherlands

(Received 5 December 2006; revised 1 November 2007; accepted 6 November 2007)

People increase their vocal output in noisy environments. This is known as the Lombard effect. The aim of the present study was to measure the effect as a function of the absorption coefficient. The noise source was generated by using other talkers in the room. A-weighted sound levels were measured in a 108 m^3 test room. The number of talkers varied from one to four and the absorption coefficients from 0.12 to 0.64. A model was introduced based on the logarithmic sum of the level found in an anechoic room plus the increasing portion of noise levels up to 80 dB. Results show that the model fits the measurements when a maximum slope of 0.5 dB per 1.0 dB increase in background level is used. Hence Lombard slopes vary from 0.2 dB/dB at 50 dB background level to 0.5 dB/dB at 80 dB. In addition, both measurements and the model predict a decrease of 5.5 dB per doubling of absorbing area in a room when the number of talkers is constant. Sound pressure levels increase for a doubling of talkers from 3 dB for low densities to 6 dB for dense crowds. Finally, there was correspondence between the model estimation and previous measurements reported in the literature. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821410]

PACS number(s): 43.55.Dt, 43.55.Hy, 43.70.Bk [RYL] Pages: 803–813

I. INTRODUCTION

In 1911, Lombard published an article on the effect of people tending to raise their voices in noisy environments, now known as the Lombard effect.[1] Since then many researchers have tried to establish the increase in human vocal output as a function of noise level, type of noise, etc. Lane and Tranel conducted an overview of the literature in 1971[2] containing over 200 citations. They found different slopes but the majority of the results is close to an increase in vocal output of approximately 0.5 dB per 1.0 dB increase in noise level. This will be denoted by 0.5 dB/dB throughout this paper, where all sound levels are assumed to be A-weighted. For low noise levels, below about 50 dB, lower slopes are found and lower slopes are also found for very high levels, over 100 dB, since there is a maximum to the level of human speech.

Lane and Tranel focus heavily on the feedback loops used by a speaker. They reject the hypothesis that a speaker reacts to the sound level of her or his own voice. Speakers rather use an internal "private" loop, based on articulatory processes but also on the tension of muscles within the body, plus an external "public" loop, which is based on the response a speaker gets from the listener about the intelligibility of her or his speech. However, there is hardly any information in the paper of the building acoustics parameters that influence the vocal output as, for instance, the amount of absorption or the reverberation time. Also the distance between talker and listener, which has a strong influence on the speech level, is treated in a few sentences only, while it may be an important part of the public loop as well.

In the same year, 1971, the results of measurements by Gardner were published.[3] Gardner measured sound levels as a function of the number of people present in dining rooms and auditoria, but he also measured the absorbing areas of the rooms. He also found an increase in vocal output of approximately 0.5 dB(A) per dB(A) increase in background sound level caused by the other talkers in the room.

In 1977, Pearsons *et al.* summarized different results from measurements at different noise types. Their modal Lombard slope was somewhat higher and given as 0.6 dB/dB.[4,5] Recently, Hodgson *et al.* have published results based on measurements taken in ten eating establishments with different acoustic characteritics.[6] The values differed from 0.40 to 2.61 dB/dB but the type of background noise was very varied, since, for instance, loud café music was also taken into account.

Sound levels in a room are affected by the total absorbing area in that room. According to the principles of acoustic theory a decrease of approximately 3 to 4 dB is found if the total absorbing area is doubled when the output power of the sound source is kept constant. In a multitalker environment, however, the output powers of the human sources depend on the sound level, so a self-reinforcing effect occurs and higher decreases are found. The Lombard effect as a function of the absorbing area has not been thoroughly investigated in the literature. Examples may be found from consulting practice of "before and after measurements" where absorption is added to improve reverberant situations. Oberdörster and Tiesler, for instance, compared two similar rooms with and without a sound-absorbing ceiling.[7]

From these findings acoustic consultants were able to derive their own rule of thumb: The sound pressure level (SPL) of multitalker speech increases by 6 dB when the re-

[a]Author to whom correspondence should be addressed. Electronic mail: l.nijs@tudelft.nl.
[b]Electronic mail: k.saher@tudelft.nl.
[c]Electronic mail: d.denouden@tudelft.nl.

verberation time (RT) is doubled. Hodgson *et al.* have developed a more general formula for the vocal output of teachers by comparing university classrooms with different absorption, and giving talker sound power as a function of the logarithm of the total absorption in square meters.[8,9] The slope in their curve (see Sec. II D for more details) is given as −9.6, which means that the Lombard effect is close to 0.5 dB/dB.

We do not know of any investigations in which vocal output is measured for a series of absorption coefficients. This lack of findings from measurements in different absorption conditions is probably due to difficulties in changing the absorbing area. We were fortunate to be able to take measurements in a test facility specially designed for the new building of the Conservatory of Amsterdam. It was designed to facilitate the study of "ideal" reverberation times for music teaching and practice. The reverberation time could be varied between 1.36 and 0.21 s. This project has been and will be discussed in other articles.[10] It is the aim of this paper to present sound power levels as a function of the amount of absorption and the number of talkers in a room. From these results a simple equation will be derived that can be used in the architectural design process.

The reason we carried out the measurements is that we try to develop architectural guidelines for architects on acoustical quality in institutions for people with intellectual disabilities. It is part of research at the Faculty of Architecture at Delft University. Institutions often comprise groups of eight to ten residents and there are frequent occasions in which a "multitalker situation" occurs if two or more people talk simultaneously. The combination of room shape and sound absorption in particular is helpful in reducing sound levels,[11] but the present paper focuses on one aspect alone: The sound level of human vocal output in relation to the amount of sound absorbing material in a rectangular room in which two or more people are talking simultaneously. Both the absorption coefficient and the number of talkers in a room are taken into account. Although the number of talkers in institutions is mainly from one to four, a simple design equation will be developed and compared to earlier results from literature where the number of talkers may be as high as 100. Vocal output will be restricted to so-called "normal" conversation, which is often below the levels of speech, and sometimes screaming, found in institutions for individuals with intellectual disabilities.

II. THEORY

A. Reverberation time and sound pressure level

The RT used throughout this paper follows Sabine's equation:

$$\mathrm{RT} = \frac{55.3V}{cA} = \frac{0.16V}{A}, \tag{1}$$

where V is the total volume of the room and c the speed of sound. If c is taken as 343 m/s the value 0.16 emerges. The influence of air absorption is omitted throughout the remainder of this paper. The surface area A represents the total sound absorbing area in the room. It is found as the sum over all surfaces in the room:

$$A = \sum_i \alpha_i S_i. \tag{2}$$

For the total room the average value of the absorption coefficient can be calculated as follows:

$$S_{\mathrm{tot}} = \sum_i S_i, \tag{3}$$

$$\alpha_{\mathrm{mean}} = \frac{A}{S_{\mathrm{tot}}}. \tag{4}$$

If a room is still in its design stage, Eqs. (2)–(4) are used in that order. Once a room is finished, it is very difficult to distinguish the absorption coefficients of each surface and α_{mean} is calculated backwards from the RT, V, and S_{tot} measurements:

$$\alpha_{\mathrm{mean}} = \frac{0.16V}{\mathrm{RT}S_{\mathrm{tot}}}. \tag{5a}$$

The remainder of this paper will use a slightly different definition of S_{tot}. In an empty room, all absorption is from the ceiling, floor, and walls only. When furniture and people are added to the room, values A and S_{tot} will increase and V may decrease. In practice, the differences are slight and we will use the following *definition* of the mean absorption coefficient, denoted by α, throughout this paper:

$$\alpha \equiv \frac{0.16V}{\mathrm{RT}S}, \tag{5b}$$

where V and S are calculated for the empty room, but RT includes the influence of furniture and people if present during the measurements.

The calculation of the SPL from a source in a room is introduced as

$$\mathrm{SPL} = L_W + 10\log(H), \tag{6}$$

where L_W represents the sound power output of the sound source. The room and source characteristics are denoted by the variable H. In most acoustics textbooks (see, for instance, Pierce [12]), H is formulated as

$$H = \frac{Q}{4\pi r^2} + \frac{4(1-\alpha)}{A}. \tag{7}$$

The first term is for the direct sound between a speaker and a listener and is dependent on the distance between the two, given by r, and a factor Q, which represents the directivity of the source. Q is a number that is sometimes expressed as the directivity factor (DI) in decibels, defined as

$$\mathrm{DI} = 10\log(Q). \tag{8}$$

Hodgson *et al.* use $Q=2.0$, or DI=3.0 dB, in front of the mouth in their article.[6] In the $\mathrm{AL}_{\mathrm{cons}}$ measuring method a value of $Q=2.5$, or DI=4.0 dB, is generally used.[13]

In most practical situations the first term in Eq. (7) is greater than the second term if r is smaller than about 1 m. If the second term is much greater than the first, at larger dis-

tances from the source, the receiver is in the so-called reverberant field. In this case, H reduces to H_{dif}, defined as follows:

$$H_{\text{dif}} = \frac{4(1-\alpha)}{A}. \tag{9}$$

B. Multiple talkers

The variables SPL and L_W in Eq. (6) are logarithmic values that can be written as

$$\text{SPL} = 10\log\left(\frac{p^2}{p_0^2}\right), \tag{10}$$

$$L_W = 10\log\left(\frac{W}{W_0}\right), \tag{11}$$

where p is the sound pressure, W is the total sound power of the source, and p_0 and W_0 are the reference variables taken as 20 μPa and 1 pW, respectively.

According to common acoustic theory, the sound powers of multiple sources in a room can be summed to find the total sound power. This is not always easy since some sources (kitchen noise in dining rooms, for instance) have an impulsive nature but these problems are considered beyond the scope of the present paper. It is assumed that all sound is produced by human speech which is more or less constant.

The representation of the total sound pressure in a room from N talkers can be found by rewriting Eqs. (6) and (7) and introducing the sum over all talkers:

$$\frac{p^2}{p_0^2} = \sum_{i=1}^{N}\left(\frac{W_i}{W_0}\left(\frac{Q_i}{4\pi r_i^2} + \frac{4(1-\alpha)}{A}\right)\right). \tag{12}$$

If all talkers are in the reverberant field, Eq. (12) reduces to

$$\frac{p^2}{p_0^2} = \sum_{i=1}^{N}\left(\frac{W_i}{W_0}\frac{4(1-\alpha)}{A}\right). \tag{13}$$

After the introduction of a mean value W_{mean}, Eq. (13) can be written as follows:

$$\frac{p^2}{p_0^2} = \frac{NW_{\text{mean}}}{W_0}\frac{4(1-\alpha)}{A} = \frac{NW_{\text{mean}}}{W_0}H_{\text{dif}}. \tag{14}$$

A new variable H_m is now introduced:

$$H_m = \sum_{i=1}^{N}\left(\frac{W_i}{NW_{\text{mean}}}\frac{Q_i}{4\pi r_i^2}\right) + \frac{4(1-\alpha)}{A}, \tag{15}$$

and hence the total SPL can be written as

$$\text{SPL} = L_{W,\text{mean}} + 10\log(NH_m), \tag{16}$$

where

$$L_{W,\text{mean}} = 10\log\left(\frac{W_{\text{mean}}}{W_0}\right). \tag{17}$$

It is in fact the main purpose of the present paper to estimate the value of $L_{W,\text{mean}}$ as a function of the number of talkers and the acoustic configuration in a room.

C. H_m vs H_{dif}

In reverberant rooms it can be expected that the second term in Eq. (15) will be greater than the first. In absorbent situations with many sound sources, however, the first term cannot be neglected. Talkers close to the listener in particular can be heard separately. It is possible to make an estimation of the near field effect with a few assumptions. The first assumption is that all talkers have the same sound power, so $W_i = W_{\text{mean}}$. A second assumption is that all talkers speak in random directions and that $Q=1$. The third assumption is that all sound power is evenly distributed in circles around the receiver. An integral equation can be formulated in terms of the distance to the listener and a solution is easily found if the boundaries are circular, which means that the floor space of a rectangular room is translated into a circular floor with the same amount of square meters. So the floor space of a room S_{floor} is represented by $\pi R_{\max}^2$.

In terms of Eq. (12) we find for the first term:

$$\frac{p^2}{p_0^2}W_0 = \frac{NW_{\text{mean}}}{\pi R_{\max}^2}\int_{R_{\min}}^{R_{\max}}\frac{2\pi r dr}{4\pi r^2}. \tag{18a}$$

Solving the integral yields

$$\frac{p^2}{p_0^2}W_0 = \frac{NW_{\text{mean}}}{2\pi R_{\max}^2}\ln\left(\frac{R_{\max}}{R_{\min}}\right). \tag{18b}$$

The average floor space taken by N_p persons is given by

$$N_p\pi\overline{R^2} = \pi R_{\max}^2. \tag{19}$$

The mean value of R is also the best estimate for the minimum value $R_{\min}$. In most cases N_p will be the number of talkers. However, we will also produce measurement results where a target source plus a listener is surrounded by noise-talkers. If the target speaker remains silent the floor area is divided into $N+1$ sections and $N_p = N+1$.

With $N-1$ noise sources, we find for Eq. (18b):

$$\frac{p^2}{p_0^2}W_0 = \frac{(N-1)W_{\text{mean}}}{2S_{\text{floor}}}\ln(\sqrt{N_p}) = \frac{(N-1)W_{\text{mean}}}{4S_{\text{floor}}}\ln(N_p). \tag{20}$$

Numerical verification with noise sources equally spaced in a rectangular room justifies the use of this term.

The calculation of the reverberant part of the noise in a room can be expressed as follows:

$$\frac{p^2}{p_0^2}W_0 = \frac{(N-1)W_{\text{mean}}}{(2S_{\text{floor}} + S_{\text{walls}})}\frac{4(1-\alpha)}{\alpha}. \tag{21}$$

And now H_m can be derived as

$$H_m = \frac{1}{4S_{\text{floor}}}\ln(N_p) + \frac{1}{(2S_{\text{floor}} + S_{\text{walls}})}\frac{4(1-\alpha)}{\alpha}. \tag{22}$$

To show the influence of the contribution of the direct sound represented by the first term, an example is given in Fig. 1 for a rectangular room as a function of the absorption coefficient. For this example the sound power level L_W is taken as a constant and rather arbitrarily as 70 dB.

As Fig. 1 shows, the addition of the direct sound to the reverberant sound plays a role for high values of the absorp-

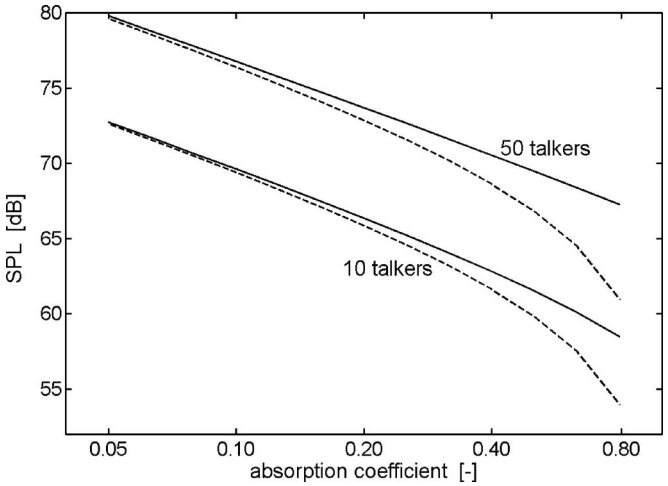


FIG. 1. Sound pressure levels for a $12\times10\times4$ m^3 room with 10 and 50 talkers. Full curves are calculated with both terms of Eq. (22); dashed lines represent the second term only. The value of L_w is taken constant as 70 dB. Absorption coefficients are taken as equal for all surfaces; they are given along a logarithmic axis.

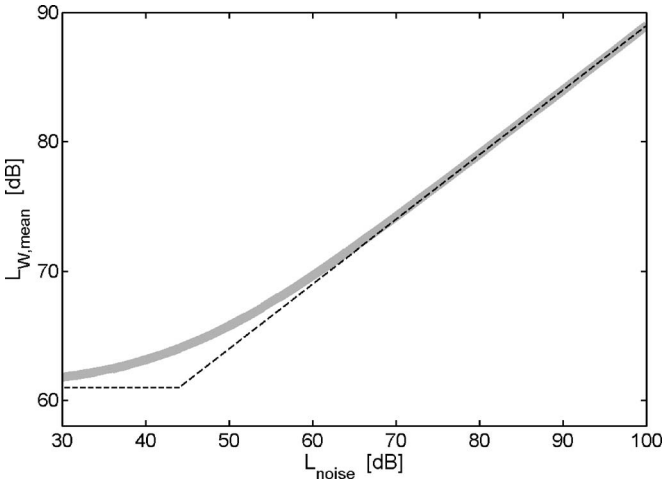


FIG. 2. An example of human vocal output as a function of background noise. The dashed curve is used in older literature. The shape of the full line represents our hypothesis for the remainder of the paper. In this example $C=61$, $D=39$, and $E=0.5$, but the actual values still have to be established.

tion coefficient or for high values of the number of talkers. In this example, 10 talkers stands for one talker per 12 m^2 floor space; 50 talkers (at least 100 people present) on 120 m^2 floor space represents a crowded cocktail party.

The absorption coefficient is given along a logarithmic horizontal axis. When only the denominator in H_{dif} is considered, a straight line will be found as a function of $\log(\alpha)$. The decrease is 3 dB per doubling of the absorption. The numerator causes a steeper descent for the higher absorption values, but the contributions of the direct sound turn the curve almost back into a straight line.

D. A model for vocal output

As a hypothesis for further investigations a model was developed at the start of the measurements. It is expressed as a simple logarithmic sum of two terms:

$$L_{W,\text{mean}} = 10\log(10^{C/10} + 10^{(D+EL_{\text{noise}})/10}), \tag{23}$$

where C, D, and E are the three values under investigation. The model is shown in Fig. 2, where two asymptotic lines are also given: $L_{W,\text{mean}}=C$ for low noise levels and $L_{W,\text{mean}}=D+E\times L_{\text{noise}}$ for higher values (dashed lines).

Equation (23) is based on previous models (see for instance Van Heusden *et al.*[14]) in which these two straight lines are used separately. For low noise levels the signal-to-noise ratio between speech and background noise is assumed sufficient and talkers do not raise their voices at all. From one specific noise level (for instance, 44 dB in Fig. 2) talkers raise their vocal output linearly. This part is the Lombard effect with slopes mainly between 0.3 and 0.6 dB/dB.

A discontinuous curve like that in Fig. 2 is based on the assumption that for noise levels below approximately 40 dB people talk at a level which is common for an anechoic chamber. It is our hypothesis that people in these situations talk somewhat louder and therefore we propose the full curve in Fig. 2 instead. A first indication of this level increase can already be found in the results by Van Heusden *et al.*[14]

When the proposed curve is used, different Lombard slopes are found. In Fig. 2 the slope is 0.6 dB/dB for noise levels above 80 dB when $E=0.6$. When noise levels are approximately 60 dB, the slope is only of the order of 0.2 to 0.3.

Recently, after we performed our measurements and curve fitting, Hodgson *et al.* have proposed a model for the total sound pressure level in an eating establishment.[6] When this model is somewhat adapted it may be used as an alternative for Eq. (23):

$$L_{W,\text{mean}} = C + \frac{\text{asym}}{\left\{1+\exp\left(\frac{L_{\text{mid}}-L_{\text{noise}}}{\text{scale}}\right)\right\}}. \tag{24}$$

The curve is shown as a dashed line in Fig. 3. The maximum slope is found when $L_{\text{noise}}=L_{\text{mid}}$. The slope itself equals asym/(4 × scale) dB/dB.

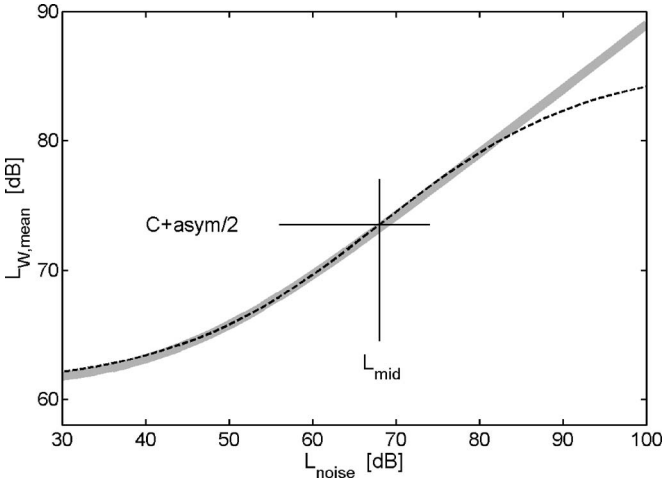


FIG. 3. Two models of human vocal output as a function of background noise. The full line is from Eq. (23); the dashed line is from Eq. (24) proposed by Hodgson *et al.* In this example $C=61$, $D=39$, and $E=0.5$ for Eq. (23); for Eq. (24) asym=25, $L_{\text{mid}}=68$, and scale=12.5 are used. The maximum slope is 0.5 dB/dB.

The curve by Hodgson *et al.* is of course more accurate at high noise levels. There is an upper limit to the vocal output of humans which is not catered for in Eq. (23). At lower noise levels differences are slight. In the example shown in Fig. 3, differences are smaller than 0.2 dB when noise levels are below 80 dB. It is not expected that noise levels from human speech will be higher in the present investigations. In this case the present model is easier to use since it requires only three parameters.

It should be noted that Hodgson *et al.* used their equation for the sound *pressure* level, while Eq. (24) is for the sound *power* output. This has no effect on the shape of the curve. The curve is only shifted vertically by 11–DI dB.

E. Feedback from the room's SPL on the vocal output

If the noise in a room is generated by other talkers only, $L_{W,\mathrm{mean}}$ in Eqs. (16) and (17) is dependent on the noise level from $N-1$ talkers. The total sound pressure level in a room, denoted by SPL, from all N speakers is as follows:

$$\mathrm{SPL} = L_{W,\mathrm{mean}}(L_{\mathrm{noise}}) + 10\log(NH_m). \tag{25}$$

In Eq. (25), SPL and L_{noise} are mutually dependent. Hence the equation must be solved recursively. In general, recursive numerical methods do not form part of the toolbox of the average architect. But there is also a mathematical problem. A curve fitting process is a recursive process as well, and hence a second recursive method is used simultaneously to derive C, D, and E from measurements.

The recursive method to find SPL can be avoided if $E=0.5$. This method was derived after the measurements were carried out and the background will be explained at a later stage. The method itself is described in the present section.

If Eqs. (10), (11), (14), and (15) are slightly rewritten, the sound power from the noise can be written as

$$\frac{p_{\mathrm{noise}}^2}{p_0^2} = \frac{W(p_{\mathrm{noise}}^2)}{W_0}(N-1)H_m, \tag{26}$$

which, in combination with Eq. (23), can be written as

$$\frac{p_{\mathrm{noise}}^2}{p_0^2} = (10^{C/10} + 10^{(D+EL_{\mathrm{noise}})/10})(N-1)H_m, \tag{27}$$

which is

$$\frac{p_{\mathrm{noise}}^2}{p_0^2} = \left(10^{C/10} + 10^{D/10}\left(\frac{p_{\mathrm{noise}}^2}{p_0^2}\right)^E\right)(N-1)H_m. \tag{28}$$

When $E=0.5$, Eq. (28) can be solved as a quadratic equation with the solution:

$$\frac{p_{\mathrm{noise}}}{p_0} = 10^{D/10}(N-1)H_m \times\left(0.5 + 0.5\sqrt{1 + \frac{4\times 10^{(C-2D)/10}}{(N-1)H_m}}\right), \tag{29}$$

and hence:

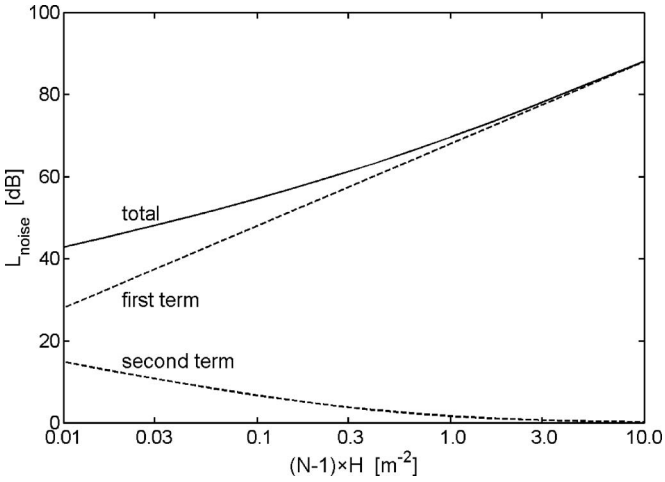


FIG. 4. L_{noise} calculated with the aid of Eq. (30) (full line) to show the difference between the first and the second term (dashed lines).

$$L_{\mathrm{noise}} = 20\log(10^{D/10}(N-1)H_m) + 20\log\left(0.5 + 0.5\sqrt{1 + \frac{4\times 10^{(C-2D)/10}}{(N-1)H_m}}\right). \tag{30}$$

Figure 4 illustrates the contributions of the first and second terms to the total noise level. If the human vocal output were not affected by noise, a straight line would be found with an increase of 3 dB per doubling of $(N-1)H_m$. Equation (30) results in a higher slope. The first term causes an increase of 6 dB per doubling of $(N-1)H_m$; the addition of the second term reduces the slope. If $(N-1)H_m$ is greater than about 2 this second term no longer has any influence and a 6 dB increase is found.

An example will serve to clarify the values of N and H_m. A rectangular room measuring $9\times6\times3$ m^3 has a total area of 198 m^2; the floor space is 54 m^2. So, if $\alpha=0.15$, $H_m=0.11$. When the absorption coefficient equals 0.30, H_m becomes 0.047.

If $(N-1)H_m>2$ the contribution of the second term in Eq. (30) is less than 1 dB and the slope of the total curve varies with $20\log\{(N-1)H_m\}$. This value is found if the number of talkers is 18 or 42 for 15% and 30% absorption, respectively. The first case is possible on 54 m^2 floor space, but the second case should be considered as a very crowded cocktail party.

When the room is scaled up by a factor of 2 to 18 $\times12\times6$ m^3, the results are the same if the number of talkers is increased to 72 and 164, respectively. So the most important factor is in fact not the number of talkers, but the number of talkers per floor space.

III. MEASUREMENTS

A. Preliminary measurements

During the measurement phase participants were asked to read out excerpts from magazines at "normal conversational level as if they were talking to a listener at 1 m distance." The sound levels when reading aloud are somewhat higher than those for conversation; differences of about 2 dB have been reported.[14] To investigate this effect, preliminary

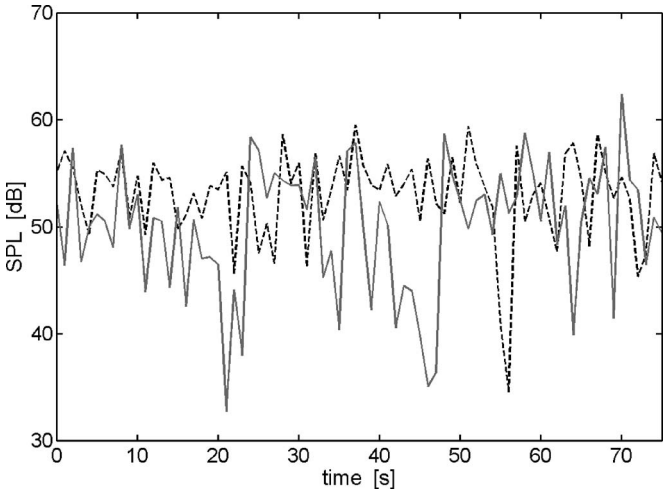


FIG. 5. SPL at 1 m in front of the head of a male test participant, recorded in an anechoic chamber, for reading (dashed curve) and normal conversation (full curve).

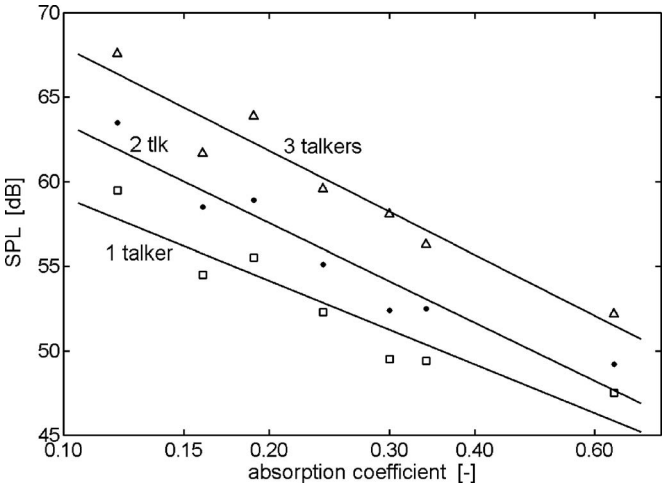


FIG. 6. SPL values from one, two, or three talkers in the reverberant field. Full lines represent best fit curves; the slopes are $-20.4\log(\alpha)$, $-19.6\log(\alpha)$, and $-16.4\log(\alpha)$ from top to bottom.

measurements were carried out in an anechoic chamber. These were also used to measure the lowest noise levels.

Six participants were asked to read excerpts from books. Equivalent sound levels were measured from 19 texts lasting approximately 2 min. The microphone was placed 1 m from participants' mouths. The measured mean A-weighted sound pressure level of the test participants was 54.6 dB, with a minimum level of 52.0 dB and a maximum level of 57.4 dB.

Figure 5 shows a typical example of the difference between reading and talking for one male test participant. The sound level of the book excerpt is fairly constant. For conversational speech the talker starts at the same level as when reading but decreases his output during the sentence, which can be observed between 15 and 22 s and between 38 and 48 s. Since the maximum levels are the same, the equivalent sound level is lower. The difference of 2 dB in equivalent sound levels agrees with that found in the literature.[14]

These findings mean that a correction is needed from reading to conversation. However, although the effect was not investigated, in the authors' opinion correction is only required at low noise levels: in noisy places talkers cannot afford to decrease their sound level during sentences, so the reading values are useful for representing conversational speech in the noisier cases without any adjustment. The difference in sound levels between participants in the anechoic chamber was up to 5.4 dB.

Measurements at 1 m in front of the mouth for 20 test persons were also done in an office room. Background noise levels varied from 25 to 80 dB. The results confirmed the Lombard slopes found in literature and are omitted here. However, the spread in the group of 20 participants is worthwhile noting. The difference between the softest and loudest voice was 12 dB when the noise level was below 30 dB. One would expect that this variation would decrease at higher noise levels as people are forced to adapt their voices to the noise. This appeared only partly the case, as differences of 9 dB are still found at a noise level of 70 dB.

B. Room with variable absorption

Speech recordings were made in a purpose-built test facility at the Conservatory of Amsterdam. The dimensions of the room were $6.4\times4.8\times3.5$ m^3, with a volume of 108 m^3 and a total surface area of 140 m^2. During the measurement phase the room contained a grand piano, a table, a few chairs, and six people.

The architect of the new Conservatory had designed special absorbers to be used in the new building, measuring 0.90×0.90 m^2 and 0.90×1.80 m^2. There were 18 large absorbers and 36 small absorbers. Special effort was invested into obtaining a flat reverberation curve as a function of frequency. Reverberation times were measured in octave bands; a mean value was calculated over the 500, 1000, and 2000 Hz bands. The mean absorption coefficients were calculated according to Eq. (5b).

One participant read a text sitting at a table. Two listeners were positioned opposite the first participant with their ears 1 m away. The noise was produced by one, two, or three real talkers at a greater distance (3 m or more) from the speaker–listener combination. They were not completely in the reverberant field of the room since the first term of Eq. (22) has some influence for the highest absorption coefficient.

Test participants were asked to read excerpts from magazines. In a test run, the main speaker at the table was asked to start reading. After approximately 30 s the noise speaker(s) started to read. After about 2 min the main speaker stopped, but the noise speaker(s) continued for a further 30 s. This method provides us not only with noise from the main speaker at 1 m, but also noise from one to three noise speakers without the main speaker.

Three runs were undertaken per absorption situation and per noise speaker. One male speaker at the table read twice; the third run was read by a female speaker.

Figure 6 shows SPLs measured at the listener's position from one, two, or three noise speakers when the target speaker remains silent as a function of the mean absorption in the room. There were seven values of the absorption co-

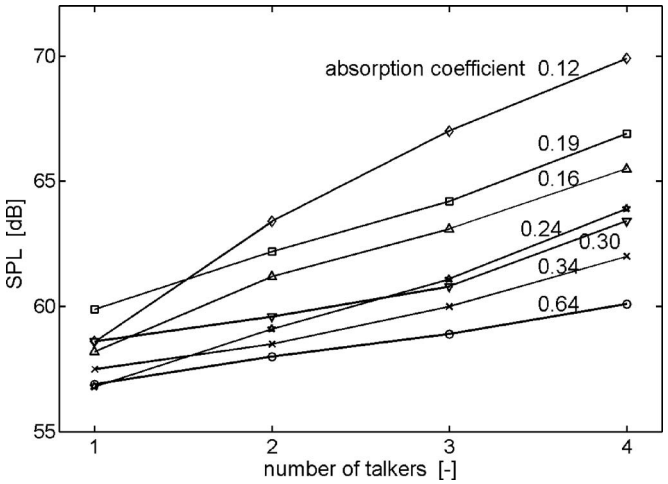


FIG. 7. SPL values at 1 m in front of the talker at the table (number of talkers=1). The total SPL is found if one, two, or three noise talkers in the reverberant field are added.

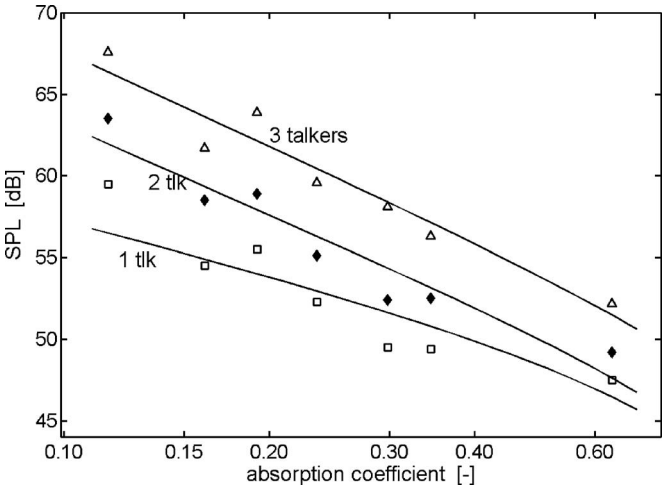


FIG. 8. Calculations of SPL for one to three talkers in the reverberant field, to be compared with Fig. 6. Triangles are for three talkers, diamonds for two talkers, and squares for one talker. See the text for C and D values.

efficient. The first situation had $\alpha=0.19$. In the second situation absorbing panels were removed and $\alpha=0.12$. Then the number of panels was increased to find $\alpha=0.16$, 0.24, 0.30, 0.34, and 0.64 in the following five situations.

The first situation, where $\alpha=0.19$, has the same amount of absorbers as the third situation, where $\alpha=0.16$. In the latter case all absorption is on the ceiling; in the first case the absorption is randomly distributed. Ray-tracing models predict the decrease of absorption as measured here.[15] However, they also predict that the SPL will stay almost constant, so the somewhat lower SPLs at $\alpha=0.16$ are not explained. More measurements are necessary to investigate this effect.

The values of NH_m are between 0.02 and 0.77, so the value of $NH_m=2$, where only the first term from Eq. (30) remains, is never reached and the second term in Eq. (30) and Fig. 1 always has an influence.

Figure 7 gives the SPL values at 1 m from the source speaker, now including the target speaker. The values are averaged over two readers in three sessions, since one reader did the test twice. Values are given as a function of the total number of speakers. So, for instance, four talkers means one target speaker plus three noise talkers.

In Fig. 7 the same increase in SPL can be observed as in Fig. 6 when the curves with $\alpha=0.16$ and 0.19 (and equal number of absorbers) are compared.

IV. CURVE FITTING

A. The curve-fitting process

In Sec. II D the background of our output curve for vocal effort was explained. Equation (23) gave the curve as a function of three variables C, D, and E. In the following, these variables will be estimated by curve fitting.

We did not find a statistical method for determining the three values simultaneously. Curve fitting methods are recursive methods, but Eq. (25) should be solved recursively as well, since SPL is on the left-hand side and the right-hand side of the equation as the noise speakers raise their voices as well. In this case a fitting process has to solve two recursive processes simultaneously.

When the noise level is kept constant, the recursive effect disappears. In this case it is possible to apply curve fitting to the data. We fitted our data using the MATLAB function lsqcurve. Babble and noise measurements were done in an office room in the early stages of our research. A comparison was also made with measurements by van Heusden *et al.*[14] Calculated E values ranged from 0.42 to 0.56, but a comparison of curves showed that if a curve was chosen with $E=0.5$, the error value produced by the fitting process increased only marginally. Differences between the best fit curves and the curves with $E=0.5$ were within 1 dB, which is well below the variation found in the measuring results. Therefore we decided to use $E=0.5$ as a given variable throughout the rest of the investigations. Equation (30) could then be applied and the recursive process for the calculation of SPL was no longer necessary, so MATLAB's lsqcurve could easily find the two remaining values for C and D.

A second problem lay in finding values for the power output (L_W) from measurements of SPL; if the microphone is close to the talker, as in Eqs. (6) and (7) and in Fig. 7, the directivity factor Q from the "target" speaker must be estimated as well.

The results from the curve fitting process will show that the value $Q=2.5$ as given in Sec. II is too low; a Q value between 4 and 5 is a better estimation, so the directivity index (DI) increases from DI=4 dB to DI=7 dB. A 2 dB increase is probably caused by the table top between the target speaker and the receiver. We measured similar differences with a loudspeaker.

B. Fitting with measurements from the Amsterdam Conservatory

The next comparison is between measurements taken at the Amsterdam Conservatory, originally given in Fig. 6, and calculations from the model, for three talkers in the reverberant field. Figure 8 shows the results of the fitting process.

Curve fitting through the measuring points for three talkers yields $C=58.3$, $D=35.2$. For the points measured with two talkers these values are almost the same: $C=57.7$, D

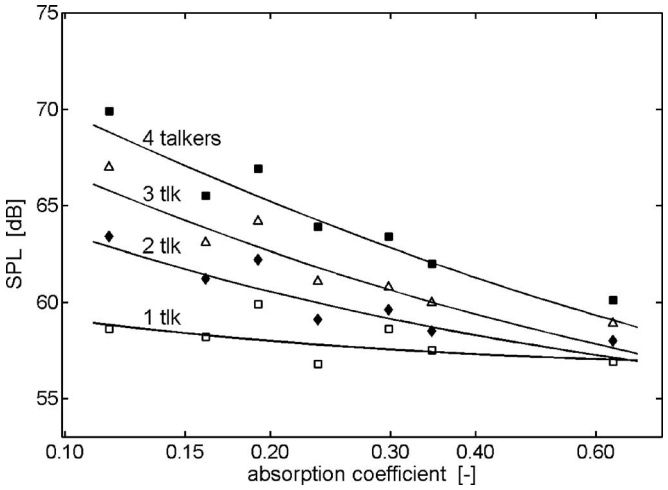


FIG. 9. Calculations of SPL at 1 m from a talker with zero, one, two, or three noise speakers in the reverberant field. The measurements from one, two, three, and four talkers are represented by open squares, closed diamonds, open triangles, and closed squares. See the text for C, D, and Q values.

=35.1. However, when there is only one talker, the noise level is very low and constant so the value of D has no meaning. The C value found in this case is 5 dB higher: C =63.0 dB. The C values for one speaker are higher than for the multitalker cases. One reason may be that our three single talkers have a voice that is louder than the mean value of three or four talkers. However, there may be another explanation. When people listen to talkers, reverberation can be regarded as noise. Common measures like speech transmission index (STI) and U_{50} for speech intelligibility are based on this principle.[16–18] Our hypothesis is that talkers react to their own sound accordingly. A simple addition to the model can be made by adding the late energy of the talker after 50 ms, as done in U_{50}. If, for instance, the late energy after 50 ms is calculated or measured as 60% of the total sound energy, a 0.6 noise speaker is added to the total number of noise speakers. In fact, it appeared possible to fit the single talker into the multitalker model. However, more measurements are required in order to check our hypothesis.

Figure 9 shows results for comparison with the readings of Fig. 7, although the results are now given with the absorption along the horizontal axis. This time the target speaker is also included: "one talker" is when he or she is the only one speaking, while two, three, and four talkers means one, two, and three noise speakers, respectively. Again, for one talker the D value has no meaning. The curve fitting process yields C=61.1 and Q=4.8. For two speakers the combination of C, D, and Q is (60.4, 32.9, 4.3), for three speakers (59.2, 33.5, 4.7), and for four speakers (59.6, 34, 4.6).

C. Gardner's results (1971)

Research undertaken at our university focuses on small numbers of talkers. It is interesting, however, to compare the model with speech from larger numbers of talkers. Gardner's results are the most appropriate, since he explicitly used multitalker backgrounds and he removed background noise from other sources.[3] More recent findings, such as those of Hodgson *et al.*, cannot be used in this instance as they include, for example, noise from radios in cafés.

Gardner investigated the Lombard effect as a function of N for some auditoria and dining rooms. In this section our intention is to estimate our C and D values from the results as given by Gardner. It is assumed that E=0.5, so all calculations are done using Eqs. (16), (22), and (30).

As an input, values are required for N, α plus the geometrical values of the rooms. N is naturally very hard to measure, so Gardner took the number of people present in the room and not the actual number of talkers. To fit our model, the value of H is needed for a specific room. This value must be derived from the measurement of the reverberation time plus the geometrical variables or at least the total area S. It is possible to calculate H_{dif} for three of Gardner's cases, since H_{dif} is very similar to the R value given by Gardner (in square feet). It is defined as

$$R = \frac{\alpha S}{(1-\alpha)} = \frac{0.05V}{\mathrm{RT}(1-\alpha)}. \tag{31}$$

For our estimation the R value itself is not enough since we require α and the area S. Gardner gives results for a 341-seat auditorium in his Fig. 5. The R value is given as 20 000 ft^2; no dimensions are given. If α is calculated from this value by assuming the dimensions that belong to this type of auditorium, something strange occurs: either the value of the volume is extremely high, or the value of α is no less than 0.7. By coincidence, there is a 336-seat auditorium in our own faculty building. When we measured that auditorium a much smaller value R=3900 ft^2 was found, or H_{dif} =0.011 m^{-2}. It was then decided to take similar measurements to Gardner's in this auditorium.

The number of individuals present at the entrance of the auditorium was counted, but to estimate the percentage of people actually talking we installed a video camera as well. Figure 10 shows the results plotted over Gardner's findings. The open circles are Gardner's; closed triangles represent our measurements.

As can be seen from Fig. 10, Gardner's measurements and our own are very similar. The percentage of people actually talking when the auditorium was almost full was 25%. Gardner mentions 30%. If this is correct, his results are 1 dB higher. The full line shows the result of curve fitting. The total sound pressure level when 320 people are present is 71 dB. If people do not raise their voices, this level would be only 59 dB, so a 12 dB increase in vocal output is found due to the Lombard effect.

In Figs. 11 and 12 Gardner's Figs. 8 and 9 are replicated with curves from our model. In this instance Gardner gives both geometrical and acoustical values. The R value in Fig. 11 (Fig. 8 in Gardner's article) is again very high, but this time it is correct, since the room has "heavy drapes" and so the absorption coefficient is no less than 0.66. In this room it is important to use both terms in Eq. (30) because this is a typical example of a room where the direct contributions are at least as important as the sound from the reverberant field.

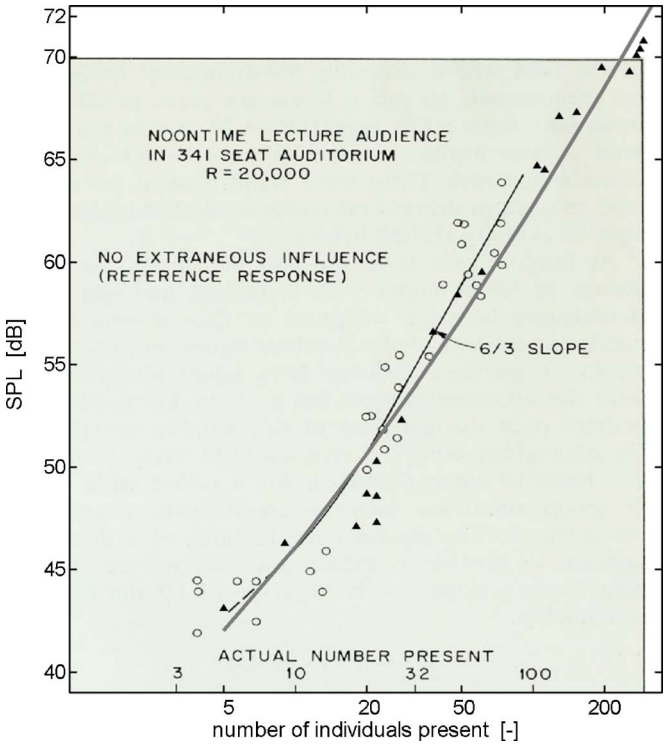


FIG. 10. (Color online) Results from Gardner's auditorium as given in his Fig. 5 (open circles). Dots are from measurements taken in a similar auditorium at our Faculty of Architecture. The full grey line gives an estimation for a rectangular room of $22\times15\times6$ m^3 where $\alpha=0.24$. It is calculated using Eq. (30) when $C=59$ and $D=35.5$. The percentage of people talking is 25%.

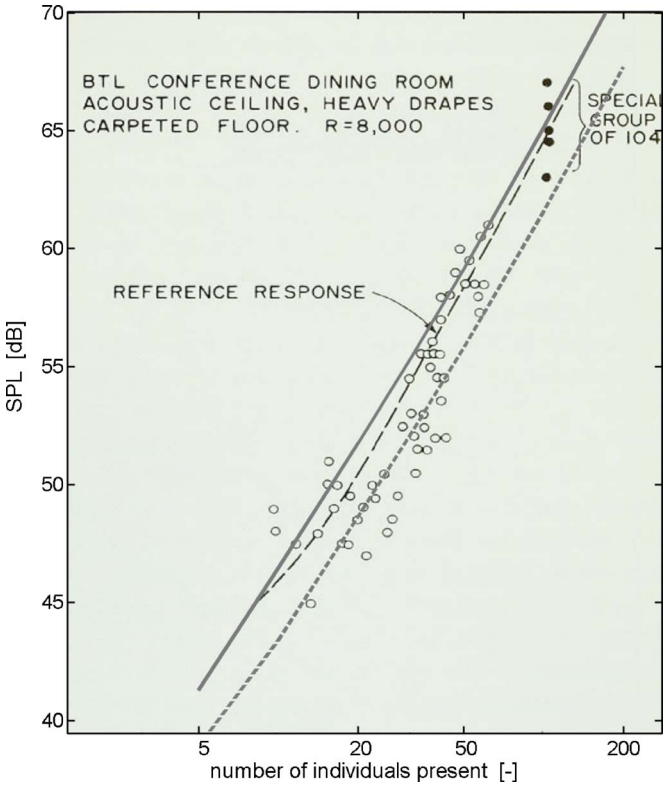


FIG. 11. (Color online) Results from Gardner's dining room as given in his Fig. 8, plus an estimation from our Eq. (30) when $C=59$ and $D=34$. The percentage of talkers is 45% for the full line and 30% for the dotted line.

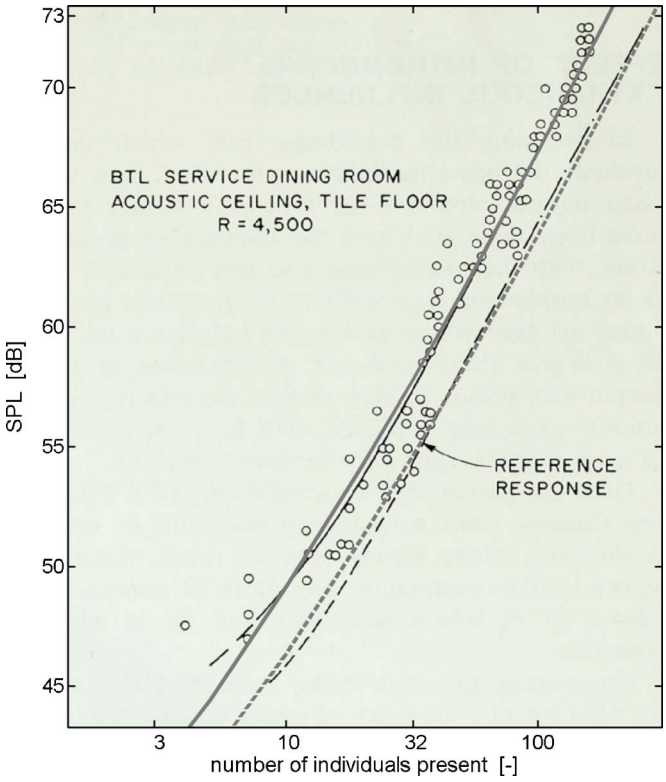


FIG. 12. (Color online) Results from Gardner's dining room as given in his Fig. 9, plus an estimation from our Eq. (30) when $C=59$ and $D=35.5$. The percentage of talkers is 45% for the full line and 30% for the dotted line.

In Fig. 12 the absorption coefficient is found as $\alpha=0.37$, which is still rather a high value. Moderately reverberant rooms were not available in Gardner's article.

The model input was $C=59$ in both cases. To fit the results with the measurements D should be chosen as 34 in Fig. 11 and as 35.5 in Fig. 12. We will return to this subject in Sec. V.

Gardner pointed out that the best fit through his measuring points in Figs. 11 and 12 should be steeper than the 6/3 slope he derived from Fig. 10, which means a 6 dB increase per doubling of the individuals present. It could be achieved in our model by increasing the E value from 0.5 to about 0.6. This could be viewed as a flaw in our model. However, another explanation could be easily observed on the video tapes we made, but can also be observed in practical situations: The percentage of talkers often increases with the number of people present. In an auditorium people arriving early are "single" and the percentage of talkers may be as low as 20%. When more people enter they join the early attendees and when the auditorium is full the percentage of talkers may be almost doubled. A similar effect can be found at cocktail parties. People often talk in groups of four to eight at the start, but these groups break up when more people enter.[19] This time the reason is acoustic: Bigger groups create bigger talker–listener distances and groups *must* break up at higher noise levels in order to maintain the minimum signal-to-noise ratio, which is found at short talker-listener distances.

In Fig. 10 our four points well below the calculated curve are explained by a lower percentage of talkers. They are measured when 22 to 24 people are in the auditorium.

TABLE I. Overview of C, D, and Q values from five cases plus the values estimated for the calculation model.

Case		Q	C	D
A	Reading in Conservatory, talkers in reverberant field		58	35
B	Reading in Conservatory, talker at 1 m included	4.6	59.5	33.5
C	Gardner's Fig. 5, normal conversation		59	35.5
D	Gardner's Fig. 8, normal conversation		59	34
E	Gardner's Fig. 9, normal conversation		59	35.5
	Estimation for architectural purpose		59	35

The model calculates a sound level of about 50 dB when five or six people are talking (which is 25%). On the videotape it is evident that only three or four people are talking simultaneously, so the decrease in SPL is explained.

Figures 11 and 12 contain two curves from calculations when 30% and 45% of the people present in the room are actually talking; the 30% curve gives the best fit for lower noise levels and the 45% curve is for higher noise levels.

Gardner mentions that he only used normal conversation by cutting out laughter, coughing, and clashing of dishes. We followed the same procedure but this was not easy. As an auditorium fills up, outbreaks of enthusiasm and laughter increase in greater and greater proportions. In our auditorium we also suffered from noise caused by footsteps in the wooden aisles. When the auditorium was full, "normal conversation" could only be measured about one-third of the time. The nonfiltered SPL values measured are about 5 dB higher.

D. Results from Hodgson *et al.* (2007)

In Eq. (24) and Fig. 3 a comparison was made with the curve used by Hodgson *et al.* If we fit this curve to our results when $C=59$ and $D=35$ and when the noise level is kept below 80 dB, the following values are found for an exponential curve: $C=58.7$, asym$=23.4$, $L_{\mathrm{mid}}=69.8$, and scale$=12.2$. The resulting maximum slope is 0.48 dB/dB. The differences between the two curves are 0.3 dB or less.

Hodgson *et al.* give measured SPL values, while our results are for L_W, but they assume that all their sound sources are in the reverberant field of the model listener and an estimation of Q is not required. Hodgson *et al.* give room dimensions plus reverberation times and number of seats, but to calculate SPL values, the number of talkers should be estimated from the number of seats. If that factor is taken as equal to 33%, differences between the results of Hodgson *et al.* and our model range from −3 to +3 dB for their cases C, B, and R. However, comparison is difficult, since Hodgson *et al.* did not restrict the noise levels to other talkers as Gardner and we did, and their variations (Table I) in SPLs are approximately 15–20 dB. They also took loud background music into account, for instance. In the two "senior residences" from Hodgson's article, the results from our model were about 10 dB too high. It is possible that the number of talkers was less (the discrepancy vanishes if only one out of ten attendees is actually speaking) or the seniors have softer voices. This is an interesting question for further research.

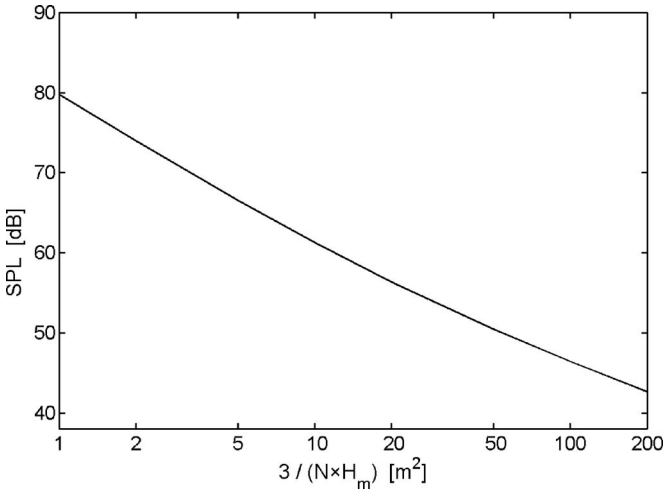


FIG. 13. The sound pressure level in a room calculated using Eq. (30). Values used in the model are $C=59$ and $D=35$. The target speaker at close distance to the listener is not included. The values plotted horizontally are almost equal to the absorbing surface per talker.

V. SUMMARIZING THE RESULTS

A. Discussion of *C* and *D* values

In order to find the C value and D value to be used in our model, the results are summarized in Table I. The values of C and D on the lowest row of Table I are calculated from the rounded means of cases A–E.

In fact, the C value represents the sound power output of human speech in an anechoic chamber. If $C=59$ and $Q=2.5$, the resulting sound pressure level at 1 m is 52 dB. This level is 2 dB lower than the mean level we measured in the anechoic chamber, but those values were from reading a book. Lower sound pressure levels, even below 50 dB, have been reported in literature.[14,20]

Case D represents a multitalker situation in a highly damped room. Now the first term of Eq. (22), predicting the influence of direct sound from all talkers in the room, is equally important as the second term for the reverberant part. Although this term from Eq. (22) has been extensively compared with numerical models, it might overestimate the direct contributions. On the other hand, an overestimation of the contribution of the reverberant part is also very likely in damped rooms, because diffuse fields with constant levels through the room are unlikely.[11,21]

B. Curve for architectural practice

If C and D are taken as 59 and 35, respectively, Fig. 4 can be redrawn. This is shown in Fig. 13, but the variable along the horizontal axis has changed. The chosen variable $3/(NH_m)$ is equal to A/N, when $\alpha=0.25$. A/N represents the absorbing surface per talker in a room and is an easy design parameter for an architect to use.

De Ruiter[19] presented similar curves, based on measurements (including Gardner's) and on the ISO 9921 standard. The results are of the same order, but a comparison is impossible since the results are given with the number of people present in the room. Unfortunately there is no indication about the percentage of people actually talking.

According to the curve in Fig. 13, it is very unlikely to find sound pressure levels from normal conversation higher than 80 dB, since one talker–listener combination represents 1 m^2 of absorption. To verify this statement, some relatively simple readings were taken with a sound level meter. A higher level was found only once during a very crowded cocktail party. The room was very reverberant, so the absorption was almost completely provided by the people standing on the floor. In this case A/N can be below 1 when people are very close to each other.

The results agree with the results given in Table 1 by Hodgson *et al.*[6] In that table, maximum L_{eq} levels are of the order of 75–79 dB, apart from one bistro with "loud music" where levels of 82 dB have been measured.

VI. CONCLUSION

The model introduced in this paper to predict the Lombard effect is based on a simple curve, with a gradually increasing slope for the vocal sound power as a function of the actual sound pressure level in the room. The model is designed to predict human vocal output for different numbers of talkers in rooms of different sizes and with different sound absorbing properties. In crowded places, outbursts of loud talking, laughter, etc., may increase equivalent sound levels, but these effects are not incorporated in the model since it is restricted to "normal conversation" only. Therefore the model does in fact predict the minimum levels for practical situations and an increase of 5 dB can easily be found.

The model shows an increase in vocal output of 0.5 dB per 1.0 dB increase of the sound pressure level when noise levels are approximately 70–80 dB. When noise levels are around 50 dB the model predicts slopes on the order of 0.2 or 0.3 dB/dB.

The model is fitted with five different sets of measuring results. A best fit is found if the sound power level is estimated as 59 dB when noise is absent. The five cases are very well matched and show differences of only 1 dB for multitalker cases. The inaccuracy in the prediction of the sound power level is about 2 dB for high noise levels, especially in nonreverberant situations, where talkers in the vicinity of the listener can be heard separately. More measurements are needed to decrease this inaccuracy.

The main reason for developing the model was to use it as a tool in the architectural design process. What happens to the sound level if the amount of absorbing area in a room is doubled, for example? The model predicts slopes as high as −6 dB per doubling of absorption if only the reverberant field is taken into account. If the influence of direct sounds is incorporated as well, this slope is lower and a value of −5 dB is found on most occasions. This value is based on the assumption that the percentage of people actually talking remains constant. In noisy conditions this percentage is always close to 50% because groups of three or four people are simply unable to understand each other. Such poor circumstances are improved by adding absorption, so the percentage of talkers may drop as well and slopes may be steeper than −6 dB per doubling of the absorption.

[1] E. Lombard, "Le signe de l'élévation de la voix [Characteristics of the elevation of the voice]," Annales des Maladies de l'Oreille et du Larynx **37**, 101–119 (1911).
[2] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 667–709 (1971).
[3] M. B. Gardner, "Factors affecting individual and group levels in verbal communication," J. Audio Eng. Soc. **19**, 560–569 (1971).
[4] K. Pearsons, R. L. Bennett, and S. Fidell, "Speech levels in various environments," U.S. Environmental Protection Agency, EPA-600/1–77–025, Washington D.C., 1977.
[5] W. O. Olsen, "Average speech levels and spectra in various speaking/listening conditions, a summary of the Pearson, Bennett, & Fidell (1977) Report," American Journal of Audiology **7**, 21–25 (1998).
[6] M. Hodgson, G. Steininger, and Z. Razavi, "Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments," J. Acoust. Soc. Am. **121**, 2023–2033 (2007).
[7] M. Oberdörster and G. Tiesler, "Acoustic ergonomics in schools," Federal Institute for Occupational Safety and Health, Dortmund, Germany, 2006.
[8] M. R. Hodgson, R. Rempel, and S. M. Kennedy, "Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures," J. Acoust. Soc. Am. **105**, 226–233 (1999).
[9] M. R. Hodgson, "Case-study evaluations of the acoustical designs of renovated classrooms," Appl. Acoust. **65**, 69–89 (2004).
[10] M. E. Valk, P. H. Heringa, and L. Nijs, "Het optimaliseren van de ruimteakoestiek voor de les- en oefenruimtes van het Conservatorium van Amsterdam," [Optimizing the room acoustics for the study and teaching rooms in the Conservatory of Amsterdam], Bouwfysica **17**, 11–15 (2006).
[11] L. Nijs, P. Versteeg, and M. van der Voorden, "The combination of absorbing materials and room shapes to reduce noise levels," 18th International Congress on Acoustics, Kyoto, 2004.
[12] A. D. Pierce, *Acoustics* (Acoustical Society of America, New York, 1985).
[13] J. van der Werff, *Speech Intelligibility, the ALcons Method Based on the Work of Victor Peutz* (Zoetermeer, Peutz, 2004).
[14] E. van Heusden, R. Plomp, and L. C. W. Pols, "Effect of ambient noise on the vocal output nd the preferred listening level of conversational speech," Appl. Acoust. **12**, 31–43 (1979).
[15] L. Nijs, "The distribution of absorption materials in a rectangular room," Internoise Congress on Noise Control Engineering, Rio de Janeiro, Brazil, 2005.
[16] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077 (1985).
[17] J. S. Bradley, "Speech intelligibility studies in classrooms," J. Acoust. Soc. Am. **80**, 846–854 (1986).
[18] J. S. Bradley and S. R. Bistafa, "Relating speech intelligibility to useful-to-detrimental sound ratios," J. Acoust. Soc. Am. **112**, 27–29 (2002). This article refers to and corrects three earlier publications.
[19] E. Ph. J. de Ruiter, *The Great Canyon, Reclaiming Land from Urban Traffic Impact Zones* (Zoetermeer, Peutz, 2004).
[20] W. T. Chu and A. C. C. Warnock, "Detailed directivity of sound fields around human talkers," National Research Council Canada, IRC-RR-104, 2002.
[21] M. Barron, *Auditorium Acoustics and Architectural Design* (E&FN Spon, London, 1993).

Acoustical determination of the parameters governing thermal dissipation in porous media

Xavier Olny
Département Villes et Territoires/Groupe AUE, Centre d'Etudes Techniques de l'Equipement de Lyon, 46 rue Saint Théobald-BP 128, 38081 L'isle d'Abeau Cedex, France

Raymond Panneton[a)]
GAUS, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

(Received 29 July 2007; revised 4 December 2007; accepted 4 December 2007)

In this paper, the question of the acoustical determination of macroscopic thermal parameters used to describe heat exchanges in rigid open-cell porous media subjected to acoustical excitations is addressed. The proposed method is based on the measurement of the dynamic bulk modulus of the material, and analytical inverse solutions derived from different semiphenomenological models governing the thermal dissipation of acoustic waves in the material. Three models are considered: (1) Champoux–Allard model [J. Appl. Phys. **20**, 1975–1979 (1991)] requiring knowledge of the porosity and thermal characteristic length, (2) Lafarge *et al.* model [J. Acoust. Soc. Am. **102**, 1995–2006 (1997)] using the same parameters and the thermal permeability, and (3) Wilson model [J. Acoust. Soc. Am. **94**, 1136–1145 (1993)] that requires two adjusted parameters. Except for the porosity that is obtained from direct measurement, all the other thermal parameters are derived from the analytical inversion of the models. The method is applied to three porous materials—a foam, a glass wool, and a rock wool—with very different thermal properties. It is shown that the method can be used to assess the validity of the descriptive models for a given material.
 [DOI: 10.1121/1.2828066]

PACS number(s): 43.55.Ev, 43.20.Gp, 43.20.Jr, 43.20.Ye [NX] Pages: 814–824

I. INTRODUCTION

In a previous paper,[1] the authors proposed an analytical inversion method to determine the viscous parameters (static airflow resistivity, viscous characteristic dimension, and tortuosity) of open-cell porous media from the acoustical measurement of the dynamic density of the material. In this paper, following a similar method, the attention is now focused on the determination of the thermal characteristic parameters used in the semiphenomenological models predicting the dynamic heat conduction phenomena in open-cell porous media.

Amongst all the models, three models have been retained for this investigation: (1) The Champoux–Allard model,[2,3] based on one parameter—the thermal characteristic dimension; (2) the Lafarge *et al.* model,[4] based on two parameters—the thermal characteristic dimension and the static thermal permeability; and (3) the Wilson model,[5] based on two adjustable parameters. An overview of methods used for characterizing these parameters can be found elsewhere.[1,5,6] Here, the goal is to focus our work on the transposition of the analytical inversion method developed in Ref. 1 to the characterization of the thermal characteristic parameters behind the three aforementioned semiphenomenological models.

[a)]Author to whom all correspondence should be addressed. Electronic mail: raymond.panneton@usherbrooke.ca

The present paper is organized as follows. In Sec. II, the descriptive semiphenomenological models are recalled to pave the way to the development of the analytical inversion method, where analytical solutions are derived from the three descriptive models. Section III estimates the error relative to the analytical inversion in the characterization of the Lafarge *et al.* parameters for two virtual materials of theoretically known properties. Section IV describes the material samples and experimental setup used to test the method. Section V presents the results of the analytical inversion applied to the tested materials. Finally, concluding remarks are given in Sec. VI.

II. THEORY

Let us consider an open-cell porous medium submitted to an acoustic excitation. Assuming the medium is homogeneous at the wavelength scale and its solid frame is rigid (i.e., motionless) or very limp (i.e., no rigidity), then only one compression wave propagates in the medium.[7] In this case, the porous medium is seen as an equivalent fluid characterized by an equivalent dynamic density $\tilde{\rho}_{eq}$ and a dynamic bulk modulus $\tilde{K}_{eq}$—the tilde indicates that the associated variable is frequency-dependent and complex-valued. These two dynamic properties are also used in Biot theory[8,9] to account for the visco-inertial and elastic interactions between air and frame in a poroelastic aggregate.

While propagating through the equivalent fluid, the compression wave is attenuated due to viscous and thermal dissipation mechanisms. The dynamic density accounts for

the viscous losses, and the dynamic bulk compression modulus for the thermal losses. In the past, many models relating these dynamic properties to the macroscopic parameters of the porous network have been worked out. In this work on the parameters governing the thermal dissipation, three semiphenomenological models are selected to describe the dynamic bulk modulus: (1) The model by Champoux and Allard, (2) the model by Lafarge *et al.*, and (3) the model by Wilson. These three models and their parameters are recalled in the following.

A. Champoux–Allard model

The semiphenomenological model by Champoux and Allard[2] is commonly used to describe the dynamic bulk modulus of the fluid saturating a porous medium. It has been found more accurate than previous models for arbitrary pore shapes.[3] In this model, using the $\exp(j\omega t)$ time convention, the dynamic bulk modulus of the fluid phase may be written as

$$\tilde{K} = \frac{\gamma P_0}{\gamma - (\gamma - 1)\left(1 - j\frac{\omega_{tA}}{\omega}\tilde{G}_A\right)^{-1}} \tag{1}$$

with the function

$$\tilde{G}_A = \left(1 + j\frac{\omega}{2\omega_{tA}}\right)^{1/2} \tag{2}$$

and the thermal characteristic frequency

$$\omega_{tA} = \frac{8\nu'}{\Lambda'^2}. \tag{3}$$

In Eqs. (1)–(3) P_0 is the static pressure, γ is the specific heat ratio, ω is the angular frequency, $\nu' = \eta/\rho_0$ Pr is the fluid thermal diffusivity, ρ_0 and η are the density and dynamic viscosity of the saturating fluid, and Pr is the Prandtl number. In Eq. (3), Λ' is the thermal characteristic dimension defined from the asymptotic high frequency behavior of $\tilde{K}$.

The dynamic bulk modulus in the porous material cannot be measured directly; however, it is possible to determine the equivalent dynamic bulk modulus of the material $\tilde{K}_{\mathrm{eq}}$. It is given by

$$\tilde{K}_{\mathrm{eq}} = \frac{\tilde{K}}{\phi}, \tag{4}$$

where ϕ is the open porosity of the material. The function $\tilde{K}_{\mathrm{eq}}$ finally depends on two intrinsic parameters of the material.

Assuming the knowledge of $\tilde{K}_{\mathrm{eq}}$ and ϕ, only one unknown remains (i.e., Λ'). After some mathematics, one can separate the real and imaginary parts in Eq. (1). This yields the following two possible solutions for Λ':

$$\Lambda'_{\mathrm{lf}} = 2\delta_t \left\{ \left[-\mathrm{Re}\left(\left(\frac{1 - \tilde{K}_{\mathrm{eq}}/K_a}{1 - \gamma\tilde{K}_{\mathrm{eq}}/K_a} \right)^2 \right) \right]^{-1} \right\}^{1/4} \tag{5}$$

and

$$\Lambda'_{\mathrm{hf}} = \delta_t \left\{ 2\left[-\mathrm{Im}\left(\left(\frac{1 - \tilde{K}_{\mathrm{eq}}/K_a}{1 - \gamma\tilde{K}_{\mathrm{eq}}/K_a} \right)^2 \right) \right]^{-1} \right\}^{1/2}. \tag{6}$$

In these expressions $K_a = \gamma P_0/\phi$ is the equivalent adiabatic bulk modulus of the equivalent fluid and $\delta_t = \sqrt{2\nu'/\omega}$ is the thermal skin depth. The two solutions Λ'_{lf} and Λ'_{hf} are *a priori* different; however, the way the model is built can help to find which one corresponds to the correct definition of the thermal length. Since Λ'_{hf} is obtained from the part that dominates at high frequency, and since Λ' appears in the high frequency asymptotic development of $\tilde{K}$, Λ'_{hf} should be used to estimate Λ'. This conclusion and the meaning of solution Λ'_{lf} will appear clearer when considering the same approach applied to the Lafarge *et al.* model.

B. Lafarge *et al.* model

Driven by the fact that the low frequency development of the Champoux–Allard model is not exact, Lafarge *et al.*[4] have proposed a new expression for the bulk modulus. Using the homogenization technique and the analogy with the Johnson *et al.* model[10] for the dynamic tortuosity, they gave a new expression for the dynamic bulk modulus introducing the parameter k'_0, called "static thermal permeability" and defined from the low frequency behavior of $\tilde{K}$. Their model summarizes by the following expression

$$\tilde{K} = \frac{\gamma P_0}{\gamma - (\gamma - 1)\left(1 - j\frac{\omega_{tL}}{\omega}\tilde{G}_L\right)^{-1}} \tag{7}$$

with the function

$$\tilde{G}_L = \left(1 + j\frac{M'}{2}\frac{\omega}{\omega_{tL}}\right)^{1/2}, \tag{8}$$

the dimensionless thermal shape factor

$$M' = \frac{8k'_0}{\Lambda'^2\phi}, \tag{9}$$

and the thermal characteristic frequency

$$\omega_{tL} = \frac{\phi\nu'}{k'_0}. \tag{10}$$

This new characteristic frequency is slightly different from ω_{tA} given in Eq. (3), and now involves k'_0. The relation between both characteristic frequencies is $\omega_{tA} = \omega_{tL}M'$. Consequently, both models are similar only when $M' = 1$ (i.e., when the dynamic heat conduction is similar to the one in a circular cylindrical pore).

The extraction of the expressions of Λ' and k'_0 from Eq. (7) is very similar to what was done for the Champoux–Allard model. Using $\tilde{K}_{\mathrm{eq}}$ instead of $\tilde{K}$, and choosing the right sign for the square roots, gives

$$\Lambda' = \delta_t \left\{ 2\left[-\mathrm{Im}\left(\left(\frac{1 - \tilde{K}_{\mathrm{eq}}/K_a}{1 - \gamma\tilde{K}_{\mathrm{eq}}/K_a} \right)^2 \right) \right]^{-1} \right\}^{1/2} \tag{11}$$

and

$$k_0' = \phi \frac{\delta_t^2}{2} \left\{ -\mathrm{Re}\left[\left(\frac{1 - \tilde{K}_{\mathrm{eq}}/K_a}{1 - \gamma \tilde{K}_{\mathrm{eq}}/K_a} \right)^2 \right] \right\}^{-1/2}. \tag{12}$$

This time, since we have two equations [real and imaginary parts of Eq. (7)], and two unknowns, Λ' and k_0', there is no ambiguity in the determination. The solution for the characteristic length given by Eq. (11) is equivalent to Λ_{hf}' in Eq. (6), and comparing Eqs. (5) and (12) reveals that

$$k_0' = \phi \frac{\Lambda_{\mathrm{lf}}'^2}{8}.$$

As shown, analytical solutions for Λ' and k_0' can be directly derived from the Lafarge *et al.* model because there are as many equations as unknowns. However, one must keep in mind that Eqs. (11) and (12) cannot be considered as exact solutions for the true parameters since they depend on the initial model, which is not exact itself. The so-obtained solutions are only the best fitted parameters for the model under consideration. It should be mentioned that the Lafarge *et al.* model was improved[11,12] in order to match more exactly the low frequency behavior of $\tilde{K}$ by introducing one more intrinsic geometrical parameter. However, due to the actual difficulties to measure or determine this parameter, it is not studied in this paper.

C. Wilson model

Wilson[5] has proposed an alternative model based on the remark that thermal and viscous effects that occur in porous media have the characteristic of a relaxation process. Contrary to Allard–Champoux and Lafarge *et al.* models, the expression of $\tilde{K}$ given by Wilson is not built to fit the low and high frequency behavior of the material, but it is attempted to match its middle frequency or transition behavior. The Wilson relaxation model can be derived from the analogy with other physical systems such as electrical networks or mechanical systems. One interest of this model is that it requires a limited number of parameters; in return the prediction should be accurate only in the middle frequency range (i.e., when the thermal skin depth is close to the pore size). Using the $\exp(j\omega t)$ time convention, the Wilson model for the bulk modulus is written as

$$\tilde{K} = K_\infty \frac{(1 + j\omega\tau_{\mathrm{ent}})^{1/2}}{(1 + j\omega\tau_{\mathrm{ent}})^{1/2} + \gamma - 1}. \tag{13}$$

According to this expression, the bulk modulus depends on two real characteristic parameters: K_∞ and τ_{ent}. Parameter K_∞ is the high frequency asymptotic value of $\tilde{K}$, and τ_{ent} is the entropy-mode relaxation time. The relation between these parameters and the microstructure of the material is not as explicit as in Champoux–Allard and Lafarge *et al.* models. However, at high frequencies $\tilde{K}$ should naturally tend to its adiabatic value so that K_∞ should be estimated by

$$K_\infty = \phi K_a = \gamma P_0. \tag{14}$$

As in Secs. II A and II B, analytical solutions for parameters K_∞ and τ_{ent} are now derived from $\tilde{K}_{\mathrm{eq}} = \tilde{K}/\phi$. In this case, the equivalent parameter $K_\infty^{\mathrm{eq}} = K_\infty/\phi$ will be used. Consequently, two unknowns are to be found. Since $\tilde{K}_{\mathrm{eq}}$ only depends on two parameters in a relatively simple way, it is possible to invert the Wilson model to find K_∞^{eq} and τ_{ent}. Assuming the equivalent dynamic bulk modulus is obtained experimentally, the analytical separation of the real and imaginary parts in Eq. (13) yields, after some mathematics, a system of two equations:

$$\begin{aligned} &\mathrm{Real} \rightarrow (X - K_\infty^{\mathrm{eq}})^2 - Y^2 - 2\omega\tau_{\mathrm{ent}} Y (X - K_\infty^{\mathrm{eq}}) - (X^2 - Y^2) \\ &\qquad \times (\gamma - 1)^2 = 0, \\ &\mathrm{Imag} \rightarrow \omega\tau_{\mathrm{ent}}((X - K_\infty^{\mathrm{eq}})^2 - Y^2) + 2Y(X - K_\infty^{\mathrm{eq}}) - 2XY(\gamma \\ &\qquad - 1)^2 = 0, \end{aligned} \tag{15}$$

where $X = \mathrm{Re}(\tilde{K}_{\mathrm{eq}})$ and $Y = \mathrm{Im}(\tilde{K}_{\mathrm{eq}})$. Introducing the expression of τ_{ent} obtained from the imaginary part, this system is rewritten as

$$\begin{aligned} &a_0 + a_1 K_\infty^{\mathrm{eq}} + a_2 (K_\infty^{\mathrm{eq}})^2 + a_3 (K_\infty^{\mathrm{eq}})^3 + a_4 (K_\infty^{\mathrm{eq}})^4 = 0, \\ &\tau_{\mathrm{ent}} = \frac{2}{\omega} \frac{XY(\gamma - 1)^2 - Y(X - K_\infty^{\mathrm{eq}})}{(X - K_\infty^{\mathrm{eq}})^2 - Y^2}, \end{aligned} \tag{16}$$

with

$$\begin{aligned} a_0 &= (X^2 + Y^2)^2 (1 - (\gamma - 1)^2), \\ a_1 &= -2X(X^2 + Y^2)^2 (2 - (\gamma - 1)^2), \\ a_2 &= 4(X^2 + Y^2) + (X^2 - Y^2)(2 - (\gamma - 1)^2), \\ a_3 &= -4X, \\ a_4 &= 1, \end{aligned} \tag{17}$$

where K_∞^{eq} is the solution of a fourth-order equation. Since the analytical solutions seem difficult to obtain, the roots of the polynomial are computed numerically. Numerical simulations showed that the equation has generally four real solutions. The most realistic solution is the one that remains the most constant with frequency and the closest to the adiabatic compression modulus of the saturating fluid. The relaxation time is then calculated using this value and the second equation of Eq. (16).

III. ESTIMATION OF THE ERROR

Estimating the error of the proposed characterization method is tricky since several assumptions are used and the precision highly relies on the ability to measure the equivalent dynamic bulk modulus. However, one important point can be addressed. It deals with the error made on the true intrinsic parameters of the material using the analytical solutions. Here, only the analytical solutions given by Eqs. (11) and (12) from the Lafarge *et al.* model are considered.

To test the error of the analytical solutions, two virtual materials are studied. The first one is a layer of parallel circular cylindrical pores of radius R. The second one is a layer of parallel slits of width $2a$. For such simple pore geometries, exact solutions for the dynamic bulk modulus are known.[4] They are recalled in the Appendix. Using these ex-

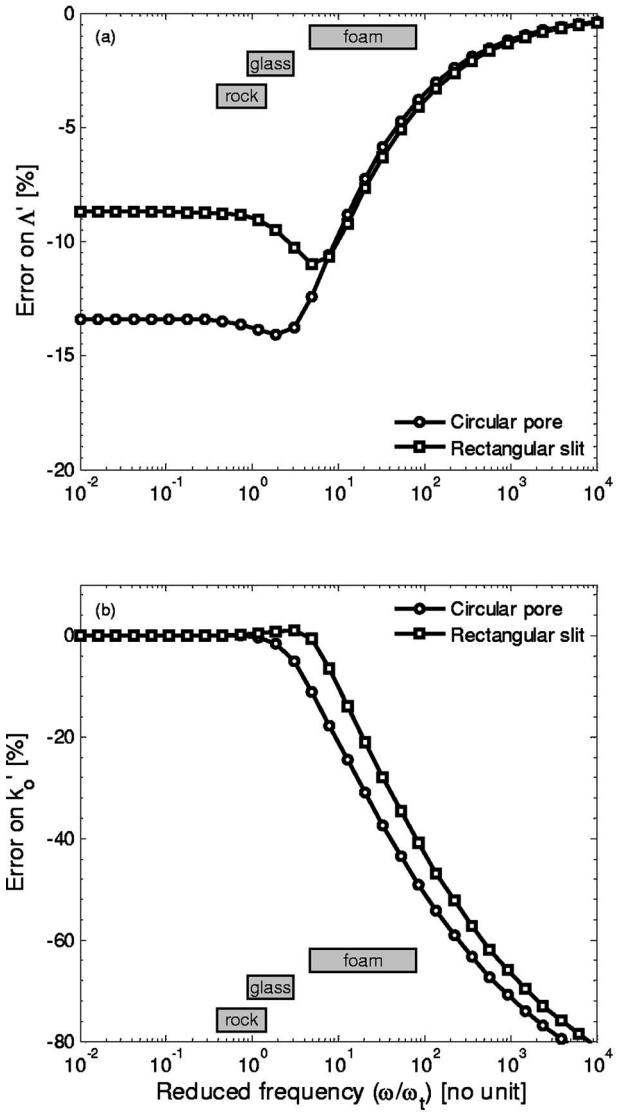


FIG. 1. Errors on the Lafarge *et al.* parameters determined from the inverse analytical formulas. The errors are relative to the theoretical parameters of two virtual air-filled porous materials having (○) cylindrical pores of circular cross section and (□) rectangular slits. (a) Error on the thermal characteristic dimension. (b) Error on the static thermal permeability. The grey rectangles represent the reduced frequency range used to compute the statistics of the characterized thermal parameters of the polyurethane foam, glass wool, and rock wool in Sec. V.

act solutions, measurements of $\tilde{K}_{\mathrm{eq}}$ can be simulated and the thermal parameters of the Lafarge *et al.* model can be determined using the developed analytical solutions of Eqs. (11) and (12).

Let us now consider the relative error made on the determination of a parameter by

$$\delta_u = \frac{\hat{u} - u}{u} \tag{18}$$

with u the theoretical value and $\hat{u}$ the value determined using Eq. (11) or (12). For the studied virtual materials, the theoretical static thermal permeabilities and thermal characteristic dimensions are given by Eqs. (A5), (A6), (A11), and (A12).

Figure 1 shows the relative errors on the analytical determinations for both materials in function of the reduced frequency ϖ. In Fig. 1(a), one can note that for both materials the thermal characteristic dimension is always underestimated. The error starts with a bias error at low frequencies, passes through a maximum (in absolute value), and then decreases to zero as the frequency increases. As for the static thermal permeability, Fig. 1(b) shows that the error tends to zero at low frequencies ($\varpi < 1$), and to a very large value (in absolute value) at high frequencies ($\varpi > 1$). This is logical since k_0' governs the low frequency behavior of the thermal dissipation, and Λ' its high frequency behavior. Consequently, one should use the low frequency range when characterizing k_0', and the high frequency range for Λ'. On the other hand, in the midfrequency range, close to the thermal characteristic frequency ($\varpi = 1$), a maximum absolute error less than 15% is observed for Λ'. This behavior at midfrequencies is linked to the maximum difference between the approximate function G_L given in Eq. (8) and the exact functions G_c and G_s given in Eqs. (A2) and (A8). It is worth mentioning that a 15% error is quite acceptable since this parameter is difficult to obtain with accuracy using other existing methods.

To conclude, the previous error analysis showed that extraction using analytical solutions should proceed preferably at high reduced frequencies ($\varpi > 1$) for Λ', and at low reduced frequencies ($\varpi < 1$) for k_0'. This would minimize the relative errors with respect to the real parameters of the material to characterize. If this cannot be respected, systematic errors on the determination of the parameters may be introduced. Also, for the determination of Λ', the frequency range in the vicinity of $\varpi \approx 1$ should be avoided.

IV. EXPERIMENTAL METHOD

A. Description of the samples

To analyze the applicability of the proposed analytical inversion method, three different porous materials will be tested: (1) One low-resistivity polyurethane foam, (2) one medium-resistivity glass wool, and (3) one high-resistivity rock wool. While the polyurethane foam has a relatively stiff frame, the two mineral wools are relatively soft and their fibers may easily vibrate under acoustic excitations. Preliminary tests have been performed to measure the open porosity of the three materials using the method proposed by Champoux *et al.*[13] The results of these tests are given in Table I. Also, for the sake of completeness, Table I presents the viscous parameters and other properties of the tested materials. The viscous parameters were found using the viscous counterpart of the analytical inversion method presented in Ref. 1.

B. Measurement of the dynamic bulk modulus

To measure the dynamic bulk modulus of each sample, the three-microphone method proposed by Iwase *et al.*[14] in conjunction with a 46-mm-diam acoustic impedance tube is used. The method allows the measurements of the characteristic impedance $\tilde{Z}_c$ and propagation constant $\tilde{\gamma}$ of a porous material. From these measurements, the equivalent dynamic

TABLE I. Open porosity, viscous parameters, and other physical properties of the tested porous materials.

	Symbol	Polyurethane foam	Glass wool	Rock wool	Units
Open porosity	ϕ	0.960	0.988	0.970	
Static airflow resistivity	σ	2 300	24 300	51 200	N s/m^4
Tortuosity	α_∞	1.28	1.01	1.06	
Viscous characteristic dimension	Λ	202	63	41	μm
Bulk density	ρ_1	60	53	183	kg/m^3
Thickness	d	12.7	12.0	11.0	mm

bulk modulus of the material can be deduced from $\tilde{K}_{\text{eq}} = j\omega\tilde{Z}_c/\tilde{\gamma}$. The cut-off frequency of the used impedance tube is f_c=4200 Hz, and its low-frequency limit is approximately 300 Hz (corresponding to a 20-mm spacing between the two upstream microphones).

C. Reduction of frame vibrations

Since the proposed characterization methods assume a rigid-frame behavior (i.e., motionless), the method proposed by Iwase *et al.*[14] to suppress or minimize the frame vibrations of the two mineral wools is used. Briefly, this method consists in sticking nails (or needles) in the porous specimen as shown in Fig. 2. Due to the friction force between the nails and the skeleton, the frame vibrations are minimized and the frame tends to be motionless. Few nails are usually needed since only the first resonant mode is to be suppressed. Higher resonant modes are usually not excited since they occur for frequencies higher than the fluid–solid decoupling frequency.[15–17] It is necessary to wonder if this operation modifies in a significant way the structure of the material so that it could bias the determination of the parameters to identify. An answer to this question can be partially given by estimating the way the porosity and the thermal characteristic dimension are changed. Assuming that the nails only replace air in the sample without modifying the total volume of the material, and that the integrity of the structure is not affected, the relative errors made on ϕ and Λ' as a functions of N (number of nails/cm^2) and d_n (nails diameter in centimeters) are given by

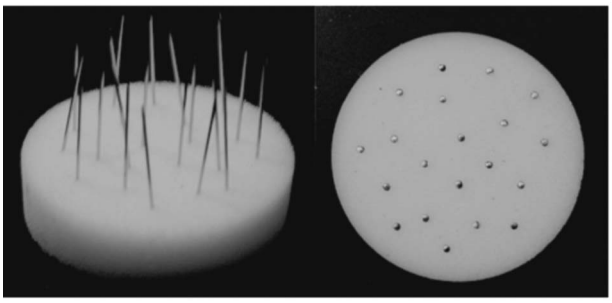


FIG. 2. Impedance tube test specimen with 20 pins used to reduce the vibration of the frame.

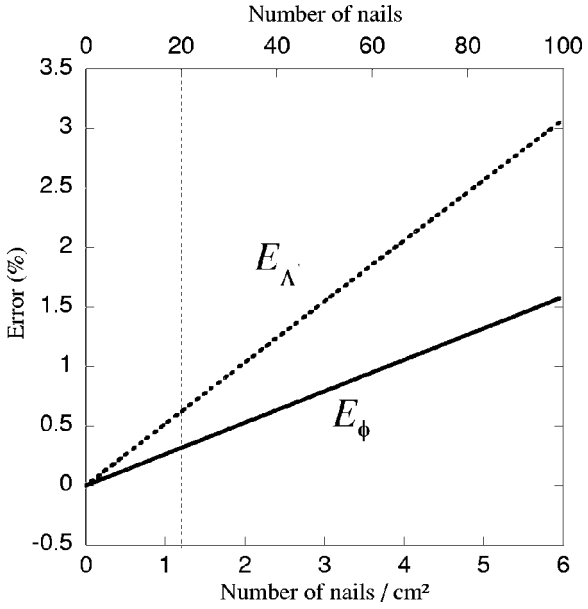


FIG. 3. Estimation of the relative errors made on ϕ and Λ' when the nail technique is used to decrease the vibration of the frame. These errors are calculated for the glass wool using ϕ=0.988 and Λ'=137 μm. The curves are given in function of the number of nails calculated when the surface of the sample equals the section of the impedance tube (≈16.7 cm^2).

$$E_\phi = \frac{\phi - \phi_n}{\phi} = \frac{10\,000}{\phi}\left(\frac{\pi}{4}Nd_n^2\right) \tag{19}$$

and

$$E_{\Lambda'} = \frac{\Lambda' - \Lambda'_n}{\Lambda'} = 10\,000\frac{N\pi d_n(d_n/4 + \Lambda')}{\phi + 10\,000Nd_n\Lambda'\pi}. \tag{20}$$

In Eqs. (19) and (20), the open porosity and thermal characteristic length are defined by $\phi = V_{\text{air}}/(V_{\text{air}}+V_{\text{solid}})$ and $\Lambda' = 2V_{\text{air}}/S_{\text{solid}}$, where V_{air} is the volume of air in the material, V_{solid} is the volume of the solid phase, and S_{solid} is the wet surface of the solid phase (solid surface in contact with the air phase). Also, ϕ_n and Λ'_n represent the porosity and the thermal characteristic dimension of the material with the nails, respectively.

From Eqs. (19) and (20), errors E_ϕ and $E_{\Lambda'}$ have been computed for the glass wool and results reported in Fig. 3. For this material, using 20 nails (i.e., N=1.2 nails/cm^2) fixed perpendicularly to the cross section leads to E_ϕ=0.32% and $E_{\Lambda'}$=0.62%. The nailing operation was also applied to the rock wool material, with the same number of nails, and the found errors were E_ϕ=0.32% and $E_{\Lambda'}$=0.46%. The influence of the nails is more difficult to estimate for parameters such as k'_0 or τ_{ent}; however, these results allow one to reasonably conclude that the nails should not affect significantly the determination of the parameters. Moreover, this preparation can help to increase the frequency range in which the porous medium behaves as rigid, and by the way it can increase the band of determination of the parameters.

V. RESULTS

The analytical inversion formulas are now applied to the three materials described in Sec. IV. For a given material,

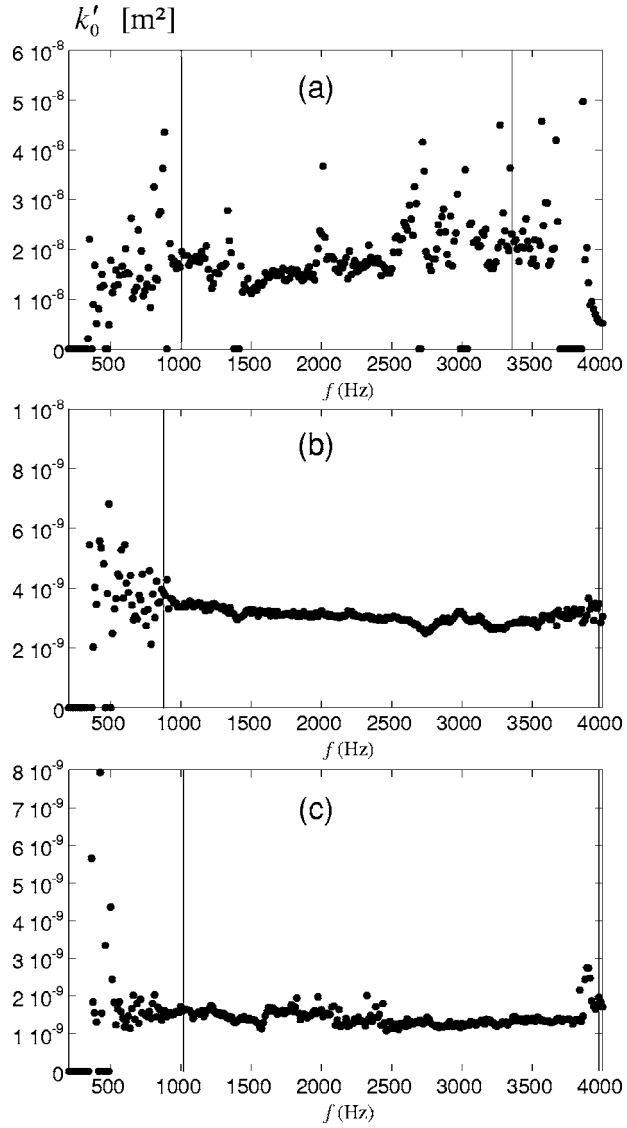


FIG. 4. Static thermal permeability k_0' deduced from the inversion of the Lafarge *et al.* model or from the Champoux–Allard model (using relation $k_0' = \phi\Lambda_{lf}'^2/8$) for: (a) the foam, (b) the glass wool, and (c) the rock wool. The two vertical lines on each graph represent the limits of the frequency range used to compute the statistics reported in Table II.

each inversion formula uses the related open porosity given in Table I, and the dynamic bulk modulus measured as per Sec. IV.

A. Static thermal permeability

Using Eq. (12) [or equivalently Eq. (5) with $k_0' = \phi\Lambda_{lf}'^2/8$] with the measured ϕ (see Table I) and $\tilde{K}_{eq}$ yields the static thermal permeability. Figure 4 shows the static thermal permeabilities found for the three materials at each frequency for which $\tilde{K}_{eq}$ has been measured. It is first noted that at low frequencies, close to and below the low frequency limit (i.e., 300 Hz) of the used impedance tube setup, the random noise is important. With a view to analyze the applicability of the proposed analytical inversion method, this noisy frequency range is ignored. Then, between this limit and the cutoff frequency of the tube, one can verify how the dynamics of the thermal dissipations are captured by the Lafarge *et al.* model. If the model finely describes the dissipation mechanism, the found k_0' should be constant in frequency. This is what is observed for the two wools. For the foam, it seems the model is still applicable; however, the variations are very important. This may be due to frame vibrations since no needles were added to stiffen the frame as done for the wools.

Even in the frequency range where the found k_0' seems stable, some local random variations are observed. To obtain a better appreciation of the static thermal permeability associated with each material, statistics are performed in the range where the model seems mostly valid (i.e., between approximately 1000 and 4000 Hz for the wools, and 1000 Hz and 3350 for the foam). As one can note, the foam has the larger k_0', while the rock wool has the lower one. From the mean values of k_0', the averaged thermal characteristic frequency ω_{tL} is computed using Eq. (10), and reported in Table II for each material. From the characteristic frequency, the reduced frequency range used to compute the statistics on each material is shown in Fig. 1 (grey rectangle). One can note that the range used to compute the statistics for

TABLE II. Thermal characteristic parameters obtained for the three materials using the analytical inversion formulas derived from the Champoux–Allard and Lafarge *et al.* models. The statistics (mean ± standard deviation) are computed in the frequency ranges shown in Figs. 4 and 5. The values in parentheses are the standard deviations in percentage.

	Symbol	Polyurethane foam	Glass wool	Rock wool	Units
Static thermal permeability	k_0'	167 ± 39(23%)	31 ± 3(10%)	12 ± 1(8%)	$\times 10^{-10}$ m^2
Thermal characteristic dimension	$\Lambda' = \Lambda_{hf}'$	376 ± 28(7%)	137 ± 8(6%)	59 ± 8(14%)	μm
Low frequency thermal characteristic dimension	Λ_{lf}'	373 ± 38	158 ± 6	10 ± 6	μm
Thermal pore shape factor	$\langle M' \rangle$	0.98	1.34	2.84	
Thermal characteristic frequency	$\langle \omega_{tL} \rangle$	~200	~1100	~2800	Hz
Reduced frequency	$\langle \varpi \rangle$	~11	~2	~1	

the foam is rather high in terms of reduced frequency. This may also explain why a larger variation ($\pm 23\%$) is obtained for the foam. This is in accordance with the error analysis presented in Sec. III, where it was shown that a more accurate analytical inversion of this low frequency thermal parameter is achieved at lower reduced frequencies.

Finally, according to Fig. 1 and the used characterization frequency range, and assuming the relaxation process follows that of a slit or circular pore, one can conclude that the analytical inversion may introduce a large systematic error on the static thermal permeability found for the foam (approximately 20% bias error), and a very small one for the wools (less than 1%). A more accurate characterization could have been achieved for the foam by using a larger impedance tube. Consequently, the mean values in Table II should be interpreted accordingly.

B. Thermal characteristic dimension

Similar to k_0', using Eq. (11) [or equivalently Eq. (6)] with the measured ϕ (see Table I) and $\tilde{K}_{eq}$ yields the thermal characteristic dimension. Figure 5 shows the analytical inversion results for the three materials as a function of the frequency. Again, noisy data are observed at low frequencies and eliminated from the analysis. As one can note, the Lafarge *et al.* model seems clearly valid on the whole frequency range analyzed for the wools. For the foam, it seems the model is applicable up to 3350 Hz. For higher frequencies, the model starts diverging (i.e., the hypotheses behind the model are no more satisfied).

Similar to k_0', the statistics on Λ' are reported in Table II. The statistics are performed in the same frequency ranges than the ones used for k_0'. Also, statistics on Λ_{lf}' are given. As one can note, the foam has the larger Λ' (i.e., large pores) and the rock wool the smaller one. Furthermore, one can note that the variation on Λ' is the larger for the rock wool ($\pm 14\%$). This can be explained by the fact the reduced frequency range on which the statistics were performed is rather low for the rock wool (see Fig. 1). Since Λ' is a high frequency thermal parameter, this is not the ideal situation for the analytical inversion.

From k_0', Λ', and Eq. (9), the averaged thermal pore shape factor M' is computed and given in Table II. It is noted that its value is close to unity for the foam and successively increasing for the other two materials. This means that the Champoux–Allard model is adequate for the foam, and inappropriate for the tested wools. This is also shown by comparing Λ_{lf}' and Λ_{hf}'. These thermal lengths are almost identical for the foam, and different for the wools. Consequently, using the Champoux–Allard model using a single thermal characteristic length (for instance, the true Λ') is not sufficient to capture the dynamics of the thermal dissipations.

Again, as for k_0', the reduced frequency range used for the analytical inversion of Λ' should be taken into consideration to assess the accuracy of the found value. In the previous analysis, looking at Fig. 1, one can deduce that the mean thermal characteristic lengths found for the three materials have an expected bias error of 9%. To reduce this error, the experimenter should *a posteriori* defines the most suitable setup (size and microphone spacing) for the impedance tube.

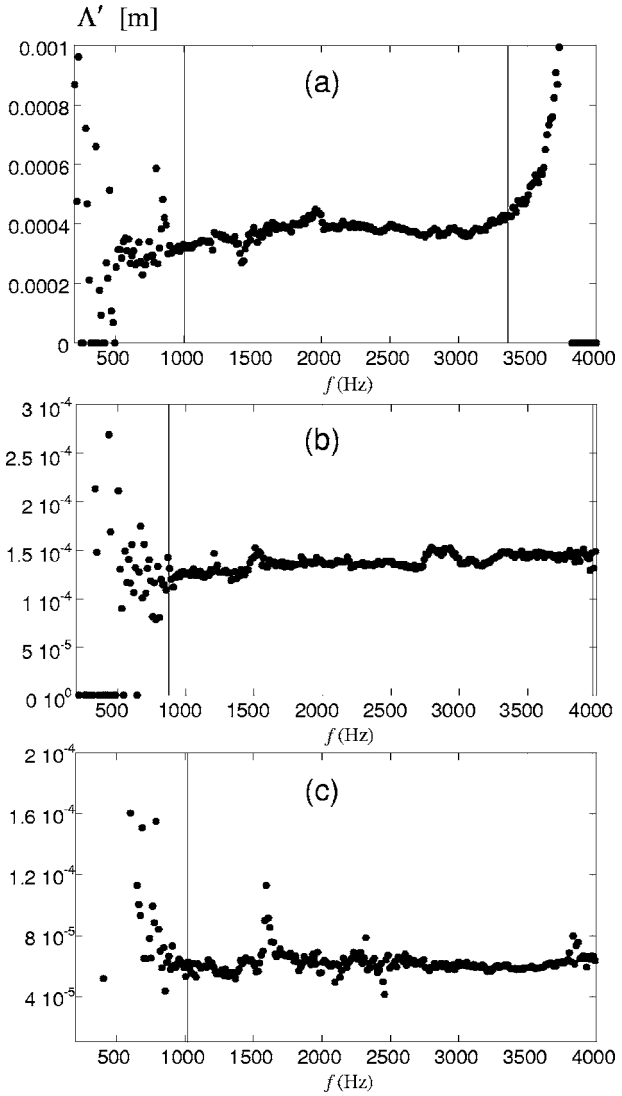


FIG. 5. Thermal characteristic dimension Λ' deduced from the inversion of the Lafarge *et al.* model or Champoux–Allard model for: (a) the foam, (b) the glass wool, and (c) the rock wool. The two vertical lines on each graph represent the limits of the frequency range used to compute the statistics reported in Table II.

C. Wilson thermal parameters

Similar to k_0' and Λ', the thermal parameters of the Wilson model are found by solving the system given in Eq. (16). The admissible results are plotted in Figs. 6 and 7. As one can note, the ranges for which the parameters are stable start at higher frequencies compared to the Lafarge *et al.* model. In fact, the values start stabilizing for frequencies greater than the thermal characteristic frequency. It is in accordance with the fact the Wilson model was developed to fit the thermal dissipation behavior in the vicinity of the transition frequency.

For the ranges where the parameters are relatively constant (i.e., 1000–3500 Hz for the foam, 1000–4000 Hz for the glass wool, and 2500–4000 Hz for the rock wool), the statistics are computed and reported in Table III. For the

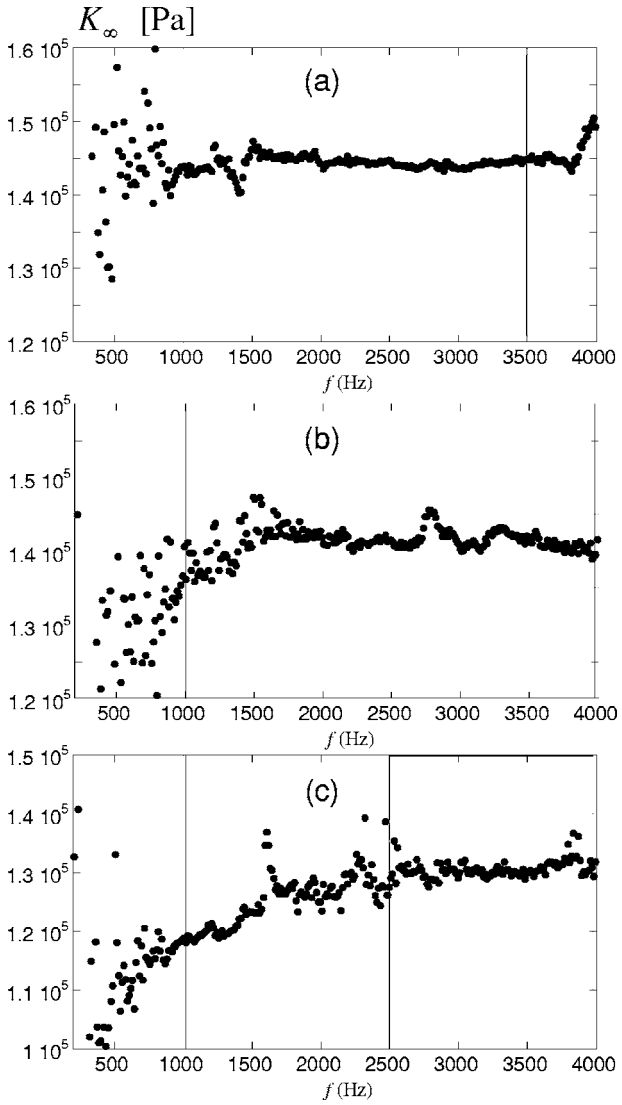


FIG. 6. Thermal parameter K_∞^{eq} deduced from the inversion of the Wilson model for: (a) the foam; (b) the glass wool, and (c) the rock wool. The two vertical lines on each graph represent the limits of the frequency range used to compute the statistics reported in Table III.

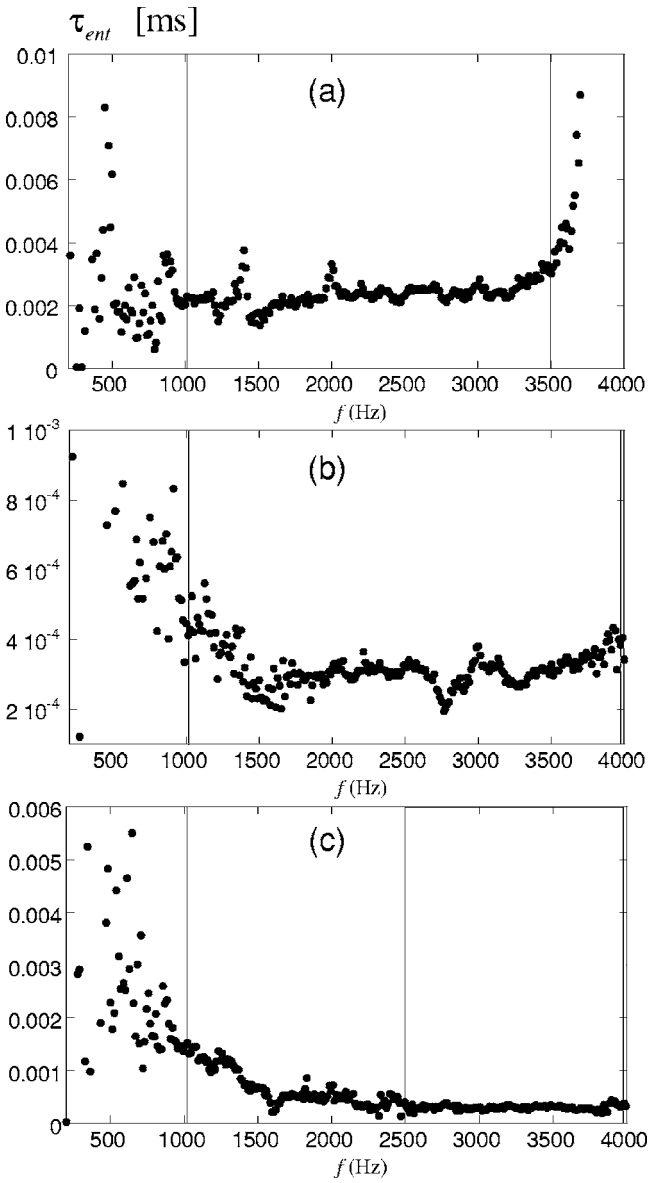


FIG. 7. Thermal parameter τ_{ent} deduced from the inversion of the Wilson model for: (a) the foam; (b) the glass wool, and (c) the rock wool. The two vertical lines on each graph represent the limits of the frequency range used to compute the statistics reported in Table III.

Wilson adiabatic bulk modulus K_∞^{eq}, the values found for the three materials tend to the adiabatic bulk modulus of the air. This confirms the use of $K_\infty^{eq}=K_a$ as proposed by Wilson.[18] In this case, only the second equation in Eq. (16) can be used. For the entropy-mode relaxation time τ_{ent}, one can observe that the foam has a very large value compared to the wools. This is explained by the fact that the thermal relaxation process is slower in larger cells, pushing the transition frequency downwards. Based on the Champoux–Allard model, Wilson proposed an approximation of this relaxation time by $\tau_A = 2/\omega_{tA}$. Similarly, one can propose a similar approximation based on the Lafarge *et al.* model by $\tau_L = 2/\omega_{tL}$. Values of τ_A and τ_L are given in Table III. One can note that τ_L is a better approximation of τ_{ent} than τ_A.

It is worth mentioning that this approach to define τ_L is not strictly rigorous. A more appropriate approach would use the extended Wilson model[18] involving two entropy-mode relaxation times, which is more compatible with the Lafarge *et al.* model. Here the authors preferred to work with the original Wilson model, since it is simpler (practically only one parameter is to be found) and because, contrary to the other models, it aims to fit the midfrequency behavior and not the low and high frequency limits.

D. Predictions of dynamic bulk modulus

Using the thermal parameters found from the analytical inversions (see Tables II and III), and the directly measured open porosities (see Table I), predictions of the dynamic bulk modulus for the three materials are compared to measurements in Figs. 8–10, respectively. Predictions using the Champoux–Allard model [Eq. (1)], the Lafarge *et al.* model [Eq. (7)], and the Wilson model [Eq. (13)] are also compared. From these comparisons, it is noted that better correlations with measurements are obtained with the Lafarge *et al.* model for the three materials and for the whole fre-

TABLE III. Thermal parameters obtained for the three materials using the analytical inversion formula derived for the Wilson model. The statistics (mean ± standard deviation) are computed in the frequency ranges shown in Figs. 6 and 7.

Relaxation time	Symbol	Polyurethane foam	Glass wool	Rock wool	Units
Wilson adiabatic bulk modulus	$\gamma P_0/K_\infty$	0.964 ± 0.007	0.984 ± 0.013	1.065 ± 0.011	
Wilson	τ_{ent}	2.39 ± 0.35	0.32 ± 0.05	0.30 ± 0.04	ms
Champoux–Allard	$\tau_A = 2/\omega_{tA}$	1.66 ± 0.25	0.22 ± 0.03	0.04 ± 0.01	ms
Lafarge *et al.*	$\tau_L = 2/\omega_{tL}$	1.63 ± 0.38	0.29 ± 0.03	0.12 ± 0.01	ms

quency range. The only exception is for the foam for frequencies greater than 3350 Hz, where an unexpected behavior is observed. The second model to better fit the measurements is the Wilson model—particularly for frequencies greater than the thermal transition frequency (identified by the maximum of the imaginary part). On the other hand, as expected, the Champoux–Allard model only fits the measurements for the foam, since it is the only material having a thermal shape factor approximately equal to unity. The discrepancies between the Champoux–Allard model and the measurements are clearly visible for the rock wool which has the largest thermal shape factor ($M' = 2.84$).

VI. CONCLUSION

In this work, an analytical inversion method was derived for the Champoux–Allard, Lafarge *et al.*, and Wilson models to characterize the parameters governing the dynamic thermal heat conduction phenomena in open-cell porous media. The method was tested on three materials of different static flow resistivities (2300 to 51 200 Ns/m^4), frame rigidities (nonresonant and resonant), and pore geometries (cells and fibers). To operate, the method requires an acoustical measurement of the dynamic bulk modulus of the tested material.

From the Champoux–Allard model, it was found that two solutions for the thermal characteristic dimension are obtained from the analytical inversion of the model—a low frequency and a high frequency solution. From the Lafarge *et al.* model, it was shown that the low frequency solution of the Champoux–Allard model can be related to the static thermal permeability of the material. This ambiguity in the Champoux–Allard model explicitly explains the limitation of

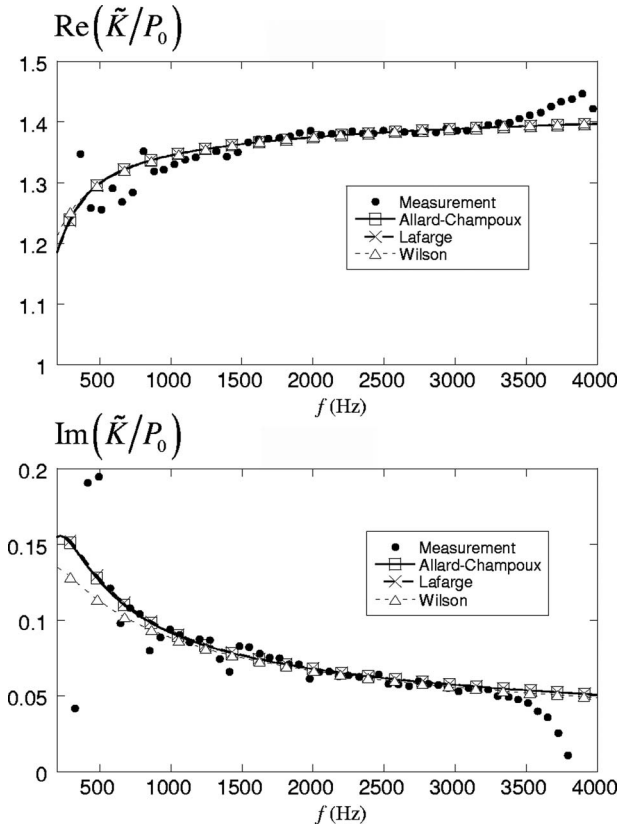


FIG. 8. Comparison between the measured and calculated normalized dynamic bulk modulus of the foam. (a) Real part. (b) Imaginary part. The parameters used in the models are the ones determined from the analytical inversion process.

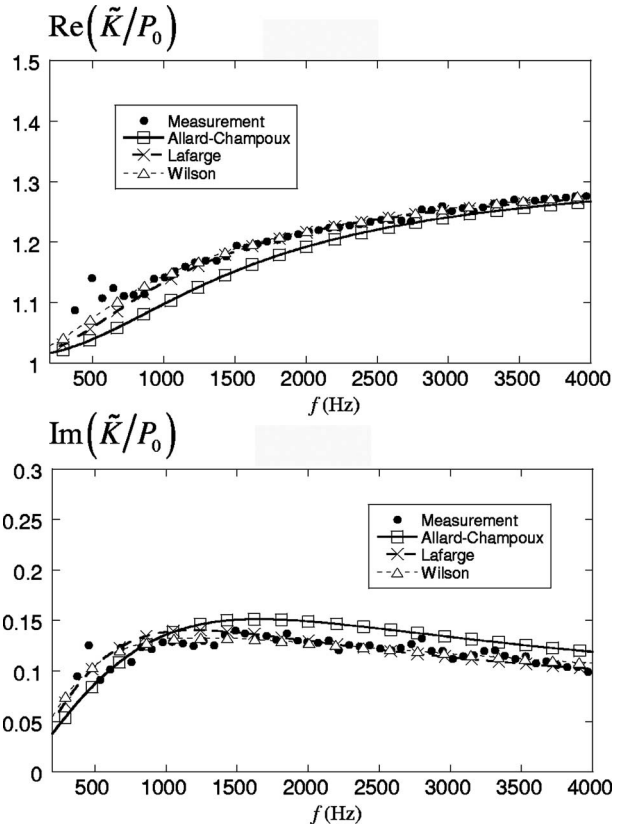


FIG. 9. Comparison between the measured and calculated normalized dynamic bulk modulus of the glass wool. (a) Real part. (b) Imaginary part. The parameters used in the models are the ones determined from the analytical inversion process.

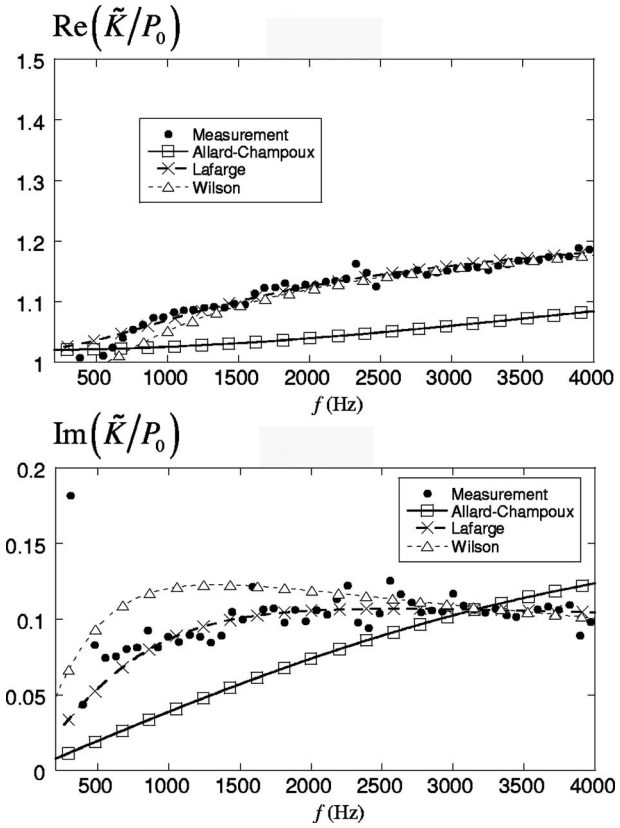


FIG. 10. Comparison between the measured and calculated normalized dynamic bulk modulus of the rock wool. (a) Real part. (b) Imaginary part. The parameters used in the models are the ones determined from the analytical inversion process.

the model to predict the thermal dissipation behavior of materials with thermal pore shape factors greater than unity (such as typical fibrous materials).

To estimate the precision of the analytical inversion method, the method was tested on two virtual materials of theoretically known properties. From this study, it was found that the method should yield small systematic errors (or bias errors) on the thermal characteristic length when tested at reduced frequencies greater than unity. The maximum error is 15% (underestimation) obtained at a reduced frequency slightly greater than unity. Conversely, small systematic errors (less than 1%) on the static thermal permeability should be obtained when the method is tested at reduced frequencies smaller than unity. In addition to the assessment of the precision of the method, Fig. 1 obtained from this error analysis can be used to select the frequency range (and the impedance tube setup) in which the extraction should proceed to minimize the error of the method.

For the Wilson model, the derived analytical inversion formula was found to be a secure characterization method to infer representative values for K_∞^{eq} and τ_{ent} in the Wilson dynamic bulk modulus. Also, it was shown that K_∞^{eq} can be replaced by the adiabatic bulk modulus of the air (K_a), and that a better estimate to τ_{ent} is given by $\tau_L=2/\omega_{tL}$ as opposed to $\tau_c=2/\omega_{tA}$ as originally proposed by Wilson. One advantage of the Wilson model over the Champoux–Allard model is that the characterization of the two Wilson's parameters does not require prior knowledge of the open porosity.

The application of the analytical inversion method for the three models has shown relative constancy in the identified thermal characteristic parameters in function of the frequency. This constancy can be used to assess the validity of: (1) the descriptive models in a given frequency range, and (2) the parameters found from the proposed characterization methods. However, one has to recall that an *a posteriori* analysis needs to be performed (once the parameters have been identified) to have an estimate of the potential bias error made on the found parameters using Fig. 1.

Once the entire model parameters have been carefully characterized, comparisons between predictions and measurements in terms of the dynamic bulk modulus showed that the Lafarge *et al.* model is the one that best fits the measurements, followed by the Wilson model. Also, it was shown that the Champoux–Allard model should be avoided, especially when characterizing materials known to have an M' very different from 1 (wools for instance).

Again, the good correlations with measurements, together with the frequency independence in the found parameters, reinforce the fact that the proposed analytical inversion method offers an elegant alternative to existing characterization methods. Moreover, since the method only relies on equations and a widespread apparatus (impedance tube), this makes the characterization of porous materials possible for many acoustic laboratories. However, one must keep in mind that the parameters determined with the method are the best parameters fitting a chosen model and that they can be slightly different from the "real" intrinsic parameters. The representation of the "frequency dependence" can help to justify or question the choice of this model.

To conclude, the authors believe that comparisons with ultrasound techniques are necessary to complete this work—this was not possible here using the available laboratory equipments. Also, to improve the reliability of the proposed method, works on improving the accuracy of the measurements of the dynamic properties (characteristic impedance, propagation constant, dynamic bulk modulus) and on the minimization of the frame vibrations are necessary.

ACKNOWLEDGMENTS

The authors would like to thank N.S.E.R.C. Canada and F.Q.N.R.T. Quebec for their financial support.

APPENDIX A: THEORETICAL MODELS

In this appendix we recall the theoretical expressions of the dynamic bulk modulus for rigid-frame materials containing air-filled (1) circular cylindrical pores, and (2) parallel rectangular slits.

1. Porous material with circular cylindrical pores

The equivalent theoretical dynamic bulk modulus of a layer having parallel cylindrical pores of uniform circular cross section is given by

$$\tilde{K}_{\mathrm{eq}}^{\mathrm{cyl}} = \frac{\gamma P_0/\phi}{\gamma - (\gamma-1)\left(1 - j\frac{1}{\varpi}G_c\right)^{-1}}, \tag{A1}$$

with

$$G_c = -\frac{s\sqrt{-j}}{4}\frac{J_1(s\sqrt{-j})}{J_0(s\sqrt{-j})} \Bigg/ \left[1 - \frac{2}{s\sqrt{-j}}\frac{J_1(s\sqrt{-j})}{J_0(s\sqrt{-j})}\right] \tag{A2}$$

and

$$s = \sqrt{8\varpi}. \tag{A3}$$

In the previous equations, J_0 and J_1 are the Bessel functions of the first kind of order zero and one, and the reduced frequency ϖ is computed using the thermal characteristic frequency given by Eq. (10).

Considering n cylindrical pores of radius R per unit area, the open porosity, static thermal permeability, and thermal characteristic dimension are, respectively,

$$\phi = n\pi R^2, \tag{A4}$$

$$k_0' = \frac{\phi R^2}{8}, \tag{A5}$$

$$\Lambda' = R. \tag{A6}$$

2. Porous material with rectangular slits

The equivalent theoretical dynamic bulk modulus of a layer having parallel rectangular slits is given by

$$\tilde{K}_{\mathrm{eq}}^{\mathrm{slit}} = \frac{\gamma P_0/\phi}{\gamma - (\gamma-1)\left(1 - j\frac{1}{\varpi}G_s\right)^{-1}}, \tag{A7}$$

with

$$G_s = \frac{s'\sqrt{j}}{3}\tanh(s'\sqrt{j}) \Bigg/ \left[1 - \frac{\tanh(s'\sqrt{j})}{s'\sqrt{j}}\right] \tag{A8}$$

and

$$s' = (3\varpi)^{1/2}. \tag{A9}$$

Considering n slits of width $2a$ per unit length, the open porosity, static thermal permeability, and thermal characteristic dimension are, respectively,

$$\phi = 2na, \tag{A10}$$

$$k_0' = \frac{\phi a^2}{3}, \tag{A11}$$

$$\Lambda' = 2a. \tag{A12}$$

[1] R. Panneton and X. Olny, "Acoustical determination of the parameters governing viscous dissipation in porous media," J. Acoust. Soc. Am. **119**, 2027–2040 (2006).
[2] Y. Champoux and J.-F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).
[3] J.-F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Elsevier Applied Science, New York, 1993).
[4] D. Lafarge, P. Lemarinier, J.-F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[5] D. K. Wilson, "Relaxation-matched modeling of propagation through porous media, including fractal pore structure," J. Acoust. Soc. Am. **94**, 1136–1145 (1993).
[6] Y. Atalla and R. Panneton, "Inverse acoustical characterization of open-cell porous media using impedance tube measurements," Can. Acoust. **33**, 11–24 (2005).
[7] R. Panneton, "Comments on the limp frame equivalent fluid model for porous media," J. Acoust. Soc. Am. **122**, EL217–EL222 (2007).
[8] M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[9] M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher-frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[10] D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[11] D. Lafarge, "Sound propagation in porous materials having a rigid frame saturated by gas," (in French), Ph.D. dissertation, Université du Maine, France, 1993.
[12] S. R. Pride, F. D. Morgan, and A. F. Gangi, "Drag forces of porous-medium acoustics," Phys. Rev. B **47**, 4964–4975 (1993).
[13] Y. Champoux, M. R. Stinson, and G. A. Daigle, "Air-based system for the measurement of the porosity," J. Acoust. Soc. Am. **89**, 910–916 (1990).
[14] T. Iwase, Y. Izumi, and R. Kawabata, "A new measuring method for sound propagation constant by using sound tube without any air spaces back of a test material," paper presented at Internoise 98, Christchurch, New Zealand, 1998.
[15] O. C. Zwikker and C. W. Kosten, *Sound-Absorbing Materials* (Elsevier, Amsterdam, 1949).
[16] D. Pilon, R. Panneton, and F. Sgard, "Behavioral criterion quantifying the edge-constrained effects on foams in the standing wave tube," J. Acoust. Soc. Am. **114**, 1980–1987 (2003).
[17] D. Pilon, R. Panneton, and F. Sgard, "Frame acoustical excitability: A decoupling criterion for poroelastic materials," Proceedings of Euronoise, Naples, Italy, 2003.
[18] D. K. Wilson, "Simple, relaxational models for the acoustical properties of porous media," Appl. Acoust. **50**, 171–188 (1997).

Effects of an air-layer-subdivision technique on the sound transmission through a single plate

Masahiro Toyoda[a)]
Kyoto University Pioneering Research Unit, B104, Kyoto University Katsura, Nishikyo-ku, Kyoto, 615-8530, Japan

Hajime Kugo and Takafumi Shimizu
Department of Urban and Environmental Engineering, Graduate School of Engineering, Kyoto University, C1-4-392, Kyoto University Katsura, Nishikyo-ku, Kyoto, 615-8540, Japan

Daiji Takahashi
Department of Urban and Environmental Engineering, Graduate School of Engineering, Kyoto University, C1-4-352, Kyoto University Katsura, Nishikyo-ku, Kyoto, 615-8540, Japan

(Received 11 June 2007; accepted 14 November 2007)

Many studies on the sound transmission through a single plate have been carried out theoretically and experimentally. The transmission-loss characteristics, in general, follow mass law. Therefore, increasing mass of a plate is a fundamental measure to improve the insulation performance. This method, however, has limitations and might not be a reasonable alternative in current standards. Furthermore, the transmission loss at the critical frequency of coincidence is deteriorated significantly even if the mass is rather large. In this paper, the effect of the air-layer-subdivision technique is studied in detail from the viewpoint of the sound transmission problem of a single plate. An analytical model of an infinite single plate with a subdivided layer is considered and the improvement of the transmission loss is estimated. The limitations of the technique are clarified with some parametric studies. In order to validate the predictions, an experiment was carried out. The transmission loss of a glass board with the air layer subdivided by acryl partitions was measured in the experiment. They were in good agreement with the theoretical ones near and above the coincidence. [DOI: 10.1121/1.2821981]

PACS number(s): 43.55.Ti, 43.55.Ev, 43.55.Rg [NX] Pages: 825–831

I. INTRODUCTION

Many studies on the sound transmission through a single plate have been carried out theoretically and experimentally. The transmission-loss characteristics, in general, follow mass law. Therefore, increasing mass of a plate is a fundamental measure to improve the insulation performance. This method, however, has limitations and might not be a reasonable alternative in current standards. Furthermore, the transmission loss at the critical frequency of coincidence is deteriorated significantly even if the mass is rather large.

In order to improve the insulation performance without increasing the mass, Mulholland[1] and Brown *et al.*[2] conducted experimental studies on the acoustic radiation from a single plate with an absorptive layer. This structure was investigated theoretically by Takahashi,[3] allowing for not only mass law but flexural vibrations. In the paper, the case that the vibrating surface itself has absorptive characteristics was also studied. Toyoda *et al.*[4,5] investigated its characteristics in detail and proposed a sound insulation structure which can provide the attenuation effects at an arbitrary frequency including low frequency. Toyoda *et al.*[6] also proposed the air-layer-subdivision technique for double-leaf structures supported in a rectangular duct of finite cross section. By constraining the particle motion in the air layer to move in the direction normal to the plates, the sound radiation is reduced because the mode resonances in the air layer are suppressed. In the paper, further possibilities of the motion constraint for increasing the transmission loss of a single-leaf window and a double-leaf wall are introduced. The technique can be achieved by attaching or inserting partitions like honeycomb structure. The configurations with combinations of partitions and fibrous materials are also introduced. However, as for the effects on transmission loss, only the calculated results are shown and the validity is not presented. The effects of such partitions have been studied on its surface wave,[7,8] which is a similar surface wave observed on an absorptive layer.[9–12] It is interesting that there is the similarity between partitions and absorbing materials although they have different characteristics on energy absorption; fibrous and porous materials can convert vibration energy into thermal energy for the sake of sound absorption, while the partitions do not absorb any energy.

In this paper, the effect of the air-layer-subdivision technique is studied in detail from the viewpoint of the sound transmission problem of a single plate. An analytical model of an infinite single plate with a subdivided layer and a decoupling air space between the plate and the layer is considered and the improvement of the transmission loss is estimated. The limitations of the technique are clarified with the parametric studies on the depth of the partition and the decoupling air space. In order to validate the predictions, an

[a)]Electronic mail: masahiro.toyoda@kupru.iae.kyoto-u.ac.jp.

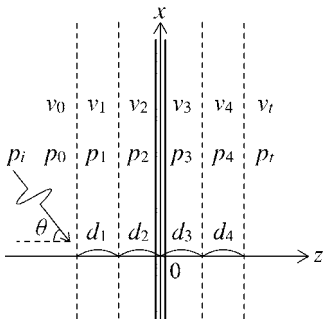


FIG. 1. An analytical model of a multilayer single plate located at $z=0$. An incident wave $p_i(x,z)$ is assumed to be plane wave with incident angle θ. Layers of depth d_1, d_2 and layers of d_3, d_4 are attached on incident side and radiation side of the plate, respectively.

experiment was carried out. The transmission loss of a glass plate with the air layer subdivided by acryl partitions was measured in the experiment. Although the improvement cannot be seen at low frequencies in the experiment due to finite dimensions of the plate, which make conditions different from the analytical model, experimental results are in good agreement with the predictions near and above the critical frequency of coincidence. The wide use of the subdivision technique is anticipated because the improvement can be achieved around the critical frequency of coincidence without fibrous and porous materials, which are not desirable under severe environments.

II. ANALYTICAL MODEL

Here, in order to investigate the possibility of the air-layer-subdivision technique, an analytical model for calculation of the transmission loss through an infinite single plate with multilayer is introduced. Motion-constraint condition, which is achieved by subdividing the air layer, is implemented by assuming that the wave propagation in the layer is confined to normal direction to the plate.

A. Formulation

Figure 1 shows an analytical model considered in this paper. A core plate is located at $z=0$ and four layers of depth d_1, d_2, d_3, and d_4 are attached on both sides of the plate as shown in Fig. 1. An incident wave $p_i(x,z)$ is assumed to be plane wave given by

$$p_i(x,z) = e^{ik_0\{x\sin\theta(z+d_1+d_2)\cos\theta\}}, \tag{1}$$

where i is the imaginary unit, k_0 is wave number of air, and θ is incident angle. The time factor $e^{-i\omega t}$ is suppressed throughout, where ω is the angular frequency.

In the following discussion, the incident side, inside of the layers, and the transmission side are indicated by subscripts 0, 1, 2, 3, 4, and t, respectively. With the aid of a Helmholtz integral formula for a two-dimensional problem, the sound pressure $p_0(x)$ at $z=-d_1-d_2$ and $p_t(x)$ at $z=d_3+d_4$ can be expressed[3] as

$$p_0(x) = 2p_i(x,-d_1-d_2) - \frac{\rho_0\omega}{2} \times \int_{-\infty}^{\infty} \mathbf{H}_0^{(1)}(k_0|x-x_0|)v_0(x_0)dx_0, \tag{2}$$

$$p_t(x) = \frac{\rho_0\omega}{2}\int_{-\infty}^{\infty} \mathbf{H}_0^{(1)}(k_0|x-x_0|)v_t(x_0)dx_0, \tag{3}$$

where ρ_0 is the density of air, $\mathbf{H}_0^{(1)}$ is the Hankel function of the first kind of order zero, and $v_0(x)$ and $v_t(x)$ are the particle velocities at $z=-d_1-d_2$ and $z=d_3+d_4$, respectively. If the layers are filled with air, the sound pressure $p_j(x,z)$ and the particle velocity $v_j(x,z)$ of the jth layer $(j=1,2,3,4)$ can be expressed in terms of unknown quantities $P_j^{\pm}$ as

$$p_j(x,z) = (P_j^+e^{ik_jz} + P_j^-e^{-ik_jz})e^{ik_j'x}, \tag{4}$$

$$v_j(x,z) = K_j(P_j^+e^{ik_jz} - P_j^-e^{-ik_jz})e^{ik_j'x}, \tag{5}$$

where k_j is the z-directional wave number, k_j' is the x-directional wave number, and $K_j=k_j/(\rho_0\omega)$. When jth layer is free field, $k_j=k_0\cos\theta$ and $k_j'=k_0\sin\theta$. When particle motion in the layer is constrained to move in the direction normal to the plate, $k_j=k_0$ and $k_j'=k_0\sin\theta$. In this case, k_j' does not mean the x-directional wave number but the distribution constant determined by boundary conditions.

The force per unit area that excites the core plate is given as the pressure difference $p_2(x,0)-p_3(x,0)$. Let the displacement of the plate excited by a unit force $\delta(x)$, which is the Dirac delta function, be denoted by $u_B(x)$; then the equation of motion of the plate is written by

$$\frac{\partial^4 u_B(x)}{\partial x^4} - k_B^4 u_B(x) = \frac{\delta(x)}{D_B}, \tag{6}$$

where $k_B^4=(\omega/c_B)^4=\rho_B h_B\omega^2/D_B$, c_B is the phase velocity of the bending wave which is depending on the frequency, ρ_B is the density, h_B is the thickness, and D_B is the flexural rigidity of the plate. From the convolution theorem, the displacement of the plate $w_B(x)$ can be expressed as

$$w_B(x) = \int_{-\infty}^{\infty}\{p_2(\xi,0)-p_3(\xi,0)\}u_B(x-\xi)d\xi. \tag{7}$$

The boundary conditions at each surface in Fig. 1 are given by

$$p_0(x) = p_1(x,-d_1-d_2), \tag{8}$$

$$v_0(x) = v_1(x,-d_1-d_2), \tag{9}$$

$$p_1(x,-d_2) = p_2(x,-d_2), \tag{10}$$

$$v_1(x,-d_2) = v_2(x,-d_2), \tag{11}$$

$$v_2(x,0) = v_3(x,0) = -i\omega w_B(x), \tag{12}$$

$$p_3(x,d_3) = p_4(x,d_3), \tag{13}$$

$$v_3(x,d_3) = v_4(x,d_3), \tag{14}$$

$$p_4(x,d_3+d_4)=p_t(x), \tag{15}$$

$$v_4(x,d_3+d_4)=v_4(x). \tag{16}$$

B. Solution

By substituting Eqs. (4) and (5) into Eqs. (8)–(16), $p_2(x,0)$, $p_3(x,0)$, $v_0(x)$, and $v_t(x)$ can be expressed with $p_0(x)$, $p_t(x)$, and $w_B(x)$ as

$$p_2(x,0)=\alpha_0 p_0(x)+i\rho_0 c_0\omega\beta_0 w_B(x), \tag{17}$$

$$p_3(x,0)=\alpha_t p_t(x)+i\rho_0 c_0\omega\beta_t w_B(x), \tag{18}$$

$$v_0(x)=\frac{\zeta_0}{\rho_0 c_0}p_0(x)-i\omega\alpha_0 w_B(x), \tag{19}$$

$$v_t(x)=-\frac{\zeta_t}{\rho_0 c_0}p_t(x)-i\omega\alpha_t w_B(x), \tag{20}$$

where

$$\alpha_{0,t}=\frac{K_{1,4}}{K_{1,4}\cosh E_{1,4}\cosh E_{2,3}+K_{2,3}\sinh E_{1,4}\sinh E_{2,3}}, \tag{21}$$

$$\beta_{0,t}=-\frac{K_{1,4}\cosh E_{1,4}\sinh E_{2,3}+K_{2,3}\sinh E_{1,4}\cosh E_{2,3}}{\rho_0 c_0 K_{2,3}(K_{1,4}\cosh E_{1,4}\cosh E_{2,3}+K_{2,3}\sinh E_{1,4}\sinh E_{2,3})}, \tag{22}$$

$$\zeta_{0,t}=-\frac{\rho_0 c_0 K_{1,4}(K_{1,4}\sinh E_{1,4}\cosh E_{2,3}+K_{2,3}\cosh E_{1,4}\sinh E_{2,3})}{K_{1,4}\cosh E_{1,4}\cosh E_{2,3}+K_{2,3}\sinh E_{1,4}\sinh E_{2,3}}, \tag{23}$$

and $E_j=ik_jd_j$ $(j=1,2,3,4)$. Combining Eqs. (17) and (18) with Eq. (7) yields

$$w_B(x)=\int_{-\infty}^{\infty}\{\alpha_0 p_0(\xi)-\alpha_t p_t(\xi)+i\rho_0 c_0\omega \times(\beta_0+\beta_t)w_B(\xi)\}u_B(x-\xi)d\xi. \tag{24}$$

The solutions can be obtained analytically in the wave number space using the Fourier transform techniques. The transform pairs with respect to x and the wave number space k are defined here as

$$w_B(x)=\int_{-\infty}^{\infty}W_B(k)e^{ikx}dk, \tag{25}$$

$$W_B(k)=\frac{1}{2\pi}\int_{-\infty}^{\infty}w_B(x)e^{-ikx}dx. \tag{26}$$

Taking the transform of Eqs. (6) and (24) gives

$$U_B(k)=\frac{1}{2\pi D_B(k^4-k_B^4)}, \tag{27}$$

$$W_B(k)=2\pi\{\alpha_0 P_0(k)-\alpha_t P_t(k)+i\rho_0 c_0\omega(\beta_0+\beta_t)W_B(k)\}U_B(k), \tag{28}$$

where $U_B(k)$, $P_0(k)$, and $P_t(k)$ are the transformed expressions of $u_B(x)$, $p_0(x)$, and $p_t(x)$, respectively.

By using the integral representation of the Hankel function, which is

$$\mathbf{H}_0^{(1)}(k_0|x-x_0|)=\int_{-\infty}^{\infty}\frac{e^{ik(x-x_0)}}{\pi\sqrt{k_0^2-k^2}}dk, \tag{29}$$

one can obtain the following expressions from Eqs. (2), (3), (19), and (20):

$$P_0(k)=\{2\sqrt{k_0^2-k^2}\delta(k-k_0\sin\theta)+i\rho_0\omega^2\alpha_0 W_B(k)\}/\epsilon_0(k), \tag{30}$$

$$P_t(k)=-i\rho_0\omega^2\alpha_t W_B(k)/\epsilon_t(k), \tag{31}$$

where

$$\epsilon_{0,t}(k)=\sqrt{k_0^2-k^2}+k_0\zeta_{0,t}. \tag{32}$$

Solving Eqs. (27), (28), (30), and (31) simultaneously yields

$$W_B(k)=F(k)\sqrt{k_0^2-k^2}\epsilon_t(k)\delta(k-k_0\sin\theta), \tag{33}$$

where

$$F(k)=\frac{4\pi\alpha_0 U_B(k)}{\epsilon_0(k)\epsilon_t(k)-2\pi i\rho_0\omega U_B(k)\{\omega\alpha_0^2\epsilon_t(k)+\omega\alpha_t^2\epsilon_0(k)+c_0(\beta_0+\beta_t)\epsilon_0(k)\epsilon_t(k)\}}. \tag{34}$$

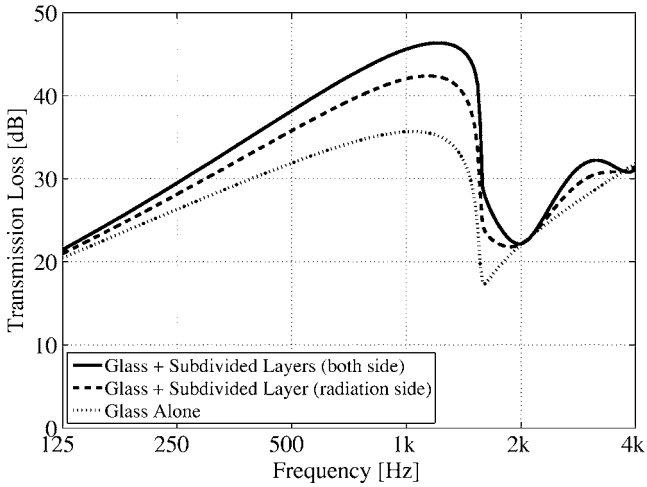


FIG. 2. Calculated results of the transmission loss through a multilayer single plate under field incidence of plane waves. The plate of thickness 8 mm is assumed to be made of window glass. The cases of the plate alone, radiation-side subdivision, and both-side subdivision are shown.

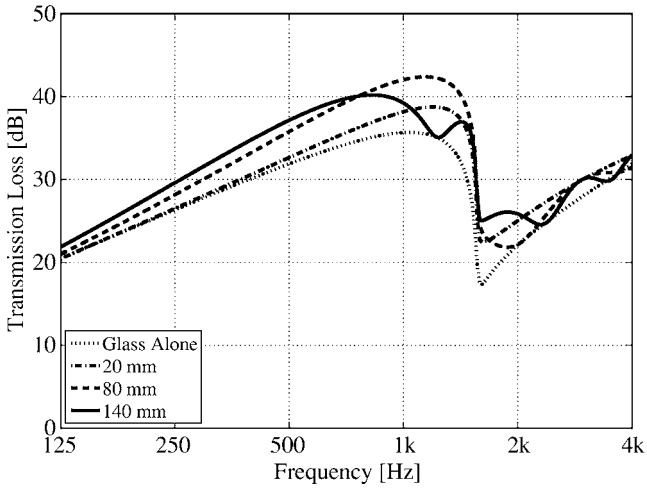


FIG. 3. Effects due to the depth of the subdivided layer. The subdivided layer is located at radiation side with a space 10 mm. Although the transmission loss increases at low frequencies as the depth is larger, resonance effects within the cell shift to lower frequencies. It is indicated that there is an optimal value of the depth in light of the maximal transmission loss at the target frequency.

From Eqs. (20) and (31)–(33), $p_t(x)$ and $v_t(x)$ can be written by

$$p_t(x) = -i\rho_0\omega^2 k_0(\cos\theta)\alpha_t F(k_0 \sin\theta)e^{ik_0 x \sin\theta}, \tag{35}$$

$$v_t(x) = -i\omega k_0^2(\cos^2\theta)\alpha_t F(k_0 \sin\theta)e^{ik_0 x \sin\theta}. \tag{36}$$

C. Transmission loss

The transmitted sound intensity $I_t(\theta)$ can be expressed as

$$I_t(\theta) = \tfrac{1}{2}\mathrm{Re}\{p_t(x)v_t^*(x)\} = \tfrac{1}{2}\rho_0 c_0^3 k_0^6 \cos^3\theta|\alpha_t F(k_0 \sin\theta)|^2, \tag{37}$$

where the asterisk denotes the complex conjugate. The incident sound intensity $I_i(\theta)$ can be written as $I_i(\theta) = \cos\theta/(2\rho_0 c_0)$. Then the transmission loss of the structure $\tau(\theta)$ can be expressed as $\tau(\theta) = I_t(\theta)/I_i(\theta)$. The mean transmission loss $\overline{\mathrm{TL}}$ can be given by

$$\overline{\mathrm{TL}} = 10\log_{10}\frac{\int_0^{\theta_{\lim}}\sin 2\theta d\theta}{\int_0^{\theta_{\lim}}\tau(\theta)\sin 2\theta d\theta}, \tag{38}$$

where $\theta_{\lim}$ is 78° for field incidence or 90° for random incidence.

III. NUMERICAL EXAMPLES AND DISCUSSIONS

In the following, some numerical examples of the calculated results are shown in order to discuss the effects of an air-layer-subdivision technique and also clarify its limitations.

Figure 2 shows the mean transmission loss under field incidence of plane waves. Parameters of the core plate, made of glass, are as follows: thickness is 8 mm; density is 2500 kg/m^3; Young's modulus is 7×10^{10} N/m^2; Poisson's ratio is 0.22; and loss factor is 0.002. Motion-constraint condition is achieved by subdividing the air layer with partitions, the cell size of which should be sufficiently smaller than the wavelength at the target frequency where improvement of sound insulation performance is desired. Partitions of depth 80 mm are assumed to be installed near the glass plate with a space 10 mm. In order to investigate only the effect of motion constraint, these spaces are inserted to avoid stiffening the core plate by the partitions. The cases of radiation-side subdivision ($d_1=d_2=0$ mm, $d_3=10$ mm, $d_4=80$ mm, $k_3=k_0\cos\theta$, $k_4=k_0$, $k_3'=k_4'=k_0\sin\theta$) and both-side subdivision ($d_1=d_4=80$ mm, $d_2=d_3=10$ mm, $k_1=k_4=k_0$, $k_2=k_3=k_0\cos\theta$, $k_1'=k_2'=k_3'=k_4'=k_0\sin\theta$) are shown in Fig. 2. It is seen that sound insulation performance is improved remarkably near the critical frequency of coincidence 1500 Hz by constraining the particle motions to move in the direction normal to the plate. The improvements are greater for the case of both-side subdivision than the case of one-side subdivision.

As a parameteric study, effects due to the depth d_4 of the subdivided layer are shown in Fig. 3. Other parameters are the same as those of the preceding example ($d_1=d_2=0$ mm, $d_3=10$ mm, $k_3=k_0\cos\theta$, $k_4=k_0$, $k_3'=k_4'=k_0\sin\theta$) and the cases of $d_4=20$, 80, and 140 mm are predicted. It is seen that the transmission loss increases at low frequencies as the depth is larger. It is because the constraint condition can be steadily achieved when the depth is sufficiently larger than the target wavelength. However, z-directional resonance effects within the cell, which can be seen in the case of $d_4=140$ mm, shift to lower frequencies as the depth is larger. Thus, there is an optimal value of the depth in light of the maximal transmission loss at the target frequency. The depth of the subdivided layer should be determined so as to achieve the constraint condition and avoid the resonance effects at the target frequency.

Figure 4 shows the effects of space between the plate and the subdivided layer. Other parameters are the same as those of the preceding example ($d_1=d_2=0$ mm, $d_4=80$ mm, $k_3=k_0\cos\theta$, $k_4=k_0$, $k_3'=k_4'=k_0\sin\theta$) and the cases of $d_3=10$, 40, 80, and 160 mm are predicted. As described earlier, the grazing radiation near the coincidence and elliptical convection below the coincidence occur in the neighborhood of

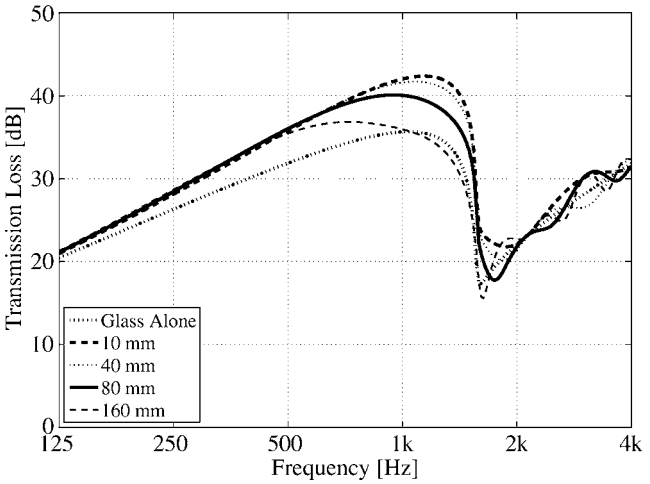


FIG. 4. Effects due to the depth of a space. The subdivided layer of thickness 80 mm is located at radiation side with the space. Resonance effects within the space occur at high frequencies. Because the resonances shift to lower frequency as the space becomes larger, the improvement obtained by motion constraint is significantly reduced at low and midfrequencies in the case of $d_3=160$ mm. The subdivided layer should be installed as near the plate as possible.

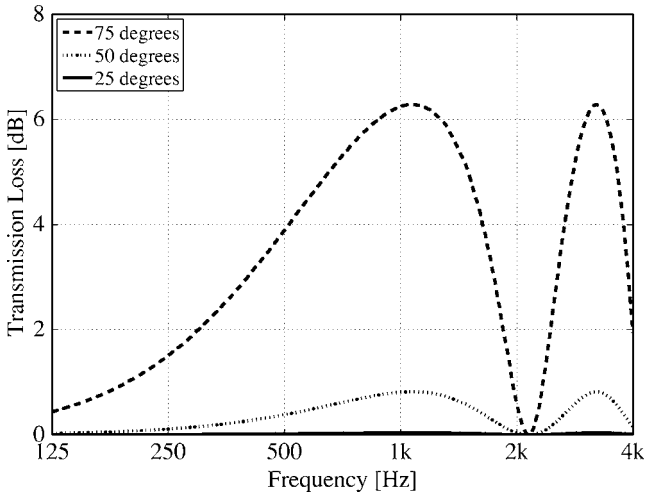


FIG. 5. Transmission loss of the subdivided layer itself. Results of the cases where incident angles are 25°, 50°, and 75° are shown, where the depth of the partitions is 80 mm. It can be seen that transmission loss become greater as the incident angle approaches the grazing angle.

the plate. It is seen from Fig. 4 that z-directional resonance effects within the space occur at high frequencies. Because the resonances shift to lower frequency as the space becomes larger, the improvement obtained by motion constraint is significantly reduced at low and midfrequencies in the case of $d_3=160$ mm. Transmission loss converges at the case of glass alone when the space is sufficiently large. Thus, the subdivided layer should be installed as near the plate as possible. Practically speaking, if the subdivided layer is tightly attached to the plate, there is no resonance within the above-described space, and furthermore, more improvement would be anticipated because the plate is stiffened by the partitions. However, in this case, the layer will also undergo uncertain factors like bending motions and lateral displacements, which might reduce the improvement. As for these points, it is necessary to do more detailed studies.

IV. MECHANISM

The following is devoted to a discussion on the mechanism of improving the sound insulation performance with a subdivided layer.

In order to clarify the role of a subdivided layer, transmission loss of the layer itself is calculated with the same assumptions considered in Sec. III. Figure 5 shows the results of the cases where incident angles are 25, 50, and 75°, where the depth of the partitions is 80 mm. It can be seen that transmission loss becomes greater as the incident angle approaches the grazing angle. As a matter of course, transmission loss is zero in the case of normal incidence. Dips seen around 2 and 4 kHz, where the depth of the partitions is equal to one-half and one wavelength, respectively, are caused by z-directional resonances inside the cell. From these characteristics, it can be said that the layer provides a kind of impedance boundary like the open-end reflection of a duct.

In Fig. 5, transmission loss at 3 kHz is similar to that at 1 kHz. However, as seen in Figs. 2–4, improvements due to the partitions are not obtained at high frequencies. This is caused by coincidence effects. At a frequency above the critical frequency of coincidence, the coincidence effect occurs at the incident angle $\theta_c=\sin^{-1}(c_0/c_B)$. At the frequency, the average transmission loss is affected dominantly by a plane wave at the incident angle θ_c. The angle θ_c approaches zero as the frequency shifts higher, while partitions have no effect in the case of normal incidence as discussed earlier. Therefore, improvements due to partitions cannot be obtained at high frequencies.

At the critical frequency of coincidence, the case of the grazing incidence is dominant. Partitions are most effective in such a case. Thus, the improvement can be obtained unless the z-directional resonance inside the cell is avoided at the frequency.

Below the critical frequency of coincidence, no coincidence effect occurs at any incident angles. In this case, the sound transmissions are generally serious near the grazing incident angle. However, as stated earlier, the subdivision technique has an advantage to insulate such plane waves near the grazing incident angle.

From these discussions, improvements are obtained near and below the critical frequency of coincidence.

V. EXPERIMENTAL STUDY

In order to validate the theory discussed in the preceding sections and examine the practical effect of the air-layer-subdivision technique, experimental results of transmission loss were obtained by using the reverberation chamber method. Five microphones were located in each chamber and a loud speaker is set up as a sound source. The size of the opening between the chambers is 0.9 m × 1.8 m. The glass board of thickness 8 mm and the acryl partition of depth 80 mm were installed at the opening as shown in Fig. 6. The cell size of the partition is 40 mm × 40 mm. In this case, the constraint condition is expected to be effective below

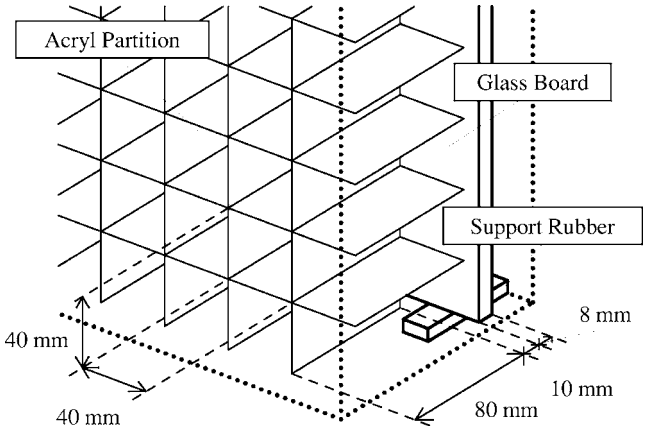


FIG. 6. Details of the glass board and the acryl partitions. The glass board and the partitions were separated with a space 10 mm to avoid stiffening the board. The glass board was supported by some small rubbers and spaces between the glass edge and the opening were filled up with clay so as to cut off the sound transmission through the spaces.

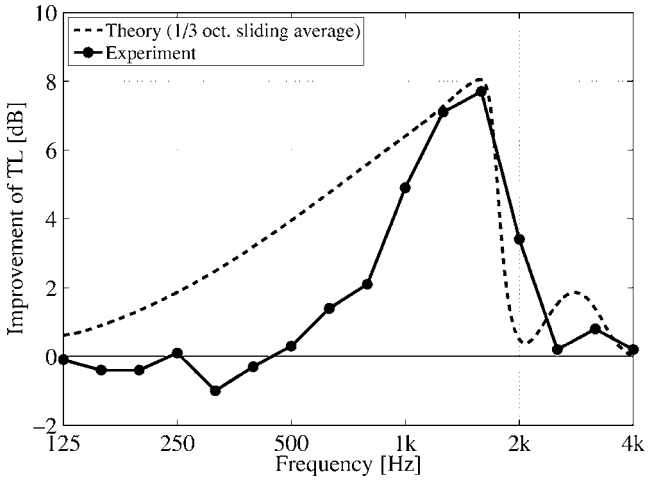


FIG. 8. Comparison of theoretical and experimental results from the viewpoint of the insulation improvement. The theoretical data are given as sliding frequency averages over a 1/3-octave band interval.

4250 Hz, where the one-half wavelength is equal to the cell size. The glass board and the partition were separated with a space 10 mm to avoid stiffening the board. The glass board was supported by some small rubbers and spaces between the glass edge and the opening were filled up with clay so as to cut off the sound transmission through the spaces. White noise is used for the sound source and spatially averaged sound pressure levels over each 1/3-octave band are measured from five micro-phones in each chamber. Transmission loss is calculated from the sound pressure levels with correction for background noises and reverberation times. The same measurements and procedures were carried out for the glass board alone in order to obtain the reference data.

Figure 7 shows the experimental results of transmission loss and Fig. 8 shows the improvement attributable to the air-layer-subdivision technique with theoretical ones, which are calculated from the results of transmission loss shown in Fig. 2. In order to compare the experimental results with the theoretical ones, the predicted data are given as sliding frequency averages over a 1/3-octave band interval in Fig. 8. In comparison with the theoretical results, the improvement cannot be observed below the critical frequency of coincidence. These disagreements would be caused by the mode vibrations of the finite glass board. An excited finite plate yields mode vibrations which are determined according to support conditions at edges of the plate. High-order-mode vibrations are dominant at high frequencies and low-order-mode vibrations are dominant at low frequencies. Because of these mode vibrations, the radiated sound does not form a plane wave, especially at low frequencies. Low-order mode vibrations significantly disturb the radiated plane-wave field and reduce the oblique-incidence components into the layer. However, a subdivided layer itself is effective only for oblique incidence as discussed in Sec. IV. Therefore, the improvement could not be obtained at low frequencies. However, near and above the coincidence, the experimental improvements are in good agreement with theoretical ones; the value of the improvement peak is about 8 dB.

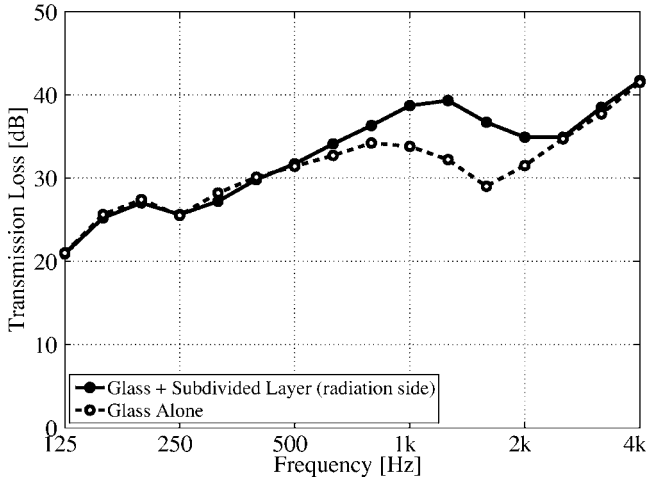


FIG. 7. Measured results of transmission loss. Although the improvement cannot be observed at low frequencies, the transmission loss significantly increases near the coincidence.

VI. CONCLUSION

In this paper, the effect of the air-layer-subdivision technique is studied in detail from the viewpoint of the sound transmission problem of a single plate. In order to estimate the effect, theoretical predictions are obtained by using an analytical model of an infinite single plate with a subdivided layer. Motion-constraint condition, which is achieved by subdividing the air layer, is implemented by assuming that the particle motion in the layer is constrained to move in the direction normal to the plate. The calculated results show the possibility of improving the sound insulation performance of a single plate especially near and below the critical frequency of coincidence. In order to clarify the limitations of air-layer-subdivision technique, some parametric studies are carried out. Consequently, it is shown that there is an optimal depth of the layer for a target frequency. In the experiment, transmission loss of a glass plate with the air layer subdivided by acryl partitions was measured by the reverberation chamber method. Experimental results near the coincidence, which were in good agreement with theoretical predictions,

validate the theory and the realization method of particle-motion-constraint condition. The wide use of the subdivision technique is anticipated because the improvement can be achieved without fibrous and porous materials, which are not desirable under severe environments.

ACKNOWLEDGMENTS

This study was supported by Program for Improvement of Research Environment for Young Researchers from Special Coordination Funds for Promoting Science and Technology (SCF) commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

[1]K. A. Mulholland, "The effect of sound-absorbing materials on the sound insulation of single panels," Appl. Acoust. **2**, 1–7 (1969).
[2]S. M. Brown, J. Niedzielski, and G. R. Spalding, "Effect of sound-absorptive facings on partition airborne-sound transmission loss," J. Acoust. Soc. Am. **63**, 1851–1856 (1978).
[3]D. Takahashi, "Sound transmission through single plates with absorptive facings," J. Acoust. Soc. Am. **83**, 1453–1457 (1988).
[4]M. Toyoda and D. Takahashi, "Reduction of acoustic radiation by impedance control with a perforated absorber system," J. Sound Vib. **286**, 601–614 (2005).
[5]M. Toyoda, M. Tanaka, and D. Takahashi, "Reduction of acoustic radiation by perforated board and honeycomb layer systems," Appl. Acoust. **68**, 71–85 (2006).
[6]M. Toyoda, M. Tanaka, and D. Takahashi, "Effects of air-layer subdivision: A new method of improving sound insulation," Build. Acoust. **13**, 49–59 (2006).
[7]K. M. Ivanov-Shits and F. V. Rozhin, "Investigation of surface waves in air," Sov. Phys. Acoust. **5**, 510–512 (1959).
[8]J. Tizianel, J. F. Allard, and B. Brouard, "Surface waves above honeycombs," J. Acoust. Soc. Am. **104**, 2525–2528 (1998).
[9]R. J. Donato, "Model experiments on surface waves," J. Acoust. Soc. Am. **63**, 700–703 (1978).
[10]B. Brouard, D. Lafarge, J. F. Allard, and M. Tamura, "Measurement and prediction of the reflection coefficient of porous layers at oblique incidence and for inhomogeneous waves," J. Acoust. Soc. Am. **99**, 100–107 (1996).
[11]J. Tizianel and J. F. Allard, "Experimental localization of a pole of the reflection coefficient of a porous layer," J. Sound Vib. **202**, 600–604 (1997).
[12]W. Lauriks, L. Kelders, and J. F. Allard, "Surface waves and leaky waves above a porous layer," Wave Motion **28**, 59–67 (1998).

Dispersion-invariant features for classification

Greg Okopal and Patrick J. Loughlin[a)]
Department of Electrical and Computer Engineering, University of Pittsburgh, 348 Benedum Hall, Pittsburgh, Pennsylvania 15261

Leon Cohen
Department of Physics, Hunter College and Graduate Center, City University of New York, New York, New York 10021

(Received 5 July 2007; revised 3 November 2007; accepted 5 November 2007)

In dispersive propagation, waves from the same source will generally differ depending on how far they have traveled. Accordingly, it is desirable for classification in such environments to either account for propagation effects, or to obtain features that are invariant to such effects. The latter approach is taken in this paper, and features are derived that are unaffected by channel dispersion, per mode. A "local" view of pulse propagation in time-frequency phase space is considered. It is shown that the local duration of a wave, obtained from its time-frequency Wigner distribution, is invariant to dispersion. While higher moments of the Wigner distribution are not invariant to dispersion, the phase space considerations suggest an approach for defining "dispersion-invariant moments" (DIMs) of any order. This approach is also used to define a dispersion-invariant correlation coefficient that can be used for classification. The classification utility of the DIMs, and of the dispersion-invariant correlation coefficient, is evaluated via simulations of acoustic scattering from steel shells in a dispersive channel model (Pekeris waveguide). Receiver operating characteristic curves quantify the improved discriminability of the DIMs versus ordinary temporal moments, and of the dispersion-invariant correlation coefficient versus the ordinary correlation coefficient. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821409]

PACS number(s): 43.60.Bf, 43.60.Hj [EJS] Pages: 832–841

I. INTRODUCTION

As a pulse propagates in a dispersive channel, such as an acoustic wave in a shallow water ocean channel, some of its basic properties will change. The fundamental reason for such a change is that in a dispersive environment, different frequencies propagate at different velocities.[1–3] The more broadband the wave, the more it will be dispersed. This effect can adversely affect detection and classification, because the observed wave depends upon the distance it has traveled. Hence, the same initial wave will not appear the same to different observers, and thus can be misclassified. It is therefore desirable for detection and classification in dispersive environments to use features that account for such propagation effects.

Two basic approaches can be taken: one is to use knowledge of the propagation environment to compensate for channel effects on the observed wave; a second is to obtain features from the wave that are invariant to channel effects. Certainly, if one has sufficient knowledge about the propagation environment to enable implementation of an accurate, realizable inverse channel model, then the former approach can be an effective means to eliminate propagation effects from an observed wave,[4,5] thereby improving classification. However, if knowledge of the channel characteristics is limited, inaccurate, or simply unavailable, the use of features that are unaffected by the channel, inasmuch as possible, becomes increasingly desirable for automatic classification. Furthermore, even when accurate inverse channel models can be used, the identification of such features is, in and of itself, of interest.

It is the aim of this paper to consider this latter approach. We give a class of temporal moment-like features that are invariant to dispersion, per mode. We examine via simulation the classification utility of these features, which we call "dispersion-invariant moments" (DIMs). We compare classification performance to that of ordinary temporal moments on numerical models of acoustic scattering from steel shells in a Pekeris waveguide. We evaluate the two-class problem of distinguishing a sphere from a cylinder, and the more challenging problem of distinguishing between two different cylinders, via their backscattered echoes at various propagation distances. The classification utility of our moment features is assessed by computing receiver operating characteristic (ROC) curves for each two-class problem. We also consider an approach based on a correlator receiver. We apply our dispersion-invariant approach to define a dispersion-invariant correlation coefficient for use as a classification feature, and show the detrimental effects of dispersion on the ordinary correlation coefficient. ROC curves show that the dispersion-invariant correlation coefficient provides performance gains over the ordinary correlation coefficient for the two-class simulations, achieving near-perfect classification for signal-to-noise ratios (SNRs) down to −5 dB.

[a)]Electronic mail: loughlin@engr.pitt.edu

II. BACKGROUND

A. Classification of underwater sounds

The variety of approaches to sonar classification that have been pursued is quite large. A common method for classification is the matched filter or correlator-receiver approach, in which an observed signal is correlated with a reference signal, and then the correlation coefficient is compared to a threshold.[6] Another method is to extract features from labeled training data, which are then used to build a partitioned feature space. Observed data are classified by computing the feature vector and choosing the label corresponding to the location of the vector in the feature space.

In one of the earlier papers on sonar classification, Hoffman[7] studied the problem in the time domain by comparing the quadrature components of a received echo to the quadrature components of reference echoes for known targets and computing a likelihood ratio. Chestnut *et al.*[8] and Chestnut and Floyd[9] studied the problem in the frequency domain using feature vectors derived from filterbanks and autoregressive modeling of the spectra of sonar echoes. The feature vector of a received echo was then compared to the feature vector of known targets via a distance metric.

Classification techniques in the joint time-frequency domain have also been explored. Altes[10] proposed a method for signal detection (or classification) using the spectrogram. In general, the method involves computing a statistic that is a function of the spectrograms of a reference signal and the received signal and comparing that statistic to a threshold. In cases where some parameters of the signal to be detected are random (e.g., delay and Doppler shift), the statistic incorporates probability distributions for each of these parameters. In the case of a known signal propagating in a noisy channel, the statistic is computed by correlating the spectrogram of the received signal with an ensemble average spectrogram of the output of the channel when the known signal is applied as input. In order to account for channel effects with this method, the scattering function of the channel must be known.

To account for bottom reverberation with no specific *a priori* knowledge of the channel, Chevret[11] proposed using a time-frequency filtering method. The Wigner distribution of the free-field response of targets of interest is used to highlight areas of interest in the time-frequency plane. These areas correspond to the regions that will be passed by the filter. To detect or classify signals, the filter representing each target of interest slides along the received signal in the time dimension. The output of the filter is then compared to the reference free-field response using an error measure. The minimum error indicates detection of the target corresponding to that filter. The possible misclassification of reverberation echoes is handled by a signal-to-reverberation ratio; possible detections of signal components whose amplitudes are less than this ratio are ignored.

In this paper, we focus not on the classifier, but on the feature extraction process to obtain features that are unaffected by dispersion. We also consider a "dispersion-invariant" correlation coefficient approach. It is not our aim to develop the "best" classifier or features. Whether or not the features we derive are useful for classification will, of course, depend on the particular problem at hand, just as sometimes objects can be distinguished by their temporal or spectral moments, and sometimes they cannot. Our approach allows for the extraction of other dispersion-invariant features besides the temporal moments we consider here.

In Sec. II B, we briefly review the normal mode solution to linear wave propagation, and then we proceed with our feature extraction processing to obtain dispersion-invariant moments and a dispersion-invariant correlation coefficient.

B. Linear wave propagation

For simplicity, we consider a single spatial dimension, and denote the wave at range x and time t by $u(x,t)$. Given the initial wave, $u(0,t)$, the wave at a different position is then given by[2,3]

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int F(0,\omega) e^{jk(\omega)x} e^{-j\omega t} d\omega \tag{1}$$

per mode, where $F(0,\omega)$ is the Fourier spectrum of the initial wave,

$$F(0,\omega) = \frac{1}{\sqrt{2\pi}} \int u(0,t) e^{j\omega t} dt \tag{2}$$

and $k(\omega)$ is the dispersion relation coupling radial (ω) and spatial (k) frequencies. The spectrum of the wave at position x is given by

$$F(x,\omega) = \frac{1}{\sqrt{2\pi}} \int u(x,t) e^{j\omega t} dt \tag{3}$$

and therefore Eq. (1) can be written as

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int F(x,\omega) e^{-j\omega t} d\omega, \tag{4}$$

where

$$F(x,\omega) = F(0,\omega) e^{jk(\omega)x}. \tag{5}$$

Thus, by Eq. (5), we see that dispersive propagation can be viewed as a filtering of the wave by a linear, time-invariant but spatially dependent filter with frequency response $H(x,\omega) = e^{jk(\omega)x}$. If there is no dispersion, then $k(\omega) = \omega/c$ and the wave propagates unchanged, with velocity c, and is given by $u(x,t) = u(0, t \pm x/c)$.

If we express the spectrum in terms of amplitude and phase,

$$F(x,\omega) = B(x,\omega) e^{j\psi(x,\omega)}, \tag{6}$$

then Eq. (5) becomes

$$F(x,\omega) = B(x,\omega) e^{j\psi(x,\omega)} = B(0,\omega) e^{j(\psi(0,\omega)+k(\omega)x)} \tag{7}$$

by which it follows that

$$B(x,\omega) = B(0,\omega), \tag{8}$$

$$\psi(x,\omega) = \psi(0,\omega) + k(\omega)x, \tag{9}$$

for real dispersion relations, which is the case we consider here.

III. MOMENTS

A. Global moments

Moments quantify characteristics of a signal and have been used for classification.[12–15] We distinguish between global moments, which are moments obtained over the duration of the signal, such as the mean, variance, skew, etc., and "local" moments, which we consider in Sec. III B and are analogous to conditional moments of a distribution.

Spectral moments, such as bandwidth (square-root of the spectral variance), can be used for classification, and indeed they are invariant to dispersion, per mode, for real dispersion relations. Specifically, the nth spectral moment is given by

$$\langle \omega^n \rangle_x = \int \omega^n |F(x,\omega)|^2 d\omega. \tag{10}$$

For real dispersion relations, it follows from Sec. II B that $|F(x,\omega)| = |F(0,\omega)|$ and therefore spectral moments are unaffected by dispersion,

$$\langle \omega^n \rangle_x = \langle \omega^n \rangle_0 \tag{11}$$

per mode.

However, as shown next, this invariance to dispersion does not extend to temporal moments. Our aim is to present a process for obtaining temporal moment features that are invariant to dispersion and hence are another possible set of features to complement the spectral moments for classification of signals in dispersive environments.

The temporal moments of a wave at location x are defined by

$$\langle t^n \rangle_x = \int t^n |u(x,t)|^2 dt \tag{12}$$

and the central temporal moments are given by

$$\mu_x^n = \int (t - \langle t \rangle_x)^n |u(x,t)|^2 dt. \tag{13}$$

In a dispersive environment, the temporal moments, like the wave, will change with propagation distance x. Consider, for example, the duration of the wave, obtained from the second central temporal moment,

$$\sigma_{t|x}^2 = \int (t - \langle t \rangle_x)^2 |u(x,t)|^2 dt. \tag{14}$$

It can be shown that the duration is given exactly by[16]

$$\sigma_{t|x}^2 = \sigma_{t|0}^2 + 2x\,\mathrm{Cov}_{t\tau|0} + x^2\sigma_\tau^2, \tag{15}$$

where $\sigma_{t|0}^2$ is the duration of the initial wave $u(0,t)$, and $\mathrm{Cov}_{t\tau|0}$ and $\sigma_\tau^2 \geqslant 0$ are quantities that depend on the initial wave and the dispersion relation (if there is no dispersion, they are zero).[16] The important point here is that indeed when there is dispersion, the duration changes with propagation x, and ultimately it will increase because the x^2 will dominate (although depending on the sign of the covariance term, the pulse may contract before eventually expanding). If there is no dispersion, then $\sigma_{t|x}^2 = \sigma_{t|0}^2$ and hence the duration could be a useful classification feature in a dispersionless environment.

Because dispersion is a frequency-dependent effect, it is illuminating to consider dispersive propagation in time-frequency (or position-wavenumber) space. This leads us to considerations of "local moments" and a procedure for obtaining temporal moment features that are invariant to dispersion.

B. Local moments

Recently, a phase-space approximation of dispersive propagation was developed in terms of the position-wavenumber or time-frequency Wigner distributions of the wave.[17,18] In particular, if $W(t,\omega;0)$ denotes the time-frequency Wigner distribution of the initial wave $u(0,t)$, then the Wigner distribution of the wave $u(x,t)$ at some other distance x is approximately given by

$$W(t,\omega;x) \approx W(t + k'(\omega)x, \omega; 0) \tag{16}$$

per mode, for real dispersion relation $k(\omega)$, where the prime ($'$) denotes differentiation. This view of dispersive propagation in terms of the Wigner distribution suggests an approach to obtaining features that are invariant to dispersion. In particular, we observe that in time-frequency phase space, the wave undergoes a frequency-dependent time shift, which depends on the dispersion relation. This observation leads us to consider the possibility that *local* central moments—that is, moments that are centered about the frequency-dependent time shift—could serve as dispersion-invariant features. To see this, consider the "local duration," that is, the standard deviation in time at each frequency, of the Wigner distribution, defined by

$$\sigma_{t|x,\omega}^2 = \langle (t - \langle t \rangle_{x,\omega})^2 \rangle_{x,\omega} = \frac{1}{\int W(t,\omega;x)dt} \int (t - \langle t \rangle_{x,\omega})^2 W(t,\omega;x) dt, \tag{17}$$

where

$$\langle t \rangle_{x,\omega} = \frac{1}{\int W(t,\omega;x)dt} \int t W(t,\omega;x) dt \tag{18}$$

is the "local mean time." For real dispersion relations $k(\omega)$, these local moments are[19]

$$\langle t \rangle_{x,\omega} = -\psi'(0,\omega) - k'(\omega)x, \tag{19}$$

$$\sigma_{t|x,\omega}^2 = \tfrac{1}{2}[B'^2(0,\omega) - B''(0,\omega)B(0,\omega)]/B^2(0,\omega). \tag{20}$$

Importantly, note that, although the global duration of the wave, Eq. (15), changes with propagation distance x, the local duration of each mode does not; that is

$$\sigma^2_{t|x,\omega} = \sigma^2_{t|0,\omega}. \tag{21}$$

Higher-order local central moments of the Wigner distribution are not invariant to dispersion, but are approximately so, which can be shown by making use of the Wigner approximation, Eq. (16). In particular,

$$\langle (t - \langle t \rangle_{x,\omega})^n \rangle_{x,\omega}$$

$$= \frac{1}{\int W(t,\omega;x)dt} \int (t - \langle t \rangle_{x,\omega})^n W(t,\omega;x)dt \tag{22}$$

$$\approx \frac{1}{\int W(t,\omega;x)dt} \int (t - \langle t \rangle_{x,\omega})^n W(t + k'(\omega)x,\omega;0)dt \tag{23}$$

$$= \frac{1}{B^2(x,\omega)} \int (t + \psi'(0,\omega) + k'(\omega)x)^n \times W(t + k'(\omega)x,\omega;0)dt \tag{24}$$

$$= \frac{1}{B^2(0,\omega)} \int (\tau + \psi'(0,\omega))^n W(\tau,\omega;0)d\tau \tag{25}$$

$$= \langle (t - \langle t \rangle_{0,\omega})^n \rangle_{0,\omega}. \tag{26}$$

Although the higher-order local central moments of the Wigner distribution are only approximately invariant to dispersion, it is in fact possible to define central moments that are exactly invariant to dispersion, as we do next.

IV. DISPERSION-INVARIANT MOMENTS

The local moments considered previously suggest an approach to obtain temporal central moments that are invariant to dispersion, for any order moment. In particular, it is suggestive to consider temporal central moments that are centered about the local mean time $\langle t \rangle_{x,\omega}$, rather than the mean time $\langle t \rangle_x$. To see how this can be done, consider again the ordinary central temporal moments, Eq. (13). Because $u(x,t)$ and $F(x,\omega)$ are a Fourier transform pair, these moments may be equivalently formulated as[20]

$$\mu^n_x = \int F^*(x,\omega)\left(j\frac{\partial}{\partial\omega} - \langle t \rangle_x\right)^n F(x,\omega)d\omega. \tag{27}$$

Accordingly, we define local moments that are centered about the local mean time by

$$A_n(x) = \int F^*(x,\omega)\left(j\frac{\partial}{\partial\omega} - \langle t \rangle_{x,\omega}\right)^n F(x,\omega)d\omega, \tag{28}$$

where $\langle t \rangle_{x,\omega}$ is given by Eq. (19), which also happens to be the group delay, $-\psi'(x,\omega)$, of the wave. By centering these moments about the local mean time, they have the property that they do not change with propagation distance x in a dispersive environment. To see this, consider that[21,22]

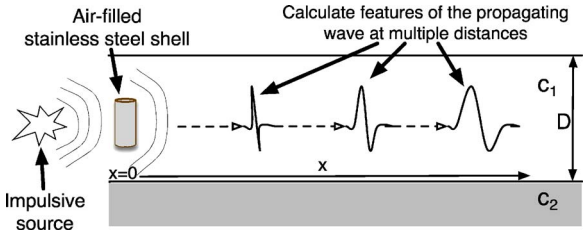


FIG. 1. (Color online) Simulation setup.

$$\left(j\frac{\partial}{\partial\omega} + \psi'(x,\omega)\right)^n F(x,\omega) = j^n B^{(n)}(x,\omega)e^{j\psi(x,\omega)}, \tag{29}$$

where superscript (n) denotes the nth derivative. Substituting this result into the definition of our local central temporal moments gives

$$A_n(x) = \int B(x,\omega)e^{-j\psi(x,\omega)} j^n B^{(n)}(x,\omega)e^{j\psi(x,\omega)}d\omega$$

$$= \int B(0,\omega)j^n B^{(n)}(0,\omega)d\omega, \tag{30}$$

where the final equation follows for real dispersion relations. Thus, $A_n(x) = A_n(0)$, and indeed these moments are invariant to dispersion, per mode.

These DIMs may be equivalently computed in the time domain, which is preferable as it avoids the direct computation of high-order derivatives of $F(x,\omega)$ in Eq. (28). Making use of Fourier relations, we find that the time-domain formulation is given by[19]

$$A_n(x) = \frac{1}{2\pi}\int\int t^n |F(x,\omega)|e^{-j\omega t}d\omega dt. \tag{31}$$

Note that for real signals, the integral over ω produces a symmetric signal—namely, a signal that is real and even—and hence this formulation makes clear that the odd-order DIMs are identically zero. Because this offers no classification utility, we calculate absolute value moments for odd n by letting $t \to |t|$ in Eq. (31).

We emphasize that the dispersion invariance of the DIMs (as well as of spectral moments) is per mode. The reason for this is that each mode has its own group velocity. If there are multiple propagating waveguide modes, then mode separation should be used at the receiver (if possible) to achieve invariance. In our simulations presented next, we consider the first propagating mode of the (Pekeris) waveguide. In a future paper, we will consider the effect of mul-

TABLE I. Geometry of steel shells used in the sphere vs cylinder simulation.

	Inner radius (m)	Outer radius (m)
Sphere	0.90	1.00
Cylinder	1.43	1.45

TABLE II. Geometry of steel shells used in the cylinder vs cylinder simulation.

	Inner radius (m)	Outer radius (m)
Cylinder 1	1.10	1.20
Cylinder 2	1.20	1.25

tiple waveguide modes, but note that preliminary results indicate that although performance of the DIMs degrades when there are multiple modes, they perform better than ordinary moments.

V. CLASSIFICATION SIMULATIONS

A. Simulation methods

We conducted classification simulations to compare the performance of the dispersion-invariant moments to that of traditional moments that have been used for classification. The simulations were conducted by numerically computing the sonar backscatter from air-filled steel shells and the subsequent propagation of the wave through a simulated dispersive shallow-water channel (Fig. 1). We considered the two-class problem of distinguishing a sphere from a cylinder and the more challenging problem of distinguishing between two different cylinders via their backscattered echoes at various propagation distances. The dimensions of the shells for the sphere versus cylinder case are given in Table I, and the dimensions of the shells for the cylinder versus cylinder case are given in Table II.

The shells were insonified by an impulsive source, and the backscattered pressure obtained by resonance scattering theory (RST)[23–27] was used as the initial wave $u(0,t)$ in the channel at position $x=0$. This initial backscattered wave was then propagated to several distances in a dispersive channel model (Pekeris waveguide). The depth of the channel (D) was fixed at 25 m. The densities and sound speeds of all materials are given in Table III. Gaussian white noise was added to the received signal to obtain a SNR of 20 dB, where SNR is defined as the ratio of the power of the propagated noise-free echo to the power of the white Gaussian noise added to the signal.

The classification features computed from each simulated propagated echo were ordinary central temporal moments given in Eq. (13) for $n=2,3,4$, and the corresponding DIMs given in Eq. (31). As mentioned previously, for real waves $u(x,t)$, odd-order DIMs are identically zero, which offers no classification utility. Accordingly, we compute absolute value DIMs by letting $t\rightarrow|t|$ in Eq. (31) for $n=3$. While ordinary central odd-order temporal moments are not identically zero, for completeness in comparisons, we computed ordinary central temporal moments for $n=2,3,4$ as well as absolute value third-order central temporal moments, given by

$$v_x^3 = \int |t-\langle t\rangle_x|^3 |u(x,t)|^2 dt. \tag{32}$$

All moments were normalized by the respective zero-order moment so that differences in signal energy between the classes would not contribute to classification performance. Specifically, the DIMs [Eq. (31)] were normalized by com-

TABLE III. Material properties.

	Density (kg/m^3)	Sound speed (m/s)
Water layer	1000	1500
Sediment layer	1800	1800
Air	1.2	340
Steel	7800	5880 (dilatational) 3140 (shear)

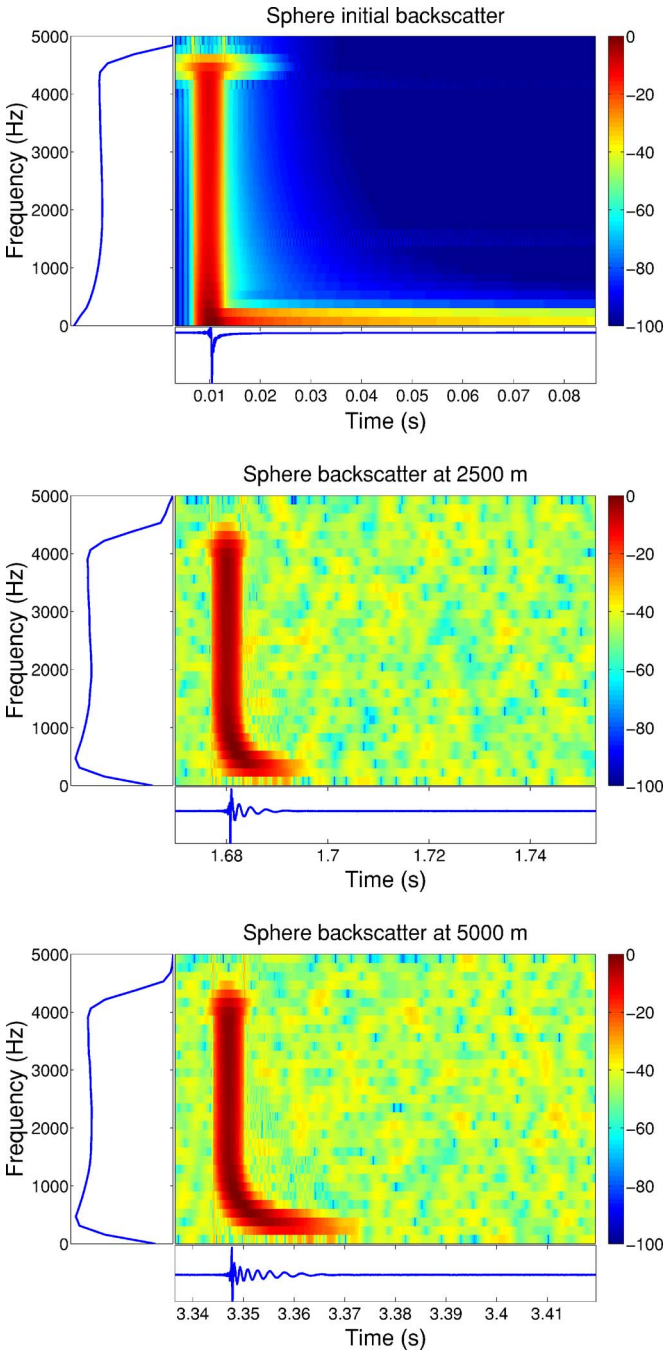


FIG. 2. (Color online) Sphere vs cylinder: Waveforms generated by the sphere (initial, 2.5 km, and 5 km).

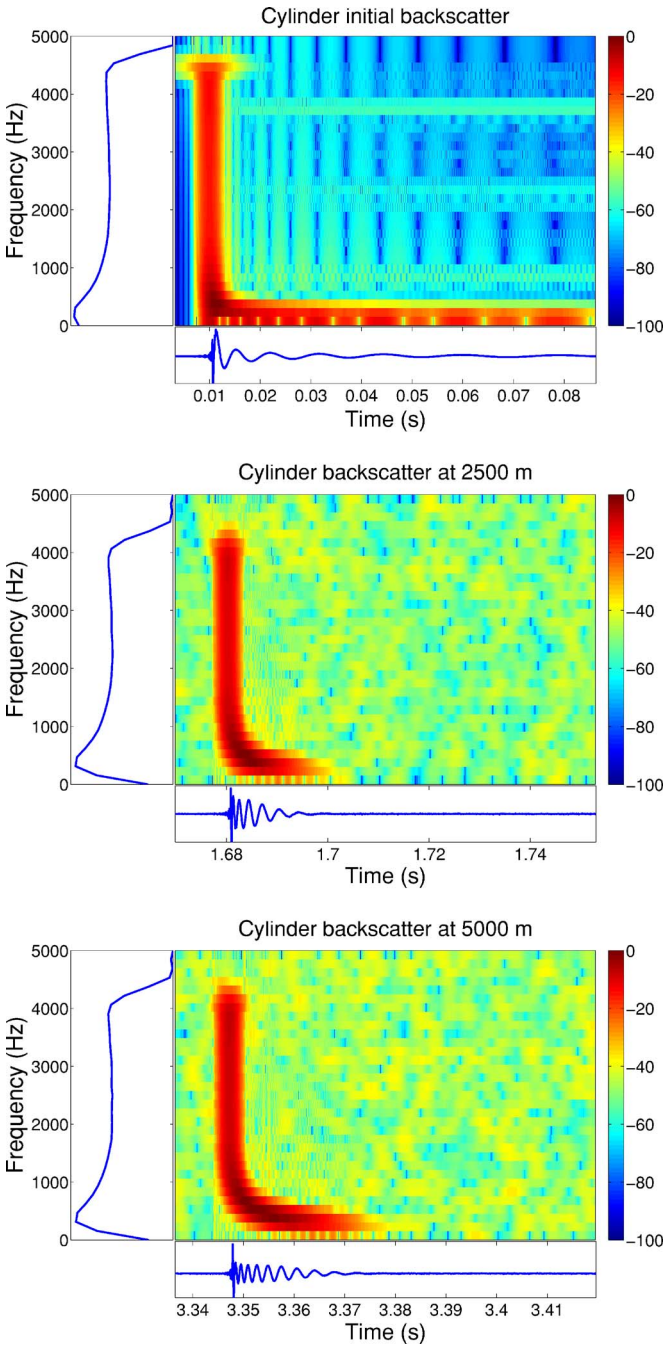


FIG. 3. (Color online) Sphere vs cylinder: Waveforms generated by the cylinder (initial, 2.5 km, and 5 km).

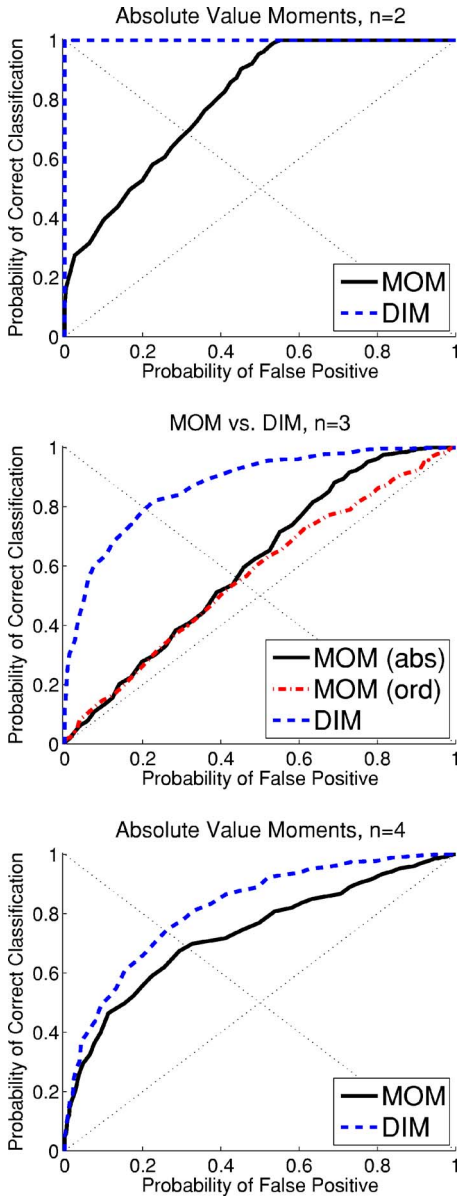


FIG. 4. (Color online) Sphere vs cylinder: Classification results for MOMs and DIMs, $n=2,3,4$. The plot for $n=3$ shows results for both the ordinary third-order central temporal moment [MOM (ord)] and the absolute value third-order central temporal moment [MOM (abs)].

puting $A_n(x)/A_0(x)$, and the ordinary moments [Eq. (13)] were normalized by computing μ_x^n/μ_x^0.

In order to compare the classification performance of the features, ROC curves were generated. The ROC curve offers a concise representation of the class separability provided by a single feature on the data set. These curves were generated as follows. For each feature, normalized histograms of the feature values for all target echoes and, separately, all clutter echoes provide approximate probability density functions (pdf) for a given feature on each class. To generate the ROC curve for a given feature, a decision threshold is swept across these pdfs, and at each threshold value, the probability of correct classification and the probability of false alarm are calculated and plotted.

Feature values were calculated at 5 m increments, starting at $x=5$, for the first propagating mode in the waveguide ($m=1$); therefore, mode separation is implicit in our simulations. The maximum propagation distance was 5 km. These simulations could represent the situation where the target range is unknown and uniformly distributed between 5 m and 5 km. The sampling frequency was fixed at 10 kHz, and an antialiasing filter was applied to limit the bandwidth of the initial backscattered waves to 5 kHz.

B. Simulation results

For the sphere versus cylinder case, the initial backscatter $u(0,t)$ computed by RST for each shell is shown in the

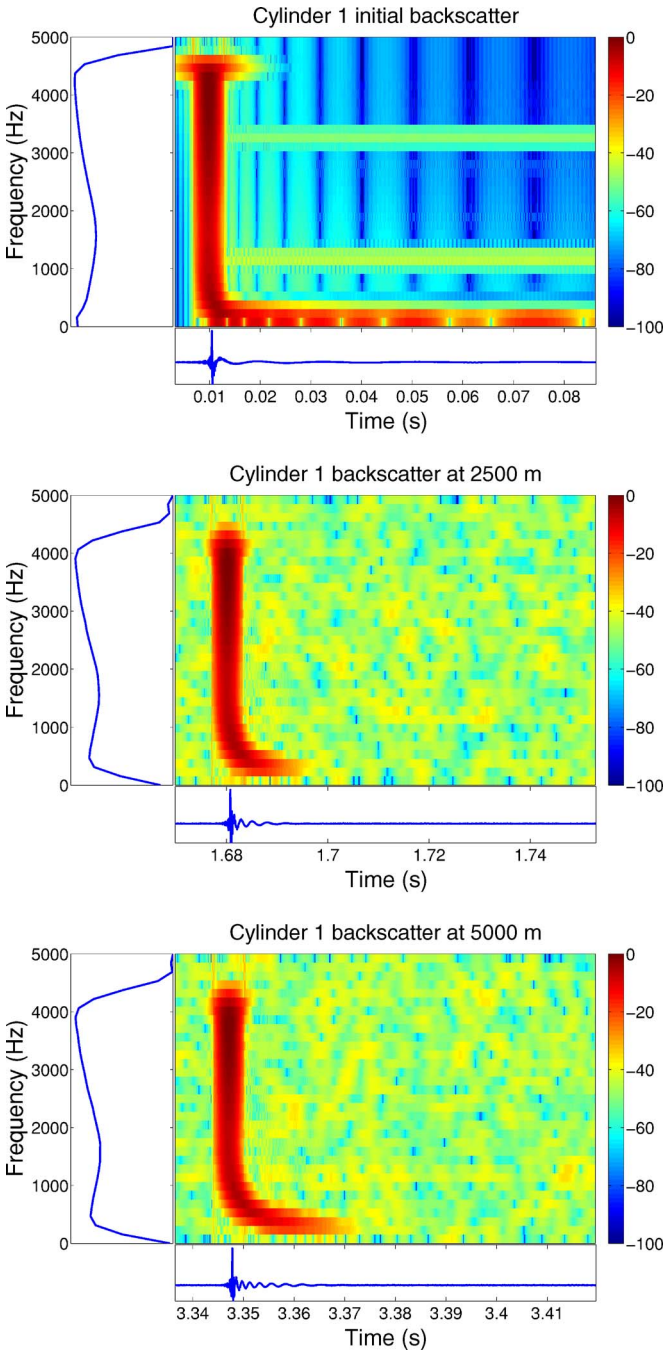


FIG. 5. (Color online) Cylinder vs cylinder: Waveforms generated by cylinder 1 (initial, 2.5 km, and 5 km).

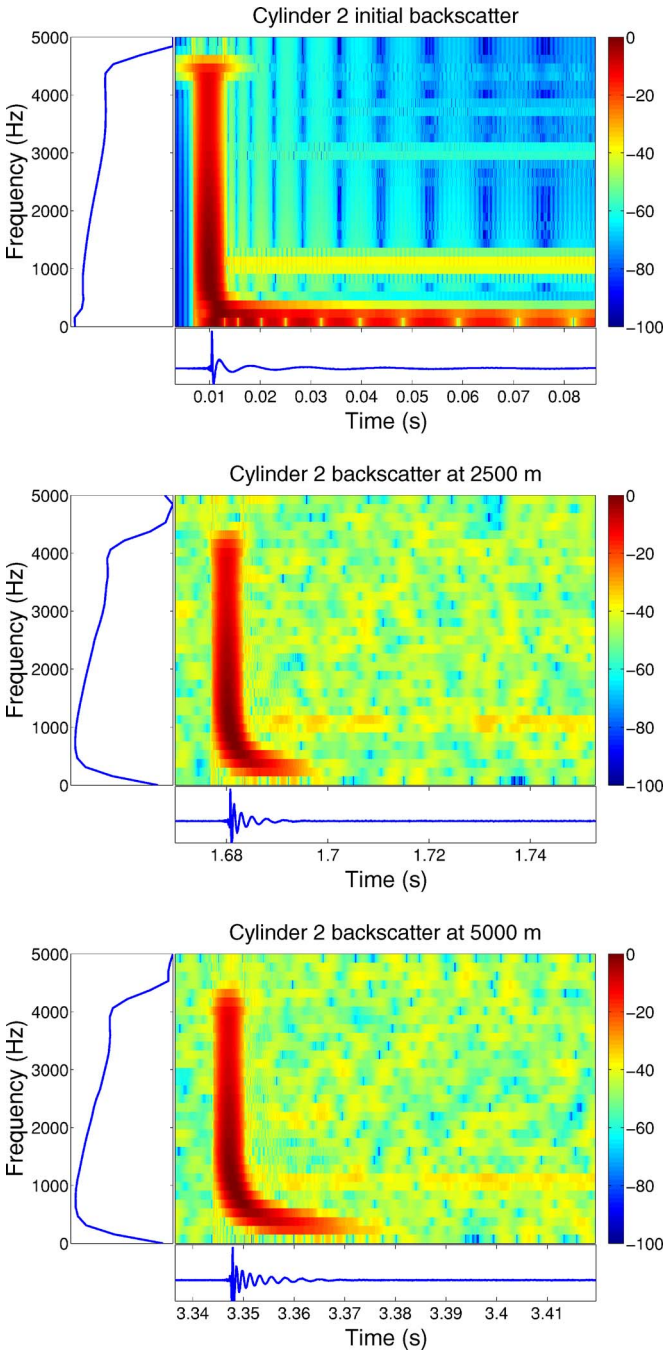


FIG. 6. (Color online) Cylinder vs cylinder: Waveforms generated by cylinder 2 (initial, 2.5 km, and 5 km).

first panel of Figs. 2 and 3. As the waves propagate through the dispersive Pekeris waveguide, the energy at low frequencies is delayed relative to higher frequencies.

ROC curves showing the separability of the distributions of the moment features indicate that, as anticipated from our theoretical considerations in the previous sections, the dispersion-invariant temporal moments (DIM) are more useful for classification (Fig. 4) than the ordinary temporal moments (MOM).

For the cylinder versus cylinder case, the initial backscatter $u(0,t)$ computed by RST for each shell is shown in the first panel of Figs. 5 and 6. The ROC curves given in Fig. 7 indicate that the dispersion-invariant temporal moments (DIM) are more useful for classification than the ordinary temporal moments (MOM).

VI. DISPERSION-INVARIANT CORRELATION COEFFICIENT

Our approach may also be applied to define a dispersion-invariant correlation coefficient for use as a feature in the two-class classification problem. First, however, we examine the effects of dispersion on an ordinary correlation coefficient used to classify the wave. Letting $u(x,t)$ denote the received wave and $u_i(0,t)$ the backscatter for shell i, the correlation coefficient is given by

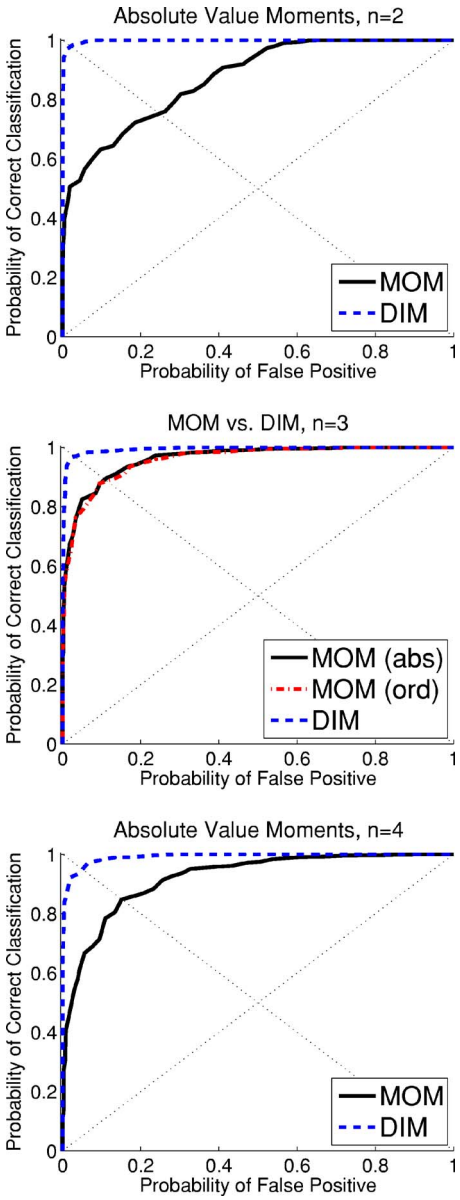


FIG. 7. (Color online) Cylinder vs cylinder: Classification results for MOMs and DIMs, $n=2,3,4$. The plot for $n=3$ shows results for both the ordinary third-order central temporal moment [MOM (ord)] and the absolute value third-order central temporal moment [MOM (abs)].

$$r = \frac{\int u^*(x,t)u_i(0,t)dt}{\int |u_i(0,t)|^2 dt}, \tag{33}$$

where integration is over the duration of the wave, and we normalize by the energy of $u_i(0,t)$ so that in the absence of noise, the correlation coefficient is $r=1$ when $u(x,t)=u_i(0,t)$. Hence, for an ensemble of noisy realizations, we will obtain a distribution of values of r, the mean of which will be 1 when the correct class is matched and dispersion is negligible.

We wish to explore the effects of dispersion on the correlation coefficient. To do this, we give the equivalent formulation in the frequency domain, which makes clear the detrimental effects of dispersive propagation on the correlation coefficient classification performance,

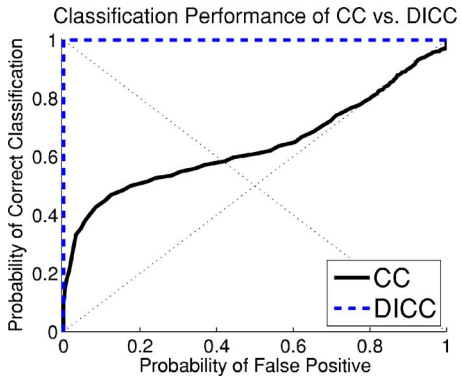


FIG. 8. (Color online) Sphere vs cylinder: DICC performance vs CC performance, no noise added.

$$r = \frac{\int u^*(x,t)u_i(0,t)dt}{\int |u_i(0,t)|^2 dt} = \frac{\int F^*(x,\omega)F_i(0,\omega)d\omega}{\int |F_i^*(0,\omega)|^2 d\omega} \tag{34}$$

$$= \frac{\int B(0,\omega)B_i(0,\omega)e^{-j(k(\omega)x+\psi(0,\omega)-\psi_i(0,\omega))}d\omega}{\int B_i^2(0,\omega)d\omega} \tag{35}$$

where we have made use of Eq. (7). When the correct class is matched, the output is

$$r = \frac{\int B^2(0,\omega)e^{-jk(\omega)x}d\omega}{\int B^2(0,\omega)d\omega}, \tag{36}$$

which clearly depends on the dispersion relation and changes with propagation distance x.

To eliminate the effects of dispersion, we apply the dispersion-invariant approach used to obtain the DIMs by implementing the correlation coefficient as

$$r = \frac{\int |F^*(x,\omega)||F_i(0,\omega)|d\omega}{\int |F_i(0,\omega)|^2 d\omega} = \frac{\int B(0,\omega)B_i(0,\omega)d\omega}{\int B_i^2(0,\omega)d\omega}, \tag{37}$$

which is independent of propagation distance x and in the absence of noise gives

$$r = \frac{\int B^2(0,\omega)d\omega}{\int B^2(0,\omega)d\omega} = 1 \tag{38}$$

when the correct class is matched.

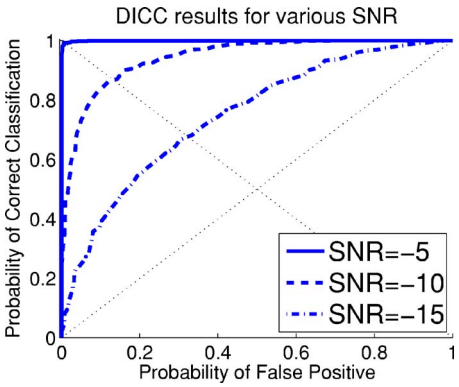


FIG. 9. (Color online) Sphere vs cylinder: DICC performance for various SNR values.

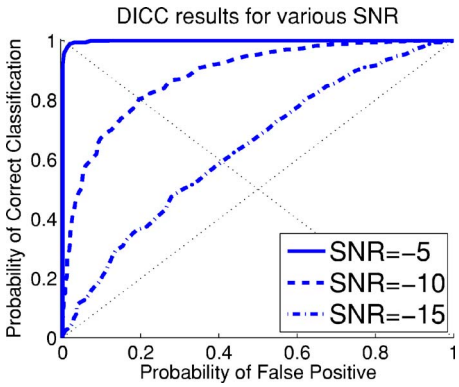


FIG. 11. (Color online) Cylinder vs cylinder: DICC performance for various SNR values.

A. Simulations

We conducted simulations to compare the performance of the dispersion-invariant correlation coefficient (DICC) to the performance of an ordinary correlation coefficient (CC) in a dispersive propagation channel. The simulation setup was identical to the simulations given in Sec. V B. For each trial, each correlation coefficient r (dispersion-invariant and ordinary) was computed. The values of r computed at each propagation distance lead to a distribution of feature values per echo class (sphere versus cylinder, or cylinder 1 versus cylinder 2). We investigated SNR levels as low as −15 dB. Classification utility of the correlation coefficients was evaluated by computing ROC curves from histograms of the feature values.

For the sphere versus cylinder case, we chose the reference signal $u_i(0,t)$ to be the initial backscatter from the cylinder. Thus, received echoes originating from the cylinder should be correctly classified as "target" echoes, while the echoes originating from the sphere should be classified as "clutter" echoes. For the cylinder versus cylinder case, the reference signal $u_i(0,t)$ was the initial backscatter from cylinder 2.

B. Results

The results of one simulation of the sphere versus cylinder are given in Fig. 8. For this trial, no noise was added to the propagated echoes. The DICC achieves perfect classification while the CC shows little classification utility.

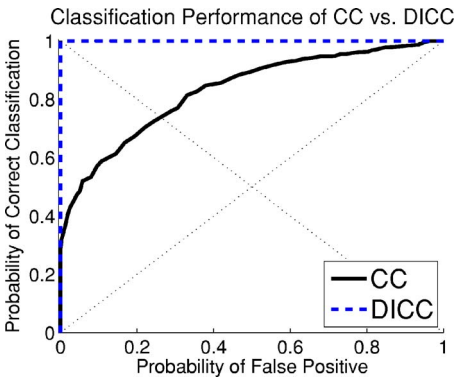


FIG. 10. (Color online) Cylinder vs cylinder: DICC performance vs CC performance, no noise added.

The performance of the DICC was also examined in a noisy environment. These trials are identical to the above-mentioned trial, with the exception that white Gaussian noise is added to each received echo $u(x,t)$ at each propagation distance. The results for three trials of the sphere versus cylinder are shown in Fig. 9. For SNR values as low as −5 dB, the DICC still achieves perfect classification.

For the cylinder versus cylinder case, the results of a noise-free trial are given in Fig. 10. The results of three noisy trials are given in Fig. 11. Similar to the sphere versus cylinder case, the DICC achieves near-perfect classification for SNR values as low as −5 dB.

VII. CONCLUSION

We defined a new class of temporal moment features that are invariant to dispersion, per mode, which we call dispersion-invariant moments (DIMs). We also showed how our dispersion-invariant approach can be applied to implement a dispersion-invariant correlation coefficient, similar to a matched filter, for use as a classification feature. Our moment features and correlation coefficient were evaluated via numerical simulations to distinguish between different steel shells in a Pekeris waveguide. The simulations demonstrated improved discriminability of the DIMs versus ordinary temporal moments and of the dispersion-invariant correlation coefficient versus an ordinary correlation coefficient.

For classification problems where ordinary temporal moments do not serve as good classification features on the initial echoes, the dispersion-invariant temporal moments will also perform poorly. For such problems, other dispersion-invariant features can be defined using our approach. The strength of the dispersion-invariant approach is the ability to preserve class separation of features independent of propagation. Areas of future work include extension of the approach to multimode propagation, and frequency-dependent attenuation.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research (Grant No. N00014-06-1-0009 (PL)), and the Air Force Office of Scientific Research (LC).

[1]J. Lighthill, *Waves in Fluids* (Cambridge University Press, New York, 1978), Chap. 3, Sec. 3.7.
[2]I. Tolstoy and C. Clay, *Ocean Acoustics: Theory and Experiment in Underwater Sound* (AIP, New York, 1987), Chaps. 2 and 4.
[3]G. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974), Chap. 11, Sec. 11.2.
[4]A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 402–424 (1993).
[5]J. M. Hovem, "Deconvolution for removing the effects of the bubble pulses of explosive charges," J. Acoust. Soc. Am. **47**, 281–284 (1970).
[6]H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 1968).
[7]J. F. Hoffman, "Classification of spherical targets using likelihood and quadrature components," J. Acoust. Soc. Am. **49**, 23–30 (1971).
[8]P. C. Chestnut, H. Landsman, and R. W. Floyd, "A sonar target recognition experiment," J. Acoust. Soc. Am. **66**, 140–147 (1979).
[9]P. C. Chestnut and R. W. Floyd, "An aspect-independent sonar target recognition method," J. Acoust. Soc. Am. **70**, 727–734 (1981).
[10]R. A. Altes, "Detection, estimation, and classification with spectrograms," J. Acoust. Soc. Am. **67**, 1232–1246 (1980).
[11]P. Chevret, N. Gache, and V. Zimpfer, "Time-frequency filters for target classification," J. Acoust. Soc. Am. **106**, 1829–1837 (1999).
[12]C. Chen, "Automatic recognition of underwater transient signals—A review," IEEE Proc. International Conference on Acoustics, Speech and Signal Processing'85, Tampa, FL, Vol. **10**, 1985.
[13]T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, A. Linney, "Classification of audio signals using statistical features on time and wavelet transform domains," IEEE Proc. International Conference on Acoustics, Speech and Signal Processing '98, Seattle, WA, Vol. **6**, May 1998.
[14]S. S. Soliman and S. Z. Hsue, "Signal classification using statistical moments," IEEE Trans. Commun. **40**, 908-916 (1992).
[15]B. Tacer and P. J. Loughlin, "Non-stationary signal classification using the joint moments of time-frequency distributions," Pattern Recogn. **31**, 1635–1641 (1998).
[16]L. Cohen, "Pulse propagation in dispersive media," Proceedings of the IEEE Statistical Signal and Array Processing Conference, Pocono Manor, PA, pp. 485–489, 2000.
[17]P. Loughlin, "Wigner distribution approximation for filtered signals and waves," J. Mod. Opt. **53**, 2387–2397 (2006).
[18]P. Loughlin and L. Cohen, "A Wigner approximation method for wave propagation," J. Acoust. Soc. Am. **118**, 1268–1271 (2005).
[19]P. Loughlin and L. Cohen, "Moment features invariant to dispersion," Proceedings of the SPIE Defense & Security Symposium, Automatic Target Recognition XIV, 2004, Vol. **5426**, pp. 234–246.
[20]L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[21]K. Davidson, "Instantaneous moments of a signal," Ph.D. dissertation, University of Pittsburgh, Pittsburgh, PA, 2000.
[22]P. Loughlin and K. Davidson, "Instantaneous spectral skew and kurtosis," Proceedings of the IEEE Statistical Signal and Array Processing Conference, Pocono Manor, PA, 2000, pp. 574–578.
[23]L. Flax and L. Dragonette, "Theory of elastic resonance excitation by sound scattering," J. Acoust. Soc. Am. **63**, 723-731 (1978).
[24]G. Gaunaurd and D. Brill, "Acoustic spectrogram and complex-frequency poles of a resonantly excited elastic tube," J. Acoust. Soc. Am. **75**, 1680–1693 (1984).
[25]G. Gaunaurd and W. Wertman, "Transient acoustic scattering by fluid-loaded elastic shells," Int. J. Solids Struct. **27**, 699–711 (1991).
[26]G. Gaunaurd and H. Strifors, "Frequency- and time-domain analysis of the transient resonance scattering resulting from the interaction of a sound pulse with submerged elastic shells," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 313–224 (1993).
[27]C. Tsui, G. Reid, and G. Gaunaurd, "Resonance scattering by elastic cylinders and their experimental verification," J. Acoust. Soc. Am. **80**, 382–390 (1986).

Performance analysis of direct-sequence spread-spectrum underwater acoustic communications with low signal-to-noise-ratio input signals

T. C. Yang[a] and Wen-Bin Yang
Naval Research Laboratory, Washington, DC 20375

(Received 15 May 2007; revised 19 November 2007; accepted 24 November 2007)

Direct-sequence spread-spectrum signals collected from the TREX04 experiment are analyzed to determine the bit-error-rate (BER) as a function of the input signal-to-noise ratio (SNR) for a single receiver. A total of 1160 packets of data are generated by adding ambient noise data collected at sea to the signal data (in postprocessing) to create signals with different input-SNR, some as low as −15 dB. Two methods are analyzed in detail, both using a time-updated channel impulse-response estimate as a (matched) filter to mitigate the multipath-induced interferences. The first method requires an independent estimate of the time-varying channel impulse-response function; the second method uses the channel impulse-response estimated from the previous symbol as the matched filter. The first method yields an average BER $<10^{-2}$ for input-SNR as low as −12 dB and the second method yields a similar performance for input-SNR as low as −8 dB. The measured BERs are modeled using the measured signal amplitude fluctuation statistics and processing gain obtained by de-spreading the received signal with the transmitted code sequence. Performance losses caused by imprecise symbol synchronization at low input-SNR, uncertainty in channel estimation, and signal fading are quantitatively modeled and compared with data.
[DOI: 10.1121/1.2828053]

PACS number(s): 43.60.Dh [EJS] Pages: 842–855

I. INTRODUCTION

Underwater acoustic communications are band limited due to the increased attenuation at higher frequencies. Phase coherent underwater acoustic communications provide an efficient use of the limited bandwidth and have received a great deal of attention recently.[1] Direct-sequence spread-spectrum (DSSS) signaling uses phase coherent signals, where the information symbols are coded/multiplied with a code sequence, commonly known as chips.[2] The signals are processed at the receiver using the code sequence as a matched filter to extract the information symbols.[3,4] Two advantages of DSSS signals are: (1) multiple access communications between the different users using different code sequences which are almost orthogonal to each other,[5,6] and (2) communications at low signal levels (e.g., below the noise level) to avoid detection and interception by an unfriendly party.[3,4] For the former, the focus is on the separation of messages (interference suppression) using code orthogonality. For the latter, the focus is on the signal enhancement for the intended receiver using the processing gain of the matched filter.

The problem for DSSS communications in an underwater acoustic channel is the multipath arrivals which create severe interchip and intersymbol interferences. As the decision feedback equalizer (DFE) has been successfully applied to phase coherent signals, the same approach has been adapted for DSSS communications by Stojanovic *et al.*[7,8] To achieve precise symbol synchronization and channel equalization, high signal-to-noise ratio (SNR) signals are required.[4] Sozer *et al.*[4] applied a RAKE receiver to combine multipath arrivals to enhance the signal SNR. Blackmon *et al.*[9] compared the RAKE receiver with the DFE approach. Iltis *et al.*[10] applied a coherent RAKE receiver coupled with an extended Kalman-filter based estimator for the channel parameters. Although Blackmon *et al.*[9] simulated the performance of the RAKE receiver for weak signals, their article, as well as other simulation work, ignored the data acquisition and symbol tracking problem; acquisition/tracking is a difficult problem for real data at low input-SNRs as the accuracy of symbol synchronization degrades significantly with decreasing SNR. This article studies the performance of DSSS signals using data collected at sea for input-SNR as low as −15 dB. Four different methods are discussed; two are investigated in detail.

The interest in low-input-SNR acoustic communications lies in some practical applications. Acoustic signals much weaker than the ambient noise (e.g., −8 dB SNR within the signal band) are difficult to detect by an unalerted listener. Noise-like signals are difficult to decode without a prior knowledge of the structure of the signal. Communications with low-input-SNR signals at the receiver are said to provide a low probability of interception (LPI) and low probability of detection (LPD).[11] The probabilities of detection and interception are a function of the input-SNR. Naturally, when the interceptor is close to the transmitter, the communication may no longer be LPI/LPD due to the increasing SNR.

To be able to decode symbols from low-input-SNR signals, the symbol energy must be brought above the noise by

[a]Electronic mail: yang@wave.nrl.navy.mil

signal processing. The ratio of the output symbol SNR over the input-SNR is called the processing gain (PG). Using the DSSS method, the received data are despread by a correlator (a matched filter), which correlates the received data with the transmitted code sequence. The despreading provides a matched filter gain (MFG) for the signal, equal, in theory, to the time-bandwidth product of the spreading code. No MFG is expected for random noise, e.g., additive white Gaussian noise (AWGN). Thus theoretically, PG is determined by the MFG; they differ by a small amount in practice.

The DSSS approach uses code "orthogonality" to minimize interference between symbols as well as between users. The code orthogonality requires that the code sequence is almost orthogonal to any of the cyclically shifted code sequences (and to the code sequence of other users). With orthogonality, the matched filtered output yields a low sidelobe level and thus ensures minimum interference. It assures accurate symbol synchronization. However, the orthogonality of the codes is severely degraded in an underwater channel due to the rich multipaths creating inter-chip interference. To mitigate the multipath, DFE and RAKE receiver have been proposed, but as remarked earlier, both require high input-SNR and synchronization at the chip level, and are not designed for communications with low input-SNRs. In this article, an approach, referred to as passive-phase conjugation (PPC),[12] is adapted, which uses a linear filter based on the estimation of the channel impulse-response. For DSSS communications, this method requires only "coarse" synchronization at the symbol level and is suitable for communications at low input-SNRs. The standard PPC method estimates the channel impulse-response from a probe signal transmitted before the communication packet.[13–15] This method is discussed first in this article. It is shown that this method does not work for the 17 kHz data reported in the following. Even at high input-SNRs, the PPC fails because it does not account for the temporal variation of the channel, when the channel coherence time is (significantly) shorter than the packet length.[14,15] The symbol phase fluctuation due to the channel is analyzed below to illustrate how the standard PPC fails. For underwater acoustic communications, the symbol phase is path dependent and the overall phase generally changes rapidly with time from symbol to symbol except for some specific environments (e.g., the Arctic) where the ocean remains stationary. The challenge for DSSS underwater communications is how to remove/compensate the phase fluctuations in a dynamic ocean when the phase change is nonnegligible and disrupts the ability to communicate using phase-shift keying.

The approach undertaken in this article uses a time-updated PPC which uses the time-varying channel impulse-response as the filter. Two particular methods which yield promising results for a single receiver are studied in detail. The first method assumes that the time-varying channel impulse-response is known via an auxiliary noise-like signal. It yields an average bit error rate $\leq 10^{-2}$ for input-SNR ≥ -14 dB. The second method assumes that the channel is slowly varying with a coherence time longer than two symbol durations. It uses the channel impulse-response estimated from the preceding symbol as the filter for the current symbol. This second method is simple and robust. It yields an average bit error rate $\leq 10^{-2}$ for input-SNR ≥ -11 dB. Compared with the previous method, the price paid due to infrequent channel estimate/update is about 3 dB. Another method, called the decision-directed PPC,[16] which estimates the time-varying channel impulse-response directly from the communication signal is discussed in this article but is not pursued because it requires a high input-SNR. For the data analyzed, the length of the spreading code is greater than the multipath delay in order to achieve a high PG and symbol synchronization at low input-SNR. The other benefit is that the intersymbol interference is minimal.

The second part of this article is devoted to signal characterization, and performance analysis, modeling, and prediction. For a fading channel, such as the underwater acoustic channel, signal fading statistics affects the performance, and needs to be known for performance modeling and prediction. In the absence of a prediction capability, it has to be measured from data. Signal fluctuation also influences the MFG and PG; the amount of MFG/PG degradation also needs to be studied. MFG/PG degradation can, in principle, be modeled in terms of the signal and noise coherence. One finds that the degradation is of the order of 1 dB as the symbol duration is less than the coherence-time, as such no theoretical modeling is needed.

In terms of performance analysis, one notes that at high-input-SNRs, the channel can be reliably estimated/updated and symbol synchronization is sufficiently accurate. The performance loss is caused predominantly by signal fading. At low input-SNR, accurate symbol synchronization is difficult; imprecise synchronization is a dominant contributor to the performance loss. A numerical approach is used to model the performance as no analytical prediction is available; although a closed form solution of the bit error rate (BER) is known for a nonfading channel under the condition of perfect symbol synchronization, the same does not exist for a fading channel and/or when synchronization is inaccurate. Using the numerical approach, the performance loss due to imprecise symbol synchronization, imprecise knowledge of the channel variation, and signal fading can be modeled individually. One finds that the loss due to imprecise symbol synchronization at low input-SNR amounts to about 5–8 dB. In comparison, signal fading causes a performance loss of ~3 dB. The result holds whether the channel impulse-response is known for the current symbol or is estimated from the preceding symbol. The numerical model shows a difference of 6 dB in performance between the two methods for BER $\leq 10^{-2}$.

It is noted that the PPC normally requires an array of receivers. Practical systems usually allow only a small number of receivers. The advantage of DSSS is that a single receiver is often sufficient. In this context, PPC is basically a matched filter, or a correlator. Since the filter uses the channel impulse-response, the method is still referred to as PPC. Obviously, the method can be applied to an array of receivers with the added benefit of minimal signal fading and reduced phase variance.[17]

This article is organized as follows. In Sec. II, the BERs of various receiver algorithms for DSSS communications are

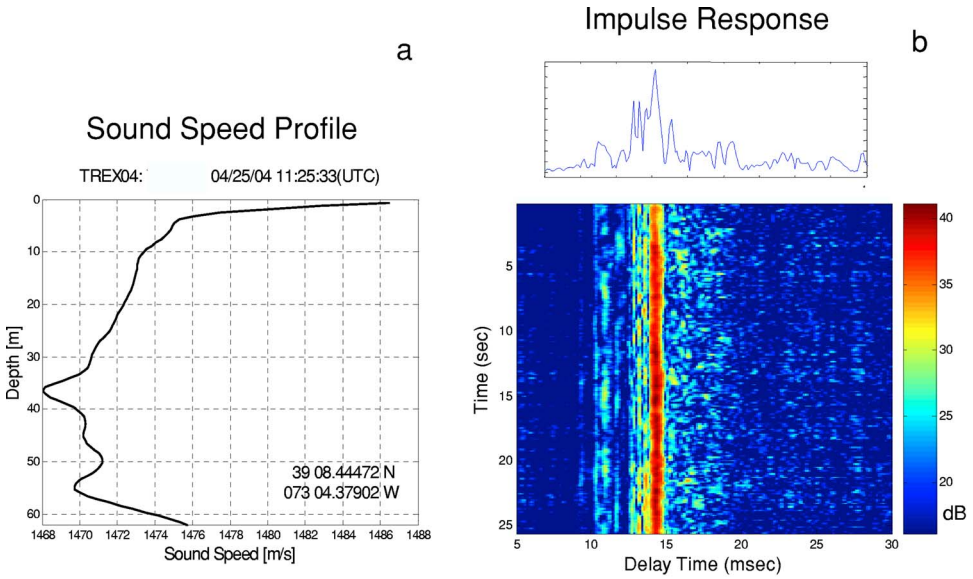


FIG. 1. (Color online) (a) Sound speed profile as a function of depth. (b) The color plot on the right-hand side shows the measured channel impulse-response as a function of time. A linear plot of the impulse-response function (upper panel).

determined as a function of decreasing input-SNR from data collected in the TREX04 experiment. To analyze the performance, the properties of the communication channel need to be studied. This is done in Sec. III where the signal fading statistics, the probability distributions of the MFG and PG are determined from data. The BER is modeled in this section as a function of decreasing SNR; the performance losses due to imprecise symbol synchronization, inaccurate channel estimation and signal fading are modeled individually. In Sec. III, LPD/LPI communications are discussed which have different requirements than that of multiaccess communications, although both use DSSS modulation. A summary of this article is given in Sec. IV.

II. TREX04 AND DATA ANALYSIS

The TREX04 (Time Reversal Experiment) was conducted by the Naval Research Laboratory in April 2004, off the coast of New Jersey, southwest of the Hudson Canyon. Figure 1(a) shows a sound speed profile based upon measurement at the site. Acoustic communication data were transmitted from a fixed source to a fixed receiver array at the range of 3.4 km. Water depth in the experimental area is about 70 m. The source and receivers were located at about 35 m depth. The vertical array has an aperture of approximately 2 m, and contains eight hydrophones with nonuniform spacing. The data presented are from the top receiver; the result applies equally well to other receivers.

The DSSS signals were centered at 17 kHz and had a bandwidth of 4 kHz. The transmitted symbols were spread with (multiplied by) an m-sequence with 511 chips. The sequence of chips was then transmitted using binary phase-shift-keying (BPSK) modulation. Each symbol has a duration of 127.8 ms, which is much longer than the multipath delay [<25 ms as shown in Fig. 1(b)]. As a result, the interference between adjacent symbols is minimal. The multipath arrival structure is estimated from the amplitude channel impulse-response by correlating the received data with the code sequence (the matched filter processor). The correlator output is shown in Fig. 2(a), displaying the amplitude fluc-

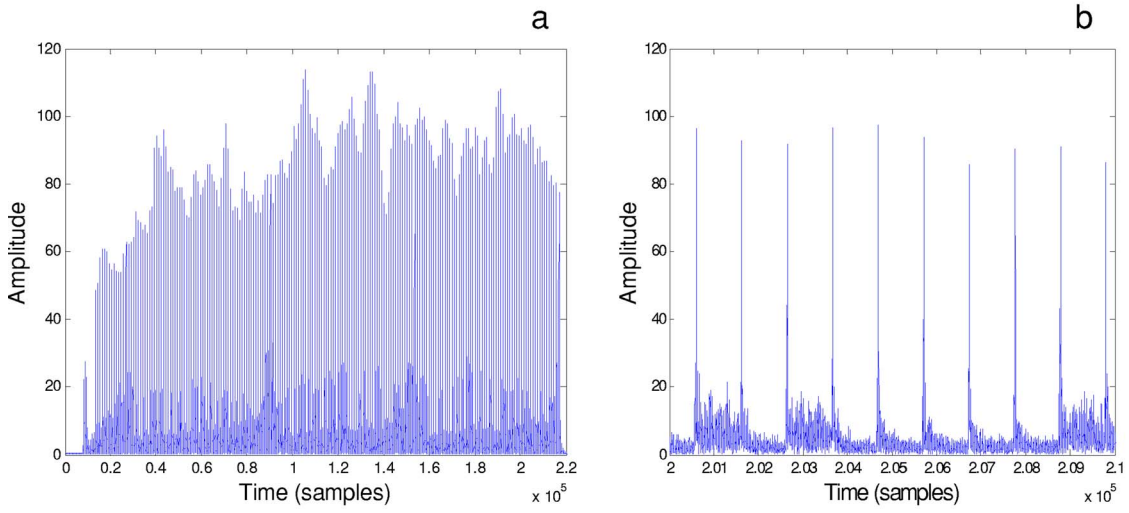


FIG. 2. (Color online) (a) Matched filter correlator output for a packet of data and (b) an expanded view of the left-hand panel near the end.

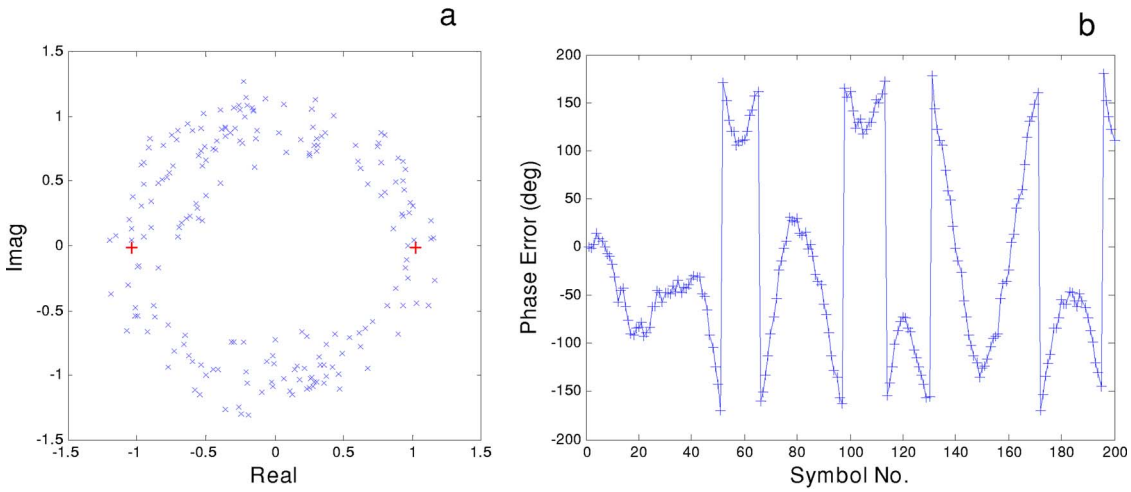


FIG. 3. (Color online) (a) Constellation plot of the symbols using the dominant arrival or the PPC method and (b) the symbol phase error shown as a function of symbol number.

tuation (fading) of the dominant arrivals. The data have an input-SNR of ~23 dB so that the effect of the noise on channel estimation is negligible.

Figure 2(b) plots on an expanded scale the impulse-responses near the end of the packet, which shows there is little intersymbol interference (ISI) between symbols of equal phase due to the near orthogonality of the *m*-sequences. Between symbols of opposite phases, there exist some small ISI as the shifted phase-modulated sequence is no longer orthogonal to the transmitted *m*-sequence. The ISI appears as a higher sidelobe in the correlator output [Fig. 2(b)]. When the symbol duration is longer than the length of the impulse-response, these sidelobes can be separated from the main lobe by time gating. The impulse-responses (Fig. 2) are often displayed in color as a function of geotime and delay time as shown in Fig. 1(b).

Each packet of data is 25.55 s long and contains 200 symbols. A total of 116 packets were transmitted over an hour of period. They are analyzed in this section. Two hours of noise data were also collected producing 278 packets of noise. Multiple packets of noise data, randomly chosen, were added to the 116 packets of signal data to generate a total of 1160 packets of data with different input-SNRs ranging from −15 to +23 dB. BER was evaluated for all the 1160 packets of data using different processors described in the following.

A. Symbol phase error and PPC

Figure 3(a) shows the symbol–constellations plot of the complex symbol amplitudes determined from the dominant arrival path [i.e., the peaks of the matched filter output in Fig. 2(a)]. The true symbols are located at +1 and −1. The received symbols in Fig. 3(a) are unduly scattered and a large number of the symbols are in error. The phase wander is caused by the propagation medium; the symbol phases are significantly modified by the signal propagation in a time varying medium. Figure 3(b) shows the phase error between the received symbols and the transmitted symbols. Symbols which have a phase error beyond ±90° are in error.

In general, the symbol phase error is path dependent as multipaths travel through different water column and have different path lengths. To mitigate the multipath-induced symbol distortions, a channel equalizer will be needed. Stojanovic *et al.*[7,8] used DFE jointly with a phase-locked loop. This method is computationally intensive and requires a high input-SNR (normally greater than 10–15 dB). As the focus of this article is on communications at low input-SNR (< −5 dB), this method will not be pursued any further.

Another method used to mitigate the multipath effect is the PPC mentioned above, also known as the passive time reversal method, which uses the channel impulse-response estimated from the probe signal, or the first symbol, as the matched filter; it can be viewed as a basic time-invariant linear equalizer. One notes that the matched filter output, Fig. 2(a) can be expressed mathematically as

$$r(t) = \sum_n h_n(t - nT)S_n G(t - nT)m(t) + N, \tag{1}$$

where S_n is the nth transmitted symbols ($S_n = \pm 1$ for binary symbols), h_n is the channel impulse associated with the nth symbol, G is a rectangle window which is zero for $t < nT - T/2$ and $t > nT + T/2$ and N denotes the noise. In Eq. (1), $m(t)$ is the autocorrelation function of the spreading code, which, in the ideal case, yields M (the number of chips) at the center of the correlator and one elsewhere.

The PPC processor convolves the correlator output, Eq. (1), with the time-reversed, complex-conjugated channel impulse-response estimated at $t=0$, denoted by $\hat{h}_0$. For simplicity (in order to illustrate the principle of the processing algorithms), the hat over h_0 will be dropped in the equations below, treating the estimated impulse-response as the same as the true impulse-response and leaving the channel estimation error to the numerical estimator. Using PPC, the nth symbol is estimated from the peak of

$$\hat{S}_n = \overleftarrow{h}_0^* \otimes (h_n S_n + N) = (h_0^* * h_n)S_n + N_h \tag{2}$$

where the superscript asterisk denotes the complex conjugation, the inverse-arrow above the impulse-response denotes the time-reversal operator, ⊗ denotes the convolution operator, and asterisk denotes the correlation operator. Note that the convolution of a time-reversed function with another

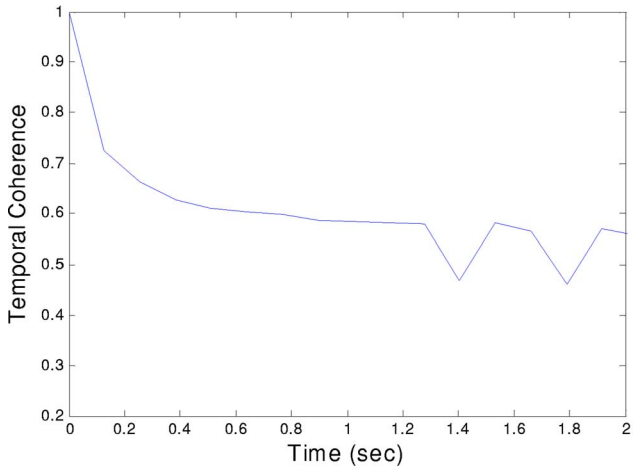


FIG. 4. (Color online) Intrapacket temporal coherence of the channel impulse-responses averaged over 116 packets.

function is the same as the correlation of the function with the other function. In Eq. (2), N_h denotes the filtered noise: $N_h = h_0^* * N$.

The PPC filter (h_0) provides a means to combine the multipath arrivals of the signal coherently.[13,14] For a time-invariant environment, it is expected to provide a higher gain than the incoherent RAKE receiver. For the present data, this standard PPC method performs poorly. The symbol constellation plot and the phase error plot are basically the same as that shown in Figs. 3(a) and 3(b), respectively, except for the first eight symbols which have near-zero phase error. The close similarity between PPC and that based on the dominant path arrival is a manifestation of the fact that the impulse-response is dominated by one major arrival. One concludes that for this environment, the standard PPC does not work.

The poor performance of the PPC method is due to the fact that the PPC filter is time invariant, and does not account for the rapid channel variation seen in the data. The phase error is due to the fact that $h_0^* * h_n$ is not real and has a phase term when the channel changes rapidly with time.

How fast the channel fluctuates can be measured in terms of the signal temporal coherence, which is shown in Fig. 4. The temporal coherence of a broadband signal is obtained by correlating the impulse-response at a reference time with the impulse-response following the reference time with a delay time.[18] The correlation averaged over the reference time is plotted in Fig. 4. Figure 4 shows that the signal temporal coherence drops to below 0.6 in 1–2 s, suggesting that the channel coherence time is of the order of 1 s (or eight symbols).

B. Time-updated PPC based on channel estimation for the current symbol

It is apparent from the previous analysis that to improve the BER (using PPC), one needs to process the data using a time-dependent channel impulse-response function or a time-dependent equalizer. Using the channel impulse at $t=nT$, the symbol is estimated from the peak of

$$\hat{S}_n = (h_n^* * h_n)S_n + h_n^* * N = (h_n^* * h_n)S_n + N_{h,n}. \quad (3)$$

As $h_n^* * h_n$ is the autocorrelation function of the channel impulse-response and is real, the symbol can be properly decoded from the phase of the $\hat{S}_n$ by determining the peak of its absolute value. In the case, the error comes entirely from the noise, and the BER, plotted as a function of the symbol energy (E_b) over the noise spectral density (N_0), should agree with the BER of a BPSK signal in free space in the presence of AWGN. Note that $N_{h,n} = h_n^* * N$ is also white temporally.

Thus, if the channel impulse-response is known as a function of time, one can extend the PPC algorithm to include the time-varying channel impulse-response. The data are divided into blocks. Each block is matched filtered with the channel impulse-response for that block time period. This method will be called the time-updated PPC. It is a channel-estimation based PPC.

To proceed with this method, one needs to estimate the (temporal variation of the) channel impulse-response. Two channel estimation methods are discussed here. One method uses a pilot signal which is transmitted concurrently with the acoustic communication signal. An example of the pilot signal will be a Gaussian white-noise like signal or a pseudo-random BPSK signals with a code sequence that is almost orthogonal to the code sequence used for communications. The other method estimates the channel impulse-response directly from the communication signal.[16] This can be accomplished on a symbol by symbol basis. Assume that the nth symbol has been decoded, one can then estimate the channel impulse-response at $t=nT$ by multiplying the correlator output, Eq. (1), at $t=nT$ with S_n. Using this channel impulse-response as the PPC filter applied to the correlator output at $t=nT$, one obtains an estimate of the symbol at $t=(n+1)T$. A decision is made on this symbol based on the true symbol with the minimum distance to the estimated symbol. Now the symbol at $t=(n+1)T$ is known (decided), the above-mentioned process is repeated to estimate the channel impulse-response at $t=(n+1)T$ and the symbol at $t=(n+2)T$, etc. This method was referred to as decision-directed PPC in Ref. 16, where the method was applied to BPSK signals. In that case, the channel estimation is based on the least mean square error between the modeled data based on a block of decision symbols and the received data. For the DSSS signals, the impulse-response is readily available from the matched filter (correlator) output as described earlier.

For the decision-directed channel estimation, the quality of the channel impulse-response estimation degrades quickly with decreasing SNR. Poor performance is expected for input signal weaker than the noise level. It is not practical for communication at low SNRs.

Using a Gaussian-random pilot signal, a low-level communication can be transmitted using the pilot signal as a cover and is thus undetectable without a prior warning. The time-varying channel impulse-response function is estimated from the pilot signal and used to decode communication signal. Using the channel impulse-response estimated from a packet of data with a high SNR, which is used here to emulate the pilot signal, one evaluates the BER for the same

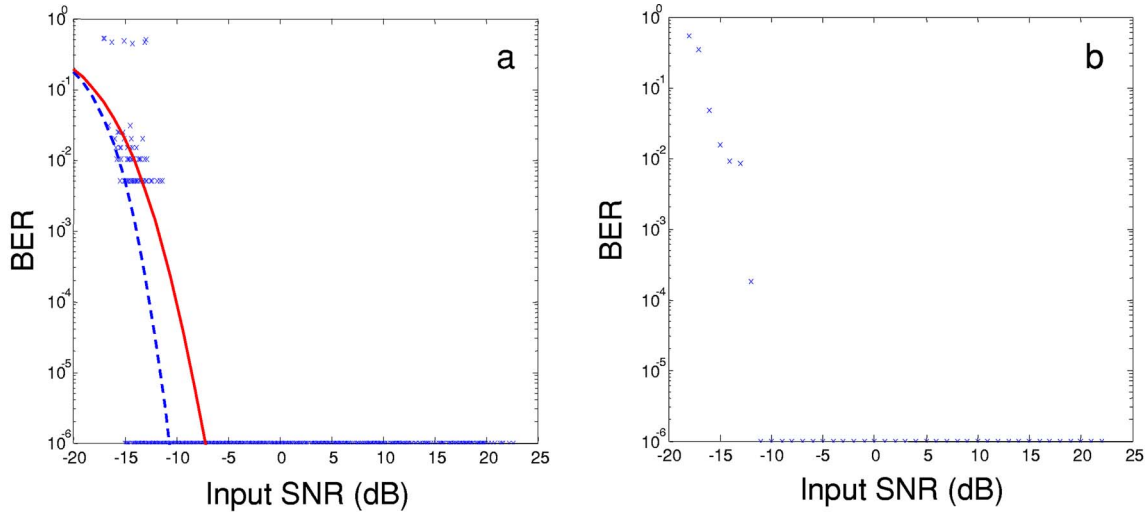


FIG. 5. (Color online) (a) BER of 1160 packets as a function of input-SNR using the method of Sec. II B assuming that the channel impulse response is known. Dashed and solid lines are modeled BER for a non-fading and fading channel, respectively. (b) The average BER plotted as a function of input-SNR. Zero BER is represented by 10^{-6} on the logarithmic scale.

packet of data with at-sea noise added to create signals with decreasing SNR. Symbol synchronization is done for both high and low SNR data by correlating the estimated (time-varying) channel impulse-response with the received data in blocks (symbol by symbol). The resulting uncoded BER is plotted in Fig. 5 as a function of the input-SNR at the receiver. Figure 5(a) shows the BER for the 1160 packets of data; there are 116 packets of data for each input-SNR. The average BER as a function of input-SNR is shown in Fig. 5(b). One finds that the average BER is less than 1% for input-SNR as low as −14 dB. All packets are error free for input-SNR >-10 dB. A substantial number of packets ($>40\%$) are error free even at input-SNR as low as −14 dB. The BER performance will be analyzed in the next section.

C. Time-updated PPC using channel impulse response estimated from the previous symbol

For practical applications, it is more convenient to apply time-update PPC using the channel impulse-response estimated from the symbol immediately preceding the current symbol. This method is simple, since as remarked above this impulse-response is readily available from the matched filter output by removing the symbol phase ($S_n = \pm 1$). It can be easily applied to communication signals with low SNRs to achieve LPI/LPD. The advantage here is that no pilot signal is needed. Note that this method should work as well as the method described in Sec. II B when the channel impulse-response changes little between two consecutive symbols. When the channel changes rapidly between adjacent symbols, there will be a penalty paid in the performance. How much of a penalty will be analyzed in the next section.

As discussed previously, assuming the nth symbol is decoded (decided), one obtains the channel impulse-response plus noise at $t=nT$ from the correlator output. Note that at low SNR, the noise contribution is not negligible. Applying PPC using this estimate of the channel impulse-response, one obtains an estimate of the $(n+1)$th symbol from the peak of:

$$\hat{S}_{n+1} = (h_n^* + N_n^*) * (h_{n+1}S_{n+1} + N_{n+1}) = h_n^* * h_{n+1}S_{n+1} + N_n^* * h_{n+1}S_{n+1} + h_n^* * N_{n+1} + N_n^* * N_{n+1}. \quad (4)$$

For high input-SNRs, the first term dominates; the terms containing the noise are responsible for the symbol error. For low SNRs, all terms contribute to the symbol error.

D. A practical recipe for differential phase-shift keying signal

The last mentioned method requires a decision of the symbol as the processor is applied to the symbols sequentially. An error in the decision will likely propagate down streams. To avoid error propagation, one may use differentially coded BPSK, referred to as DPSK, where the data is carried by the relative phase of the transmitted symbols. The processor output is given by[19]

$$\hat{D}_n = (h_{n-1}^* S_{n-1}^* + N_{n-1}^*) * (h_n S_n + N_n) = h_{n-1}^* * h_n S_{n-1}^* S_n + N_{n-1}^* * h_n S_n + h_{n-1}^* * N_n S_{n-1}^* + N_{n-1}^* * N_n, \quad (5)$$

where the peak of the correlation (the right-hand side) is used to determine the symbol $\hat{D}_{n+1}$. For DPSK, the transmitted symbol S_n is related to the data symbol D_n by

$$S_n = D_n S_{n-1} \quad \text{for } n > 1$$

with $S_1 = 1$.

Using the method described in this section, the symbol constellation for the DPSK signal and the (differential) symbol phase error are plotted in Figs. 6(a) and 6(b), respectively, as a function of the symbol number for the same packet of data analyzed in Fig. 3. One observes that compared with Fig. 3(b), the phase error is significantly reduced. All symbols are correctly determined. Figure 7(a) shows the histogram distribution of the symbol phase error plotted in Fig. 6(b). The phase error follows a Gaussian normal distribution[17] with zero mean and a standard deviation of $\sim 20°$. As noted previously, the signal has an input-SNR of

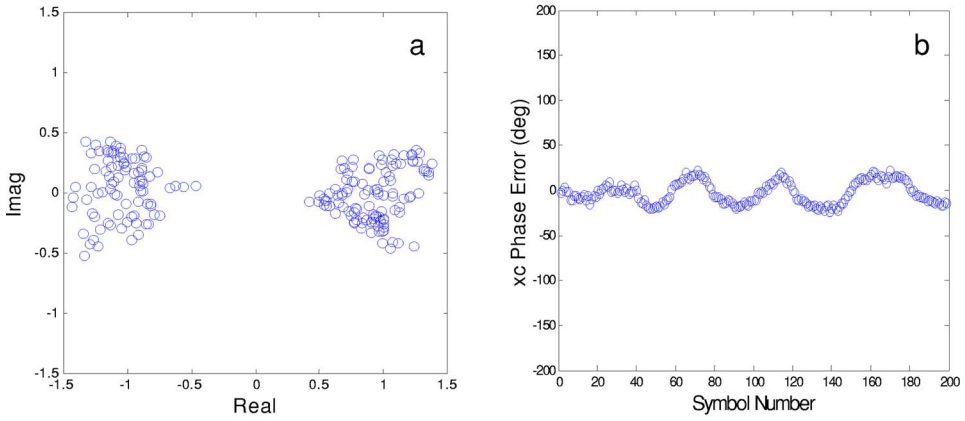


FIG. 6. (Color online) (a) Constellation plot of the symbols using the DPSK method of Sec. II D and (b) the symbol phase error shown as a function of symbol number.

~23 dB, so that noise is not a significant contributor to the phase error. This phase error is due to the channel change from one symbol to the next [the first term of Eq. (5), for high SNR cases].

As symbol phase error is a good indicator of the processor performance,[17] it is interesting to compare this phase error [Fig. 7(a)] with the phase error using the method described in Sec. II B, which is shown in Fig. 7(b) for an input-SNR of −14 dB. Note that using the method of Sec. II B, the phase error at an input-SNR of ~23 dB is negligible small and can be treated as zero for all practical purposes. In this case, the channel impulse-response is known and the phase error is due predominantly to the ambient noise at sea. As the phase error increases with decreasing SNR, eventually the phase error (using the method of Sec. II B) will reach a high standard deviation as shown in Fig. 7(a), this happens at an input-SNR of −14 dB as shown in Fig. 7(b). Note that, in Fig. 7(b), despite the low-level input-SNR, the output signal has a SNR of ~15 dB, assuming a theoretical PG of 27 dB. As a result, one can still achieve a reasonable phase error.

Comparing Fig. 7(a) with Fig. 7(b), one finds that the penalty for not tracking the temporal variation of the channel impulse-response between adjacent symbols is an increased input-SNR to achieve the same variance in phase error. The penalty in input-SNR to achieve the same BER is more complicated and will be analyzed using a numerical model later in the article.

The above-mentioned packet was also processed using the method described in Sec. II C. No bit error was found. It is found that the phase errors for the BPSK symbols (not shown here) have approximately the same variance as the phase errors of the DPSK symbols shown in Fig. 6(b). The problem using the method in Sec. II C is, as mentioned earlier, error propagation once a symbol is incorrectly identified. This happens often when the input-SNR is low. Hence, that method is not pursued any further in this paper. In contrast, error for DPSK signaling using method described in this section is local, i.e., nonpropagating. The average BER is therefore lower and the performance is more robust at low input-SNR.

The BER results using DPSK signaling [Eq. (5)] are shown in Fig. 8 as a function of input-SNR. Figure 8(a) shows the BER for 1160 packets and Fig. 8(b) shows the BER averaged over packets having the same input-SNR. One finds all packets are error free when input-SNR ≥−6 dB. The average BER is less than 1% for input-SNR ≥−11 dB.

Figure 9(a) shows the frequency spectrum of the input signal as a function of time (the spectrogram) for a packet with an in-band input-SNR of −8.6 dB. The signal covers a bandwidth from 15 to 19 kHz. One sees no trace of the communication signal in the spectrogram. Figure 9(b) shows that the symbols are detectable after the matched filter processing; the symbols are coarsely identified from the peak of the time series. Figure 9(a) shows that an unalerted party, not knowing the arrival time of a packet, will be unaware of the presence of a communication signal. It will be difficult for a spectrum-based energy detector to search and detect the signal in blind (LPD). It is even harder for a time-domain energy detector without knowing the signal bandwidth as the SNR in the time series before band pass filter is less than

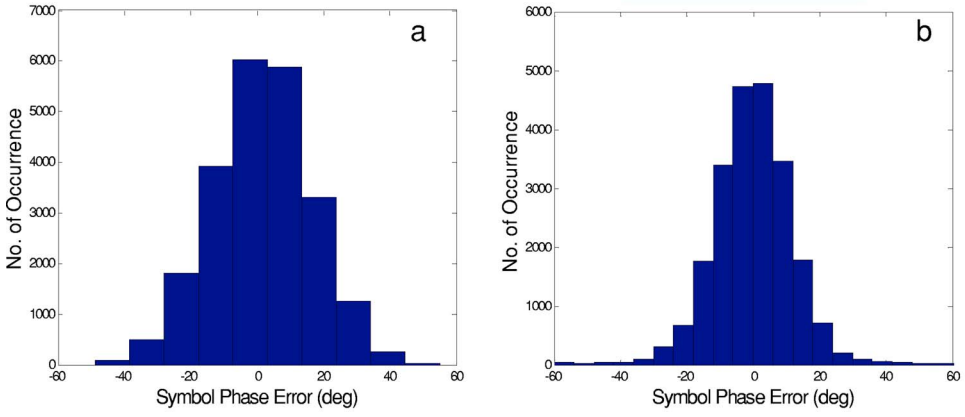


FIG. 7. (Color online) (a) Histogram distribution of phase errors using the DPSK method for input signal SNR ~23 dB. (b) Histogram distribution of phase errors using KCIR method for input signal SNR of −13.7 dB.

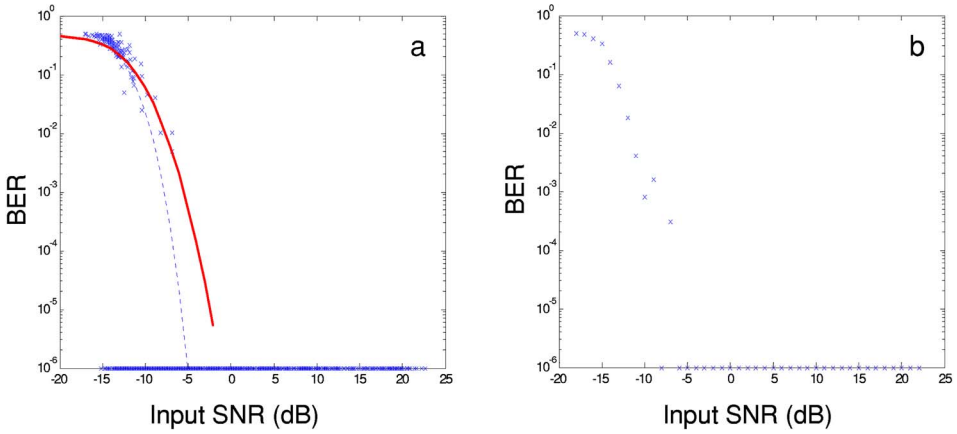


FIG. 8. (Color online) (a) BER of 1160 packets as a function of input-SNR using the DPSK method. Dashed and solid lines are modeled BER for a nonfading and fading channel, respectively. (b) The average BER as a function of input-SNR. Zero BER is represented by 10^{-6} on the logarithmic scale.

−35 dB. The signal will be hard to intercept (LPI) without a prior knowledge of the code length and/or code sequence.

Note that the above analysis used an impulse-response of approximately 20 ms duration covering the dominant arrivals (see Fig. 1). The later (small) arrivals (reverberation returns) are treated as part of the noise. One can choose a different length of impulse response, the trade off between the length of the impulse-response and the contribution of the noise will be discussed in Sec III C.

For the rest of the article, the focus is on the method described in Sec. II B, i.e., time-updated PPC with known channel impulse-response, and the method described in Sec. II D, i.e., DPSK using channel impulse-response estimated from the preceding symbol. They will be henceforth referred to, in short, as the known channel impulse-response method (KCIR) and the DPSK method, respectively.

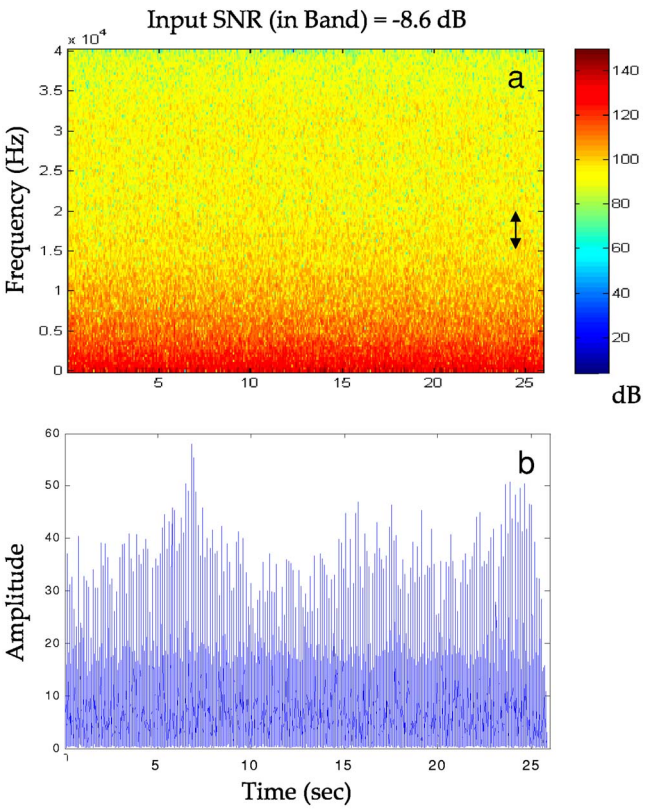


FIG. 9. (Color online) (a) Lofargram of the received signal with an in-band input-SNR pf −8.6 dB. The signal occupying a band from 15 to 19 kHz is not visually observable. (b) The symbols are detectable from the correlator output.

III. DATA CHARACTERIZATION AND PERFORMANCE MODELING

One notes in Figs. 5(a) and 8(a) that, above a certain input-SNR threshold [−10 dB for Fig. 5(a) and −6 dB for Fig. 8(a)], all packets are error free. At input-SNR below the threshold, some packets are error free while others have BER greater than 1%. The reasons for the bit errors can be traced to signal fading, processing gain loss, symbol synchronization error and channel estimation error when channel is known. The corresponding performance loss is analyzed and modeled in this section. Another performance loss is due to channel impulse mismatch using the impulse-response estimated from the preceding symbol. The loss is analyzed for the DPSK method described in Sec. II D.

A. Signal fading statistics

One notes that for packets in which the symbol energy is approximately uniformly distributed (little or no signal fading), the BER is either very low when the signal energy after despreading is above a certain threshold or quite high when the signal energy is significantly below the threshold. For packets in which the symbol energy varies significantly within the packet (significant signal fading), a significant part of the symbols are error free when their energies are above the threshold and the rest are in error because of insufficient energy (when the signal fades). Thus, to model the performance loss, one needs to know the signal-fading statistics.

To determine the input signal amplitude fluctuation statistics, the input data is band-pass filtered, and shifted to the base band. The matched filter output, Eq. (2) provides a coarse synchronization of the signal arrival time and division of the data into blocks, each one of duration T. The input signal level is determined from the signal variance for each block of data. The histogram of the square root of the signal level yields the input signal amplitude fluctuation distribution as shown in Fig. 10(a). A total of 200×116 blocks (samples) are included in Fig. 10. One finds that the distribution of the input signal amplitude is well modeled using the log normal distribution,

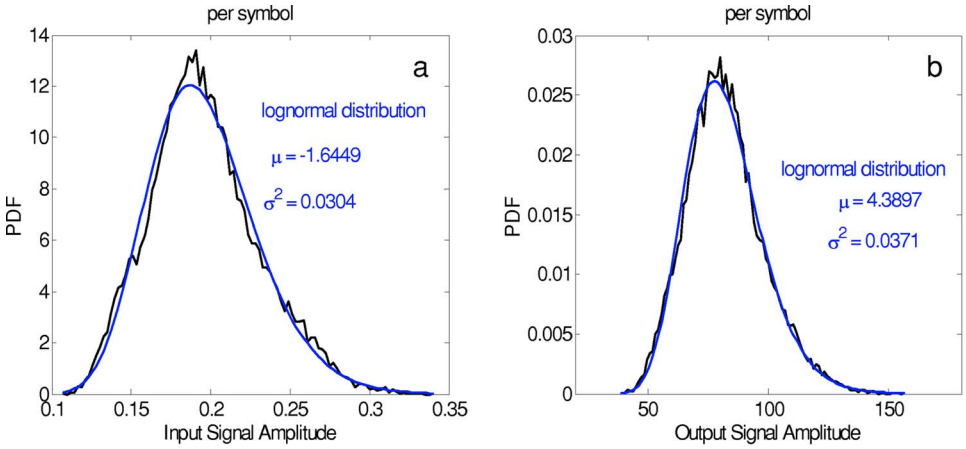


FIG. 10. (Color online) Probability distribution of signal amplitude for (a) input and (b) output signals compared with the log normal distribution.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} x^{-1} \exp\left(-\frac{(\ln x - \mu_x)^2}{2\sigma_x^2}\right), \tag{6}$$

where x is the signal amplitude. A fit to the data is shown in Fig. 10 using $\mu_x = -1.6449$ and $\sigma_x^2 = 0.0304$. The amplitude distribution of the output signals is shown in Fig. 10(b). It also follows a log normal distribution with parameter values given by $\mu_y = 5.0829$ and $\sigma_y^2 = 0.0371$.

As shown previously, the matched filter output, Eq. (2), yields an impulse. The pulse energy is the symbol energy, conventionally denoted as E_b, which can be easily obtained (numerically) from the peak value of the autocorrelation of the impulse-response functions, $E_b(h_n) = \Sigma_i |h_{n,i}|^2$, where i is the index for delay time. The histogram distribution of the square root of E_b yields the amplitude distribution of the (output) symbol, which is shown in Fig. 10(b).

For performance prediction (and modeling) one is interested in the probability distribution in terms of the input- or output-SNR, which is shown in Figs. 11(a) and 11(b), respectively, in terms of the SNR in decibels for the data analyzed. The majority of the packets have input-SNR below 0 dB. Figure 11 shows that the input- and output-SNR has a similar distribution, shifted by approximately 27 dB, the theoretical PG.

B. Measurement of MFG and PG

MFG is by definition the processing gain using the matched filter. For high SNR (or no noise), it is the ratio of the matched-filter signal output-level over the input signal level. For low SNR, the interest is in the enhancement of the SNR using the matched filter; in this case, the noise property also influences the outcome. To distinguish the two cases, the processing gain for the signal (high SNR case) will be referred to as MFG and the processing gain for the SNR will be defined as the PG in this paper.

In the absence of multipath, the matched filter output for a block of signal containing one code sequence is a simple impulse, the pulse energy over the input signal level (the variance of the input signal over the block) is the MFG. In this case, the MFG is given by the time-bandwidth product of the signal, which is equal to the number of chips in the code sequence. In the presence of multipath, the matched filter output contains many multipath arrivals. The MFG can be measured by the energy of the impulse, E_b over the variance of the input signal. Using the 23 200 symbols spread over 116 packets, the histogram of the measured MFG is plotted in Fig. 12.

MFG can be measured in two ways. Figure 12(a) shows the MFG for blocks of signals having approximately the same input signal level. The signal level is divided into 16 intervals. The measured MFG results are shown in Fig. 12(a) for six intervals; the edge intervals have a small number of samples and are ignored. One finds that the MFG for different input signal levels have approximately the same distribution (within the statistical error associated with the limited sample size).

The other method measures the MFG from 116 packet of (high SNR) data in terms of the average output-SNR (the average E_b over the number of samples in the packet) over

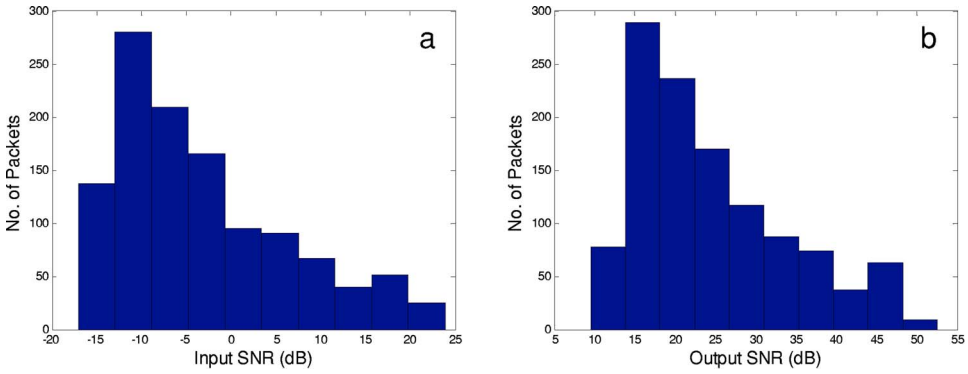


FIG. 11. (Color online) Probability distribution of the (a) input-SNR and (b) output-SNR in decibels.

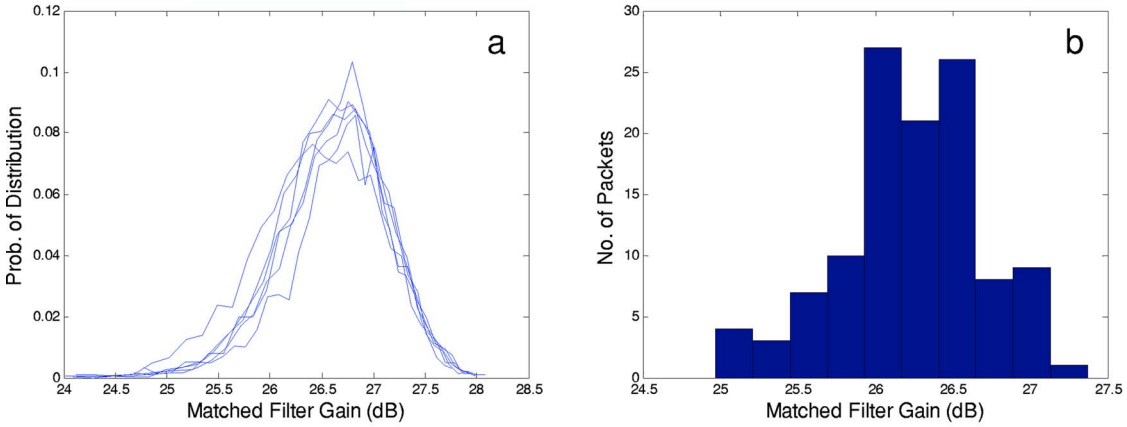


FIG. 12. (Color online) (a) Distribution of matched filter gain measured from symbols of approximately the same input signal levels; the different curves have different signal amplitudes ranging from 0.17 to 0.23 [see Fig. 10(a)]. (b) Distribution of matched filter gain measured via the ratio of mean output signal level and mean input signal level for each packet of data.

the average input-SNR (the variance of the input signal over the entire packet length). The result is shown in Fig. 12(b), using 116 packets.

The later method is used to measure the PG by measuring the average output-SNR over the input-SNR for each packet. All 1160 packets are used. The result is shown in Fig. 13(a) to compare with the MFG of Fig. 12. The PG can be fitted with a Gaussian distribution,

$$p_z(z) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left[-\frac{(z-\mu_z)^2}{2\sigma_z^2}\right], \tag{7}$$

with μ_z=596.0(~27.75 dB) and σ_z=76.7. A fit to the data is shown in Fig. 13(b) in a linear scale.

One finds the mean MFG is about 424.2 or 26.3 dB, about 0.7 dB below the theoretical MFG of 27 dB. The mean PG is 596.0 or 27.75 dB, about 0.7 dB higher than the theoretical PG of 27 dB. Both the MFG and PG have a variation of about ±0.5 dB. The small variance suggests that MFG and/or PG can be approximated by a constant in system design. The fluctuation of MFG or PG is small compared with the fluctuations of the signal amplitude due to signal fading.

C. BER modeling in a time invariant environment

In this and next sections, the BER is modeled using numerical simulations. For clarity, all the BERs in this article are the uncoded BERs, as no error-correction coding is used here. The modeling work shall focus on the two methods analyzed in details in Sec. II, namely, the KCIR method and the DPSK method. To separate the effect of signal fading from the other sources for bit error, a time invariant channel is used first. The focus here is on the performance loss due to imprecise or coarse synchronization of symbol timing and imperfect knowledge of the channel impulse-response. Performance loss is here measured in terms of the increase in input-SNR required to achieve the same BER.

One starts by simulating the matched filter output, Eq. (1), for a random binary signal of amplitude ±1. The model assumes that the channel has a fixed (complex) channel impulse-response as shown in Fig. 1 (upper right). The matched filter output is then obtained by modulating the impulse-response with a random sequence of binary symbols (of amplitude ±1). Randomly generated white noise having

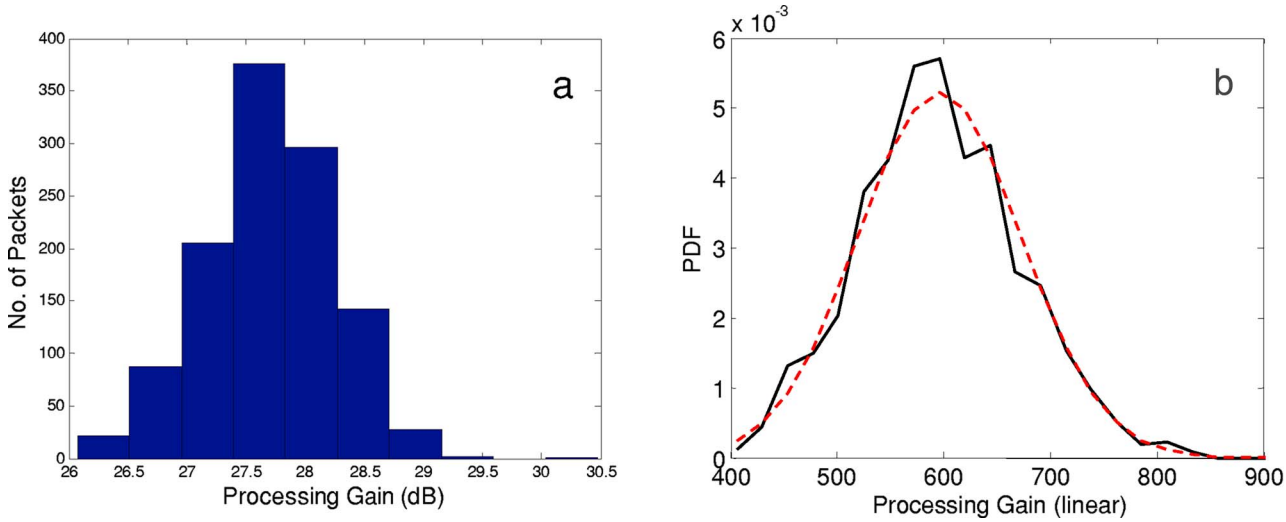


FIG. 13. (Color online) (a) Distribution of processing gain measured from the ratio of mean output-SNR and mean input-SNR for each packet of data. (b) Data, the wiggly curve, compared with a Gaussian distribution, shown as the dashed curve.

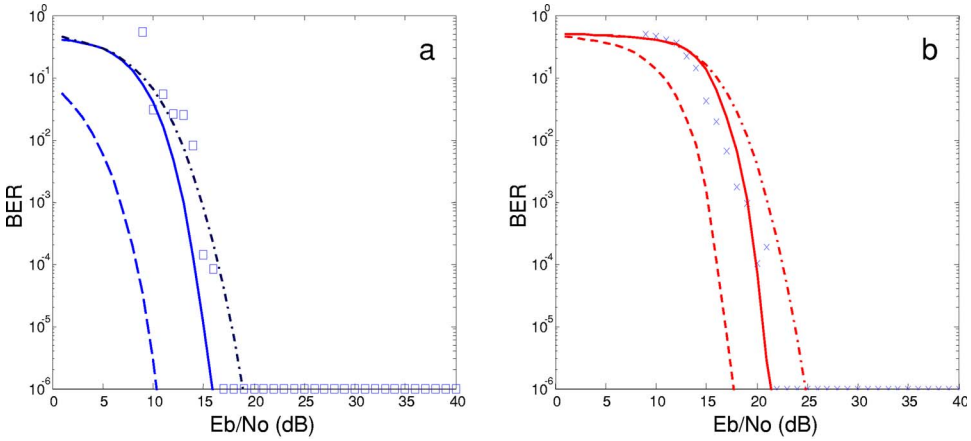


FIG. 14. (Color online) BER plotted as a function of output-SNR (E_b/N_0) and compared with modeled calculations using (a) the KCIR method and (b) the DPSK method. The left-most curves assumed a nonfading channel with perfect symbol synchronization. The middle curves use a correlator for symbol synchronization (coarse synchronization) for a non-fading channel. The right-most curves use a correlator for symbol synchronization (coarse synchronization) for a fading channel.

the same bandwidth (as the signal) but different amplitude is added to the signal to generate signals of different SNR. This signal is then processed using the KCIR and DPSK methods. To model BER to the level of 10^{-5}, a total number of 10^7 symbols is generated.

Figure 14(a) shows the BER as a function of the symbol E_b/N_0, using the KCIR method. The result assuming perfect symbol synchronization is shown by the left-most curve. In this case, the BER agrees with that predicted for BPSK in free space in the presence of AWGN (as has been previously expected). Although matched filter correlation yields reasonably accurate symbol synchronization at high SNRs, the precision degrades severely as SNR decreases. At low input-SNRs, the peaks of the matched filter output, [Fig. 2(a)], are often shifted/adjusted by the noise, which is as strong as the signal. Likewise, the peak of the PPC output (the correlation of the simulated data with the impulse-response) can also be shifted from the true location. As a result, the symbol identified by the peaks of the PPC output can incur a phase error from the noise and is consequently incorrectly decoded when the phase error is large. The cause of the bit error is due to coarse (imprecise) synchronization using the matched-filter at low input-SNR. The BER using coarse synchronization is shown as the middle curve in Fig. 14(a).

A discussion on the noise contribution is in order here. Recall in Eq. (3), the noise contribution is given by $N_{h,n} = h_n^* * N$. Note that the impulse-response has a multipath spread of ~20 ms which is in this case much shorter than the symbol duration $T=127.8$ ms. To minimize the N_h, one should limit the time span of h to ~20 ms (zero filled the rest) in the processor [Eq. (3)]. This seems desirable as it improves the output-SNR. However, a small number of noise samples are signal-like; the signal-like noise is responsible for some of the synchronization errors at low input-SNR.

Next, the BER using the DPSK method of Sec. II D is simulated. In this case, the correlator output of the previous symbol, Eq. (1), is used to correlate with the correlator output of the current symbol, as described in Eq. (5). Even with precise synchronization, additional bit errors occur due to the additional noise terms in Eq. (5) as compared with Eq. (3). The BER using this method, assuming perfect symbol synchronization, is shown by the left-most curve in Fig. 14(b) as a function of the symbol E_b/N_0. The BER using coarse synchronization is shown by the middle curve in Fig. 14(b).

D. BER model including signal fading statistics

To compare with the measured BER reported in Sec. II, the BER will be modeled using Eq. (6), which includes the signal fading distribution of the output signal. The results are shown in Figs. 14(a) and 14(b) by the right-most curves. Also shown in these figures, are the measured BER, as reported in Figs. 5(b) and 8(b), now plotted in terms of the output-SNR: E_b/N_0. Note that if the PG were a constant, the data in Figs. 14(a) and 14(b) would be just a linear shift from Figs. 5(b) and 8(b) by the PG. One finds that the modeled BER results are in good agreement with the data for the KCIR method [Fig. 14(a)]. For the DPSK method, the modeled BER agrees with data for $E_b/N_0 < 15$ dB but is ~3 dB lower than the data for $E_b/N_0 \geqslant 15$ dB. The reason is not yet understood.

From Fig. 14, the modeled results suggest, for both KCIR and DPSK methods, that performance degradation is 5–8 dB due to coarse (imprecise) symbol synchronization at low input-SNR, that is, to achieve the same BER an extra 5–8 dB is required. Performance loss is up to 3 dB due to signal fading. Performance loss due to signal fading decreases with decreasing E_b/N_0. This is expected as noise plays an increasingly important role at lower SNR.

Next, the performance prediction capability will be addressed. For practical applications, BER performance prediction is expressed in terms of input-SNR. The BER as a function of input-SNR, μ_x, is by definition the product of the probability of bit error, $p_e^{\text{in}}(x)$, for a given input-SNR and the probability distribution of the input signal

$$\text{BER}(\mu_x) \equiv \int p_e^{\text{in}}(x) p_X(x, \mu_x) dx. \tag{8}$$

The BER can also be expressed as a function of the output-SNR, μ_y,

$$\text{BER}(\mu_y) = \int p_e^{out}(y) p_Y(y, \mu_y) dy$$

$$= \int \int p_e^{\text{out}}(zx) P_Z(z) p_X(x, \mu_x) dz\, dx, \tag{9}$$

where $p_Z(z)$ is the probability distribution of the PG, given by Eq. (7), $p_e^{\text{out}}(y)$ is the probability of error, given an output-SNR, y, for a nonfading environment. Note that the PG is

assumed to be a function of the ratio of output- and input-SNR, $z=y/x$, only (Fig. 12). The relationship between the two BERs is as follows:

$$p_e^{\text{in}}(x) = \int p_e^{\text{out}}(zx)p_Z(z)dz. \tag{10}$$

Having measured the fading statistics of the input signals, Eq. (6), and the PG, Eq. (7), and the probability of error $p_e^{\text{out}}(y)$, for a nonfading environment, one can model BER in terms of the input-SNR using Eqs. (8) and (10). First, assume that the channel impulse-response is known but the symbol synchronization is coarse, as precise synchronization is difficult in practice at low SNR. This is referred above as the KCIR method. Using the middle curve in Fig. 14(a) for $p_e^{\text{out}}(y)$, the BER is calculated as a function of input-SNR. This is shown as the solid curve in Fig. 5(a). One finds that the majority of the data are in good agreement with the model. A small number of packets (the upper left corner) show a much higher BER than anticipated, most likely due to gross error in the initial data synchronization. Also shown is the BER for a nonfading channel, i.e., Eq. (10), plotted in Fig. 5(a) as the dashed line.

Next, one models the data using the DPSK method, which uses the correlator output of the previous symbol as the PPC filter. One is reminded that the symbol synchronization using the correlator is imprecise at low SNR. Using the middle curve in Fig. 14(b) for $p_e^{\text{out}}(y)$, one can calculate the BER as a function of input-SNR, shown as the solid curve in Fig. 8(a). One finds that the modeled calculation for the average BER is in good agreement with the data. The BER for a nonfading channel, i.e., Eq. (10), is also shown in Fig. 8(a) (the dashed line). As expected, it shows a lower BER than that determined from the data.

A remark is in order here. In Eq. (8), signal fading is in terms of the input signal. The BER, $p_e^{\text{in}}(x)$, is for a nonfading channel, which is related the BER of the output signal, $p_e^{\text{out}}(y)$, also for a nonfading channel. An alternative approach is to express the signal fading in the output signal. In that case, $p_e^{\text{out}}(y)$, is the average BER for a fading channel, as expressed by the right-most curve in Fig. 14(b). Using Eq. (10), $p_e^{\text{in}}(x)$ should then represent the average BER for the input signal for a fading channel. One notes that the BERs derived using the two approaches should be the same (although not immediately apparent from the equations), if the probability of signal fading is independent of the statistics of the processing gain, which seems to be supported by the data. Numerically, one finds that the BERs calculated using two approaches are very close to each other.

The conclusion of this section is that signal fading is nonnegligible in real data and can cause performance loss by ~3 dB. A close form expression of BER is often known for a nonfading channel, but a similar expression is often not known for a fading channel. One finds that the BER for a fading channel can be adequately predicted given a theoretical expression of BER for a nonfading channel, and experimental measurement of the signal fading statistics.

IV. DISCUSSION

As mentioned in Sec. I, a well known application of the DSSS signaling is multiple-access communications where different users use different code sequences. This method of multi-access communications is known as the code-division multiple-access (CDMA) communications. In a simple propagation environment (no multipath), using orthogonal codes, the message can be extracted by projecting the signal to the user's code space, using a correlator with the user's code. The other application is for LPD/LPI communications, where the received signal is extracted from the noise by the correlator using the user's code. In this case, the focus is on the PG achieved by de-spreading the signal. The two different applications thus have different requirements.

Multiple-user CDMA communications require precise symbol synchronizations. This is usually done with a probe signal place ahead of the data packet. Code sequences used for multiuser CDMA communications need to be orthogonal to each other (at arbitrary delay time) and to the cyclically shifted version of itself. To ensure code orthogonality, the signal needs to be synchronized at the chip level. This is sometimes difficult in a time-varying, multipath-rich environment, such as the underwater channel, when the arrival of the highest amplitude fluctuates from one path to the other or when the interference of the multipaths changes the peak locations of the correlator outputs. The problem is more pronounced when the multipaths have different Doppler shifts and/or when the Doppler is changing rapidly with time. This is the symbol synchronization and tracking problem. Using the DFE, the channel needs to be updated at either the symbol rate or the chip rate depending on the channel coherence time.[5,6] For multiaccess communications interferences from other users are as critical an issue as intersymbol interference of its own data. DFE requires a significant number of training data (e.g., 500 chips) depending on the convergence rate of the equalizer. The processing is computationally complex. High SNR (>10 dB) signals are usually required for precise symbol synchronization and DFE.

For LPD/LPI communications with an input signal much weaker than the ambient noise at the intended receiver, the DFE approach is not applicable. The objective of DSSS communications lies, in this case, in the processing gain. Because of the low input-SNR, LPD/LPI communications are constrained by the following:

(i) The lower the input-SNR, the higher the PG needed.
(ii) To minimize interference between symbols, the code sequence length needs to be longer than the multipath time delay. The longer the code sequence, the higher the PG, and consequently the lower the data rate.
(iii) To avoid detection, no probe signals with a high SNR should be used. Precise symbol synchronization is thus difficult at low SNR.
(iv) For low data rate communications, zero training data are desired. Error correction coding further decreases the data rate and thus the uncoded BER becomes more critical.
(v) Because of the low data rate, either a long packet or multiple packets of sufficient length would need to be trans-

mitted for communicating a decent-size message. Hence, the communication method must not be constrained by the channel coherence time.
(vi) From the signal processing point of view, channel estimation is inaccurate at low SNR.
(vii) From the practical point of view, LPD/LPI communications are often limited by the available hardware. In many applications, only one receiver is available, as (widely spaced) multiple sources/receivers are operationally unavailable.
(viii) The (modem) processor can support only a simple algorithm since the power supply is limited. Yet, it requires robust performance (zero BER) under low SNR signal conditions.

The DSSS signaling method meets all the above-mentioned requirements; the data rate is reduced depending on the amount of PG required. Among the different processing methods discussed in Sec. II, the DPSK method is the simplest and thus most favored for practical use; it offers a robust performance at the expense of a 4–6 dB loss in SNR compared with the KCIR method. When only one receiver is available, one needs an additional SNR (~3 dB in Fig. 14) compared with a nonfading channel. Note that although signal fading degrades the performance at high SNR, the performance loss at low SNR is due predominately to coarse (imprecise) symbol synchronization. (Signal fading can be mitigated using spatial diversity or frequency diversity, when multiple receivers or frequency bands are available.) For symbols with low SNR, coarse synchronization is associated with incorrect timing of the peaks of the matched filter output [Fig. 2(a)], which have been shifted/adjusted by the noise, as well as incorrect timing based on the peaks of the PPC output (the correlation of the impulse-responses between the current and previous symbol), which have been shifted by noise as well as inherent channel temporal variations. One can improve the symbol timing of the low SNR symbols based on the timing (determined from the matched filter output) of adjacent symbols with higher SNR as the symbol period is known. To improve the bit errors of low SNR symbols, the symbol phase can be estimated from the PPC output based on the symbol period instead of using the peak amplitude. This works if the channel impulse response has not changed significantly. Under the circumstances, the bit error rate can be significantly improved [~5 dB, see Fig. 14(b)].

Although, as shown previously, the DPSK method has successfully mitigated the phase fluctuation problem for fixed source and fixed receiver transmission, the method fails to achieve a reasonable BER when the same signal is transmitted from a moving source to the (same) fixed receiver. The reason is a much larger phase fluctuation which is not completely removed by the DPSK method with Doppler correction. Because the Doppler shift is dependent on the arrival angle of the multipath, the interference between the multipath arrivals causes a rapid change in the (phase of the) channel impulse-response between symbols, exceeding error tolerance for symbol-decision. It is anticipated that the KCIR method could still work under the circumstance. However, this method is complex and not practical; whereas it is LPI, it is not LPD. A new and different method is proposed for moving source DSSS communications and presented in a separate article. This does not diminish the value of the DPSK method as there exists many applications for LPI/LPD communications between fixed sources and receivers, such as between the nodes of a bottom deployed network.

Last, note that DSSS yields a high PG at the expense of the (reduced) data rate. Fortunately, neither the KCIR nor the DPSK method is limited by the channel coherence time and thus can be used to send a packet containing a long message. For the method to claim LPD/LPI, a quantitative evaluation of the probability of detection and interception will be needed. This topic will also be addressed in the moving source paper as the probability of detection and interception is range dependent. It involves detailed numerical calculations and is beyond the scope of this article.

V. SUMMARY

The DSSS method spreads the communication symbols by a code sequence and despread the signal at the receiver to recover the transmitted symbols. Despreading is done by a matched filter using the code sequence during transmission; it is referred to as the correlator. In the presence of multipath, the correlator output includes the multipath arrivals of the same symbol as well as intersymbol interference. An equalizer is needed to mitigate the multipath induced distortion and combine the multipath arrivals to recover the transmitted symbol. Multichannel DFE has been applied to DSSS signals but the method is computationally intensive and requires a high input-SNR. An approach using the concept of time-updated PPC is adapted here for communications at low SNRs. It uses a linear equalizer based on channel estimation as a function of time. Two specific methods are addressed in this article. One requires a supplementary channel estimation for each symbol time frame, the other uses the channel impulse-response or the correlator output from the previous symbol as the filter for the current symbol. These methods are evaluated against data collected from the TREX04 experiment. Many packets of ambient noise data are added to the signal data to create input data for input-SNR as low as −15 dB. The first method yields a good performance (average BER $<10^{-2}$) at input-SNR as low as −14 dB. The second method is simple and robust at the expense of a somewhat higher SNR (average BER $<10^{-2}$ at input-SNR as low as −11 dB). It is more suitable for practical applications.

The BER performance is modeled in this paper as a function of input-SNR and output-SNR (E_b/N_0). BER for a fading channel is influenced by the signal fading statistics, which characterizes the probability of signal fading. In the absence of a modeling capability, the fading statistics at a time scale of ~128 ms (symbol duration) is measured; it follows a log normal distribution. Signal fluctuation also degrades the MFG of the correlator. One finds the degradation is ~1 dB as the signal coherence remains above 0.8 for the symbol duration. The measured PG exceeds the theoretical value by ~1 dB due, presumably, to noise not being white.

With the aid of a numerical model, the performance loss: (1) due to imprecise (coarse) symbol synchronization at low input-SNR, (2) when the channel impulse-response is estimated from the preceding symbol (the DPSK method), and (3) between a fading and non-fading channel, can be modeled separately. The results suggest that performance degradation is 5–8 dB due to coarse (imprecise) symbol synchronization at low input-SNRs, that is, to achieve the same BER an extra 5–8 dB is required (for the two methods). Performance loss is up to 3 dB due to signal fading. Performance loss due to signal fading decreases with decreasing E_b/N_0. The model also suggests that the DPSK method which estimates the channel impulse-response from the preceding symbol would require ~6 dB more SNR than the KCIR method for which the channel impulse-response is known for the current symbol. It is found that the measured BER using the KCIR method agrees with that modeled. Using the DPSK method, the measured BER, for E_b/N_0 between 15 and 23 dB, is somewhat (2–3 dB) better than that modeled (Fig. 14). This suggests that the synchronization error is not as bad as modeled at these values of E_b/N_0; whether this holds in general needs more investigation.

In conclusion, this article demonstrated that DSSS works at low input SNR with only one receiver. The PG derived from the (long) code sequence is realized in real data and could prove useful for practical applications.

ACKNOWLEDGMENTS

This work is supported by the Office of Naval Research. The authors are grateful to the NRL Acoustic Division personnel who supported and conducted the TREX04 experiment, and to D. Green for discussions on LPD/LPI.

[1]D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).
[2]J. G. Proakis, *Digital Communications* (McGraw–Hill, New York, 2001).
[3]M. Stojanovic, J. G. Proakis, J. A. Rice, and M. D. Green, "Spread spectrum underwater acoustic telemetry," *Proceedings of MTS/IEEE OCEANS '98*, 1998, Vol. **2**, pp. 650–654.
[4]E. M. Sozer, J. G. Proakis, M. Stojanovic, J. A. Rice, A. Benson, and M. Hatch, "Direct sequence spread spectrum based modem for underwater acoustic communication and channel measurements," *Proceedings of MTS/IEEE OCEANS '99*, 1999, Vol. **1**, pp. 228–233.
[5]C. C. Tsimenidis, O. R. Hinton, A. E. Adams, and B. S. Sharif, "Underwater acoustic receiver employing direct-sequence spread spectrum and spatial diversity combining for shallow-water multi-access networking," IEEE J. Ocean. Eng. **26**, 594–603 (2001), and references therein.
[6]J. A. Ritcey and K. R. Griep, "Code shift keyed spread spectrum for ocean acoustic telemetry," *Proceedings of MTS/IEEE OCEANS '95*, 1995, Vol. **3**, pp. 1386–1391.
[7]M. Stojanovic and L. Freitag, "MMSE acquisition of DSSS acoustic communications signals," *Proceedings of MTS/IEEE Oceans '04*, 2004, Vol. **1**, pp. 14–19.
[8]M. Stojanovic and L. Freitag, "Hypothesis-feedback equalization for direct-sequence spread-spectrum underwater communications," *Proc. of MTS/IEEE OCEANS '00*, 2000, Vol. **1**, pp. 123–129.
[9]F. Blackmon, E. M. Sozer, M. Stojanovic, and J. Proakis, "Performance comparison of RAKE and hypothesis feedback direct sequence spread spectrum techniques for underwater communication applications," *Proceedings of MTS/IEEE OCEANS: '02*, 2002, Vol. **1**, pp. 594–603.
[10]R. A. Iltis and A. W. Fuxjaeger, "A digital DS spread-spectrum receiver with joint channel and Doppler shift estimation," IEEE Trans. Commun. **39**, 1255–1267 (1991).
[11]G. D. Weeks, J. K. Townsend, and J. A. Freebersyser, "A method and metric for quantitatively defining low probability of detection," *IEEE Military Communication Conference Proceedings*, 1998, Vol. **3**, pp. 821–826.
[12]D. R. Dowling, "Acoustic pulse-compression using passive phase-conjugation processing," J. Acoust. Soc. Am. **95**, 1450–1458 (1994).
[13]A. Silva, S. Jesus, J. Gomes, and V. Barroso, "Underwater acoustic communication using a 'virtual' electronic time-reversal mirror approach," in *Proceedings of the Fifth European Conference on Underwater Acoustics*, edited by M. E. Zakharia, European Commission, Luxembourg, 2000, pp. 531–536.
[14]D. Rouseff, D. R. Jackson, W. L. J. Fox, C. D. Jones, J. A. Ritcey, and D. R. Dowling, "Underwater acoustic communication by passive-phase conjugation:Theory and experimental results," IEEE J. Ocean. Eng. **26**, 821–831 (2001).
[15]P. Hursky, M. B. Porter, J. A. Rice, and V. K. McDonald, "Passive phase-conjugate signaling using pulse-position modulation," in *Proceedings of the MTS/IEEE OCEANS'01 Conference*, 2001, Vol. **4**, pp. 2244–2249.
[16]J. A. Flynn, J. A. Ritcey, D. Rouseff, and W. L. J. Fox, "Multichannel equalization by decision-directed passive phase conjugation: Experimental results," IEEE J. Ocean. Eng. **29**, 824–836 (2004).
[17]T. C. Yang, "Temporal resolution of time-reversal and passive-phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**, 229–245 (2003).
[18]T. C. Yang, "Measurements of temporal coherence of sound transmissions through shallow water," J. Acoust. Soc. Am. **120**, 2595–2614 (2006).
[19]After the completion of this work, the authors became aware of the paper: P. Hursky, M. B. Porter, M. Siderius, and V. K. McDonald, "Point-to-point underwater acoustic communications using spread-spectrum passive phase conjugation," J. Acoust. Soc. Am. **120**, 247–257 (2006), who applied the DPSK method to high SNR data. For high SNR cases, the method works like a coherent Rake receiver, the performance of which has been extensively investigated in the literature. The performance issues are significantly different for low SNR data as discussed in this article.

Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx)

Aijun Song[a] and Mohsen Badiey
College of Marine and Earth Studies, University of Delaware, 114 Robinson Hall, Newark, Delaware 19716

H. C. Song and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Michael B. Porter and the KauaiEx Group
Heat, Light, & Sound Research Inc., 3366 North Torrey Pines Court, Suite 310, La Jolla, California 92037

(Received 6 September 2007; revised 15 November 2007; accepted 24 November 2007)

During the July 2003 acoustic communications experiment conducted in 100 m deep water off the western side of Kauai, Hawaii, a 10 s binary phase shift keying signal with a symbol rate of 4 kilosymbol/s was transmitted every 30 min for 27 h from a bottom moored source at 12 kHz center frequency to a 16 element vertical array spanning the water column at about 3 km range. The communications signals are demodulated by time reversal multichannel combining followed by a single channel decision feedback equalizer using two subsets of array elements whose channel characteristics appear distinct: (1) top 10 and (2) bottom 4 elements. Due to rapid channel variations, continuous channel updates along with Doppler tracking are required prior to time reversal combining. This is especially true for the top 10 elements where the received acoustic field involves significant interaction with the dynamic ocean surface. The resulting communications performance in terms of output signal-to-noise ratio exhibits significant change over the 27 h transmission duration. This is particularly evident as the water column changes from well-mixed to a downward refracting environment. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2828055]

PACS number(s): 43.60.Dh, 43.60.Gk, 43.60.Fg, 43.60.Mn [EJS] Pages: 856–865

I. INTRODUCTION

Underwater acoustic channels are challenging for coherent digital communications because of severe multipath spread and limited bandwidth. Further, the variability of the ocean environment can cause fast fluctuations of acoustic channels and these fluctuations result in additional limitations on digital communications. Since the 1990s, various investigations on decision feedback equalizers (DFEs) and time reversal approaches have contributed to the advancement of coherent underwater acoustic communications. It has been shown that the DFE coupled with a phase tracking process presents a practical solution to the multi-path spread and fast phase fluctuation of underwater acoustic channels.[1] The DFE with joint phase tracking can be extended to multiple hydrophone channels to form a multichannel DFE.[2] In the multichannel DFE, the total number of adaptive feedforward taps increases with the number of channels. The implementation complexity of the multichannel DFE is high for a moderate to large number of hydrophone channels, which often are needed to achieve reliable performance in dynamic ocean environments. In order to alleviate the complexity issue, channel estimation can be incorporated into the DFE structure.[3]

Time reversal processing is able to achieve pulse compression for transmissions that have been spread by propagation through a multipath ocean environment. The time reversal concept was first demonstrated in the ocean in the 1960s.[4,5] Since the late 1990s, applications of the physics-based time reversal principle in underwater acoustic communications have shown success in at-sea experiments. Both active[6–8] and passive[9–11] time reversal methods have been investigated where in the latter, only one-way transmissions are used at a receiving array to implement the time reversal process. More recently, the time reversal approach has been combined with DFEs to improve the receiver performance while providing low implementation complexity.[12–15] In these studies, the acoustic channels usually are assumed time-invariant or slowly varying.

Over the last several years, a few studies have shown that correlation exists between high frequency acoustic fluctuations and environmental characteristics.[16–18] For example, the effects of tidally driven temperature fluctuations on underwater coherent acoustic communications have been studied at a carrier frequency of 18 kHz.[16] However, the relationship between environmental fluctuations and the performance of coherent underwater acoustic communications is not fully understood yet.

The Kauai experiment (KauaiEx) was conducted in June and July of 2003 to study high frequency (8–50 kHz) acoustic communications in 100 m deep water near Kauai, Hawaii.[19,20] During KauaiEx, extensive acoustic measurements were conducted while the ocean environment was monitored. The binary phase shift keying (BPSK) signals which were transmitted over an extended period (27 h) during KauaiEx are discussed in this paper. The communications signals are demodulated by time reversal multichannel

[a]Electronic mail: ajsong@udel.edu.

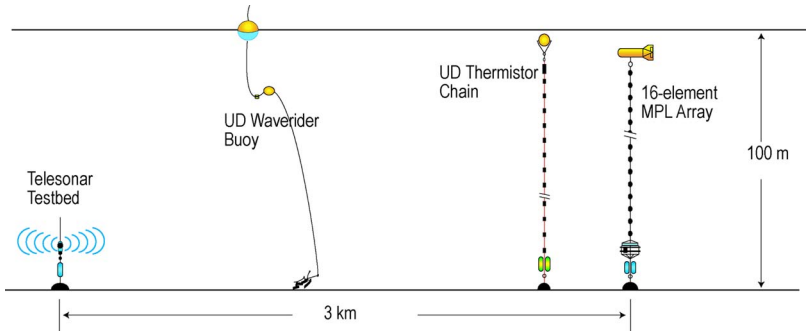


FIG. 1. (Color online) KauaiEx was conducted west of Kauai, HI, in 100 m deep water. The bottom-mounted Telesonar Testbed acoustic source was deployed 5 m above the sea bottom. At about 3 km range, the 16 element MPL autonomous receiving array spanned the entire water column. The wave-rider buoy was deployed about 1.2 km from the source. The thermistor chain shown was about 400 m from the MPL array.

combining followed by a single channel DFE. Acoustic channels in KauaiEx exhibited fast variations at a center frequency of 12 kHz in dynamic ocean environments. Therefore, continuous channel updates along with Doppler tracking are required prior to time reversal combining in order to track channel fluctuations. The performance of the communications receiver in terms of output signal-to-noise ratio (SNR) is reported over the entire 27 h transmission duration.

This paper is organized as follows. In Sec. II, a brief introduction to KauaiEx is presented. The receiver structure is presented in Sec. III. The acoustic channel and the BPSK performance are shown in Sec. IV. In the paper, a variable with superscript (i) denotes the value at the ith hydrophone. A variable with a caret denotes the estimate of the variable. c^* denotes the complex conjugate of a complex number c. $a(n)*b(n)$ denotes the convolution of two sequences $a(n)$ and $b(n)$. All time information regarding the experiment is in Greenwich Mean Time (GMT) if not otherwise stated.

II. KAUAIEX

KauaiEx was conducted from June 22 to July 9, 2003, west of Kauai, HI, in a shallow water waveguide.[19] The experiment data during July 2 through July 3, 2003, are of

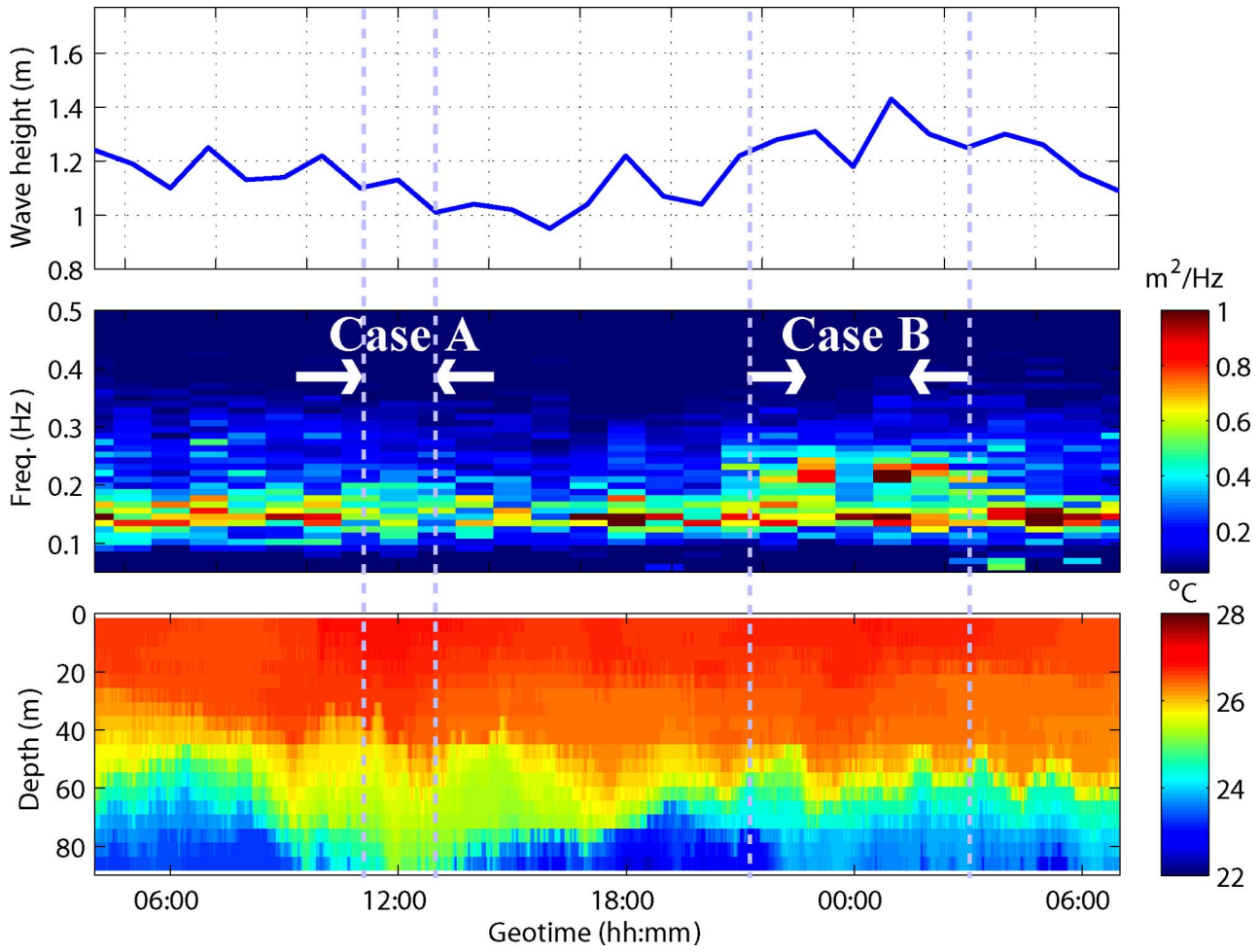


FIG. 2. (Color online) Environmental characteristics. Top panel: Significant wave height; Middle panel: Surface wave spectrum; Bottom panel: Temperature profile from 04:00 on July 2 to 07:00 on July 3, 2003, in KauaiEx. Note around 12:00 on July 2 (Case A), the water column was well mixed and the sea surface was relatively calm. From 21:00 on July 2 to 03:00 on July 3 (Case B), the water column was stratified and the sea surface was rougher.

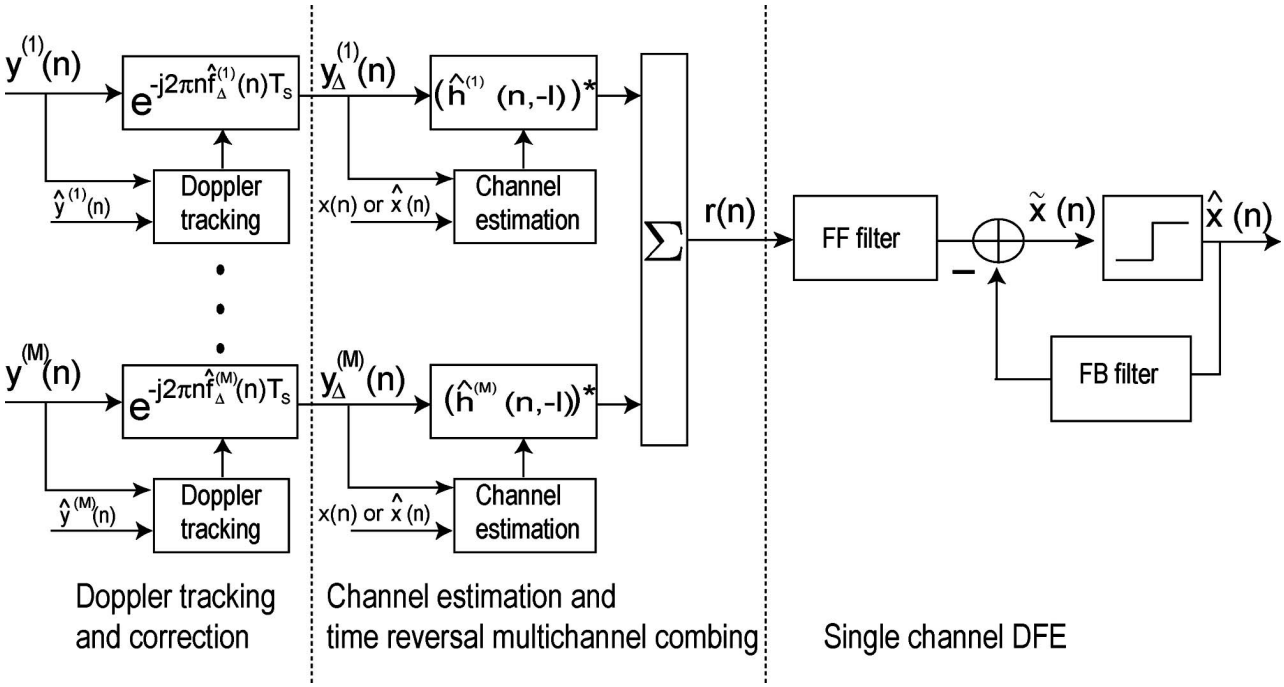


FIG. 3. The proposed receiver is composed of three parts: (1) Doppler tracking and correction, (2) channel estimation and time reversal multichannel combining, and (3) single channel DFE.

interest in this paper. As shown in Fig. 1, the water depth of the experimental site was 100 m. The Telesonar Testbed[21] was deployed 5 m above the sea bottom and served as the acoustic source. The source power level was 183 dB re 1 μPa at 1 m. At about 3 km range, a 16 element Marine Physical Laboratory (MPL) autonomous receiving array spanned the entire water column. The spacing of the hydrophones was 5 m. The top hydrophone was 16.5 m below the sea surface. Of a series of high frequency acoustic signals transmitted by the Telesonar Testbed, the 10 s BPSK signal will be used in the analysis. The carrier frequency of the BPSK signal is f_c=12 kHz and the symbol rate is R =4 kilosymbols/s. The square-root raised cosine shaping filter is used with an excess bandwidth[22] of 100%. The 10 s BPSK signal is referred to as a BPSK packet. The BPSK packet was transmitted and received every 30 min for 27 h from 04:00 on July 2 to 07:00 on July 3, 2003. In addition, 8–14 kHz, 50 ms long linear frequency modulated (LFM) chirps were transmitted a minute prior to every BPSK packet. The received waveforms on the MPL array were sampled at f_s=50 kHz.

Along with acoustic measurements, the surface wave spectrum and the two-dimensional water temperature profile were measured by a wave-rider buoy and a series of thermistor chains deployed along the propagation path. The wind speed underwent a late morning (Hawaii local time 10:00 a.m., e.g., 20:00 GMT) increase and a late night decrease as indicated by the sea surface wave spectrum in the middle panel of Fig. 2. The corresponding significant wave height varied from 1.4 to 1.0 m as shown in the top panel of Fig. 2. The water temperature profile shown in the bottom panel of Fig. 2 was measured by a thermistor chain deployed about 400 m from the MPL array. Note that for most of the time the water column was well mixed down to about 50 m depth. A cold layer (about 4–5 °C lower than the mixed layer) emerged at nearly tidal cycles. In addition to the communication results for the entire 27 h period, acoustic data during two contrasting environmental conditions marked as Case A and Case B in Fig. 2 will be discussed. Case A corresponds to around 12:00 on July 2 when the sea surface was relatively calm and the water column was well mixed. Case B is from 21:00 on July 2 to 03:00 on July 3 when the sea surface was slightly rougher and the water column was stratified. The main contrast between Cases A and B is the stratification of the water column. The significant wave height during Case A was about 1.1 m versus about 1.3 m during Case B.

III. THE RECEIVER STRUCTURE

Consider an underwater acoustic transmitter and receiver deployed in shallow water. At the source, a binary information sequence $x(n)$ is transformed into the baseband continuous wave $x(t)$. Then $x(t)$ is modulated onto the carrier frequency f_c and transmitted from a sound transducer. The receiver usually is equipped with multiple hydrophones. Let the total number of the hydrophones be M and $y^{(i)}(t)$ be the received baseband signal at the ith hydrophone. The effect of the transmission medium between the source and the ith hydrophone can be characterized by a time-varying channel impulse response (CIR) function, $h^{(i)}(t,\tau)$. The analog waveform $y^{(i)}(t)$ is sampled at a fractional symbol interval to provide robustness to carrier phase fluctuations in the underwater acoustic channel.[2,22] However, for notation convenience, symbol spaced signals are used throughout the paper. Therefore,

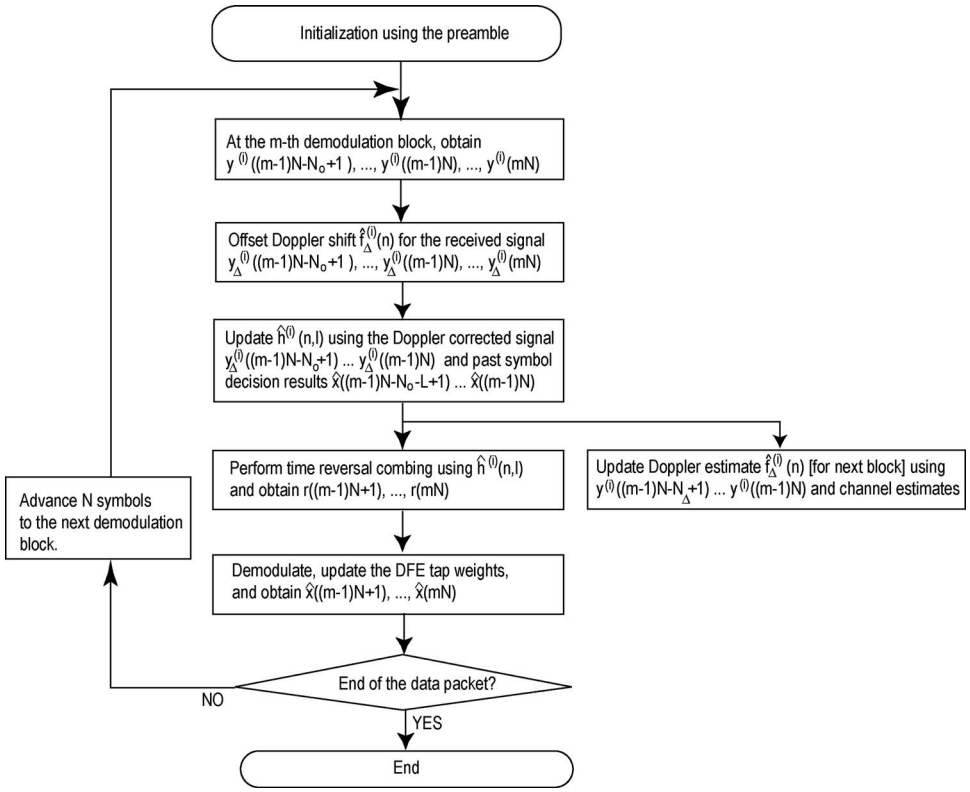


FIG. 4. The flow diagram of the receiver. Without loss of generality, $N_0 \geqslant N_\Delta$ is assumed.

$$y^{(i)}(n) = e^{j\theta^{(i)}(n)}[x(n) * h^{(i)}(n,l)] + v^{(i)}_{\text{amb}}(n), \tag{1}$$

where $y^{(i)}(n)$ is the discrete time representation of the analog signal $y^{(i)}(t)$, $\theta^{(i)}(n)$ is the instantaneous carrier phase offset, and $T_s = 1/R$ is the symbol duration. $v^{(i)}_{\text{amb}}(n)$ represents the ambient noise. $h^{(i)}(n,l)$, $0 \leqslant l \leqslant L-1$, is the discrete time baseband CIR function where L is the duration in symbols. $h^{(i)}(n,l)$ includes the combined effects of transmitter/receiver filters and the CIR function.

To recover the transmitted symbols which have been passed through the time-varying multi-path acoustic waveguide, a new multichannel receiver structure is proposed. Several features are incorporated into the receiver structure: (1) continuous Doppler tracking and correction is used to compensate for any observed linear trend in the carrier phase offset, (2) frequent channel estimation is used to track channel fluctuations,[3,23,24] and (3) compensation for residual phase fluctuations and intersymbol interference after time reversal combining is done using a DFE.[1] As shown in Fig. 3, the receiver consists of three parts: Doppler tracking and correction, channel estimation and time reversal multichannel combining, and single channel DFE.

The proposed receiver is a channel estimation based structure. At the beginning of a data packet, a preamble, or a sequence of known symbols, is used to perform initial channel and Doppler estimation and to train adaptively the DFE tap weights. After the preamble, channel and Doppler estimation are frequently updated. The most recent channel estimate on the ith channel is denoted by $\hat{h}^{(i)}(n,l)$ and the most recent Doppler estimate is denoted by $\hat{f}^{(i)}_\Delta(n)$. The three major parts of the receiver now will be discussed, followed by the overall implementation procedure.

A. Doppler tracking and correction

The Doppler estimate at the ith channel is obtained by

$$\hat{f}^{(i)}_\Delta(n) = \arg \max_{f^{(i)}_0 - (1/2)\delta f < f < f^{(i)}_0 + (1/2)\delta f} \left| \sum_{p=0}^{N_\Delta - 1} y^{(i)}(n-p)(\hat{y}^{(i)}(n-p)e^{j2\pi pfTs})^* \right|, \tag{2}$$

where $\hat{y}^{(i)}(n) = x(n) * \hat{h}^{(i)}(n,l)$ during the preamble and $\hat{y}^{(i)}(n) = \hat{x}(n) * \hat{h}^{(i)}(n,l)$ after the preamble. In Eq. (2), N_Δ is the Doppler observation block in symbols, $f^{(i)}_0$ is the coarse Doppler estimate and δf is the Doppler search range. Various Doppler estimation approaches exist in the literature, for example the ambiguity function method.[25]

At the beginning of the BPSK packets, $f^{(i)}_0$ is assumed to be zero as no *a priori* information is available and the Doppler shift should be obtained by searching over a large range of values. After the initial synchronization, as Doppler is estimated frequently, $f^{(i)}_0$ is set to the previous Doppler estimate and δf can be small. The Doppler correction is performed by offsetting the received signal $y^{(i)}(n)$ by the estimated Doppler shift, i.e., $y^{(i)}_\Delta(n) = y^{(i)}(n)e^{-j2\pi n \hat{f}^{(i)}_\Delta(n) T_s}$.

B. Channel estimation and time reversal multichannel combining

The channel estimate $\hat{h}^{(i)}(n,l)$ can be obtained from the Doppler corrected received signal $y^{(i)}_\Delta(n)$ and the previously detected symbols $\hat{x}(n)$ or the known symbols $x(n)$ during the preamble. Various least squares algorithms can be used for channel estimation. In this paper, the iterative least squares

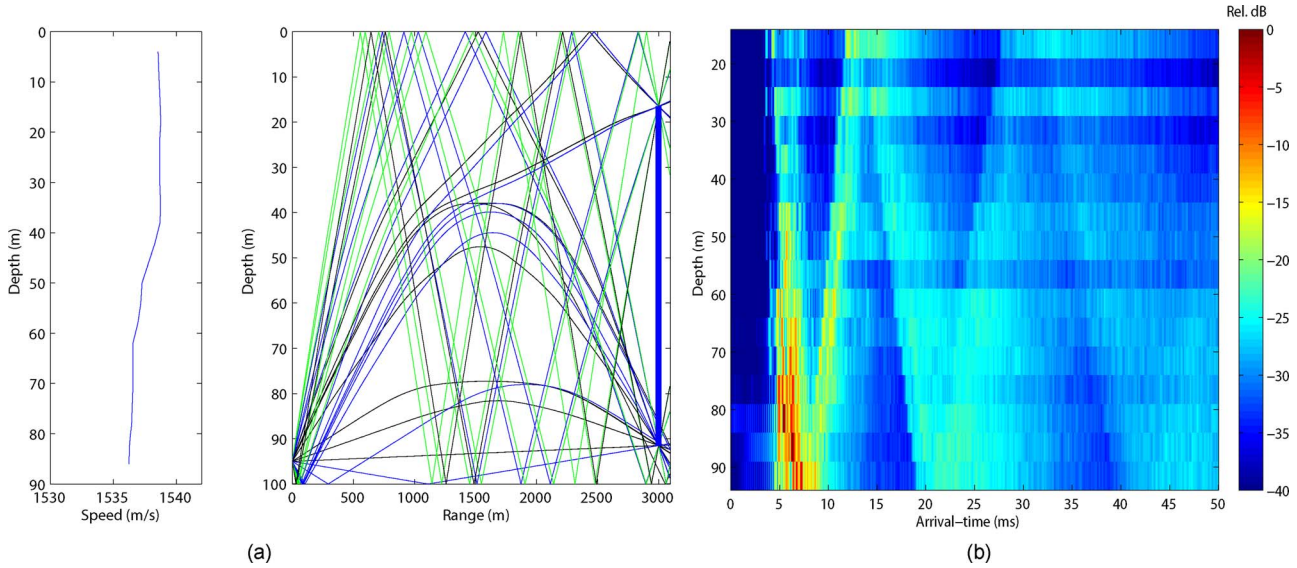


FIG. 5. (Color online) Acoustic propagation characteristics. (a) The sound speed profile (left-hand panel) and the ray diagram (right-hand panel) for the top and bottom hydrophones of the MPL array at 12:00 on July 2, 2003. (b) The CIR function obtained from the received LFM signals on the MPL array at 12:00 on July 2, 2003.

QR (LSQR) algorithm is used.[26] The channel estimation block size is chosen to be twice the channel length, i.e., $N_0 = 2L$.

Time reversal multichannel combining uses $(\hat{h}^{(i)}(n,-l))^*$ to match-filter the Doppler-corrected signals on each channel $y_\Delta^{(i)}(n)$ and then combines the results.[6,7,9] The output of time reversal combining is

$$r(n) = \sum_{i=1}^{M} (\hat{h}^{(i)}(n,-l))^* * y_\Delta^{(i)}(n) = x(n) * q(n,l) + w(n), \tag{3}$$

where $w(n)$ is the noise component,

$$w(n) = \sum_{i=1}^{M} (\hat{h}^{(i)}(n,-l))^* * (v_{\text{amb}}^{(i)}(n) e^{-j2\pi n \hat{f}_\Delta^{(i)}(n) T_s}), \tag{4}$$

and $q(n,l)$ is the effective CIR function, or the q-function,[10,11,13]

$$q(n,l) = \sum_{i=1}^{M} (\hat{h}^{(i)}(n,-l))^* * h^{(i)}(n,l), \tag{5}$$

assuming the Doppler correction completely removes the instantaneous carrier phase offset $\theta^{(i)}(n)$.

C. The single channel DFE

A single channel DFE with joint phase tracking[1] is used to equalize the residual intersymbol interference in $r(n)$. The exponentially weighted recursive least-squares (RLS) algorithm is used to update the equalizer tap weights. The residual carrier phase offset in $r(n)$ is compensated for by a second order phase locked loop (PLL) embedded in the adaptive channel equalizer. The phase correction based on the PLL output is implemented at the input to the DFE feedforward filter.

D. Implementation procedure

Figure 4 shows the implementation procedure for the proposed receiver. After the preamble at the beginning of each data packet, the receiver performs these tasks based on the previously detected symbols. Let the channel estimation update interval be N symbols with the receiver processing N symbols as a demodulation block each update. At the mth demodulation block, the receiver has obtained previously detected symbols through the $(m-1)$th demodulation block, i.e., $\hat{x}(n)$, $n \leqslant (m-1)N$, is known, and the objective is to recover the current N symbols $x(n)$, $n=(m-1)N+1,\ldots,mN-1,mN$. First, Doppler correction is made for the received signal at individual channels based on the most recent Doppler estimate $\hat{f}_\Delta^{(i)}(n)$. Subsequently, the channel estimate $\hat{h}^{(i)}(n,l)$ is updated using the Doppler corrected signals and the previously detected symbols. Note that Doppler tracking is performed again using the updated channel estimate for use in the next demodulation block. Then time reversal combining and equalization are conducted and the N symbol estimates $\hat{x}(n)$ are obtained. The algorithm advances to the next demodulation block if the end of the data packet has not been reached.

In the literature, existing DFE approaches include: (1) time reversal DFEs[12–14] and (2) multichannel DFEs developed by Stojanovic *et al.*[2,3] Although time reversal based, the proposed receiver has a different structure than the referenced time reversal DFEs where multichannel combining is performed based on channel probes or the known symbols at the beginning of the data packet. In the referenced time reversal DFEs,[12–14] phase tracking or Doppler tracking usually is performed after time reversal combining. Then an adaptive DFE is used to compensate for residual sidelobe structure in the q-function [Eq. (5)] and phase fluctuations. Compared with the referenced time reversal DFEs, the proposed receiver performs continuous Doppler tracking and channel estimation to combat fast fluctuations which occur over the

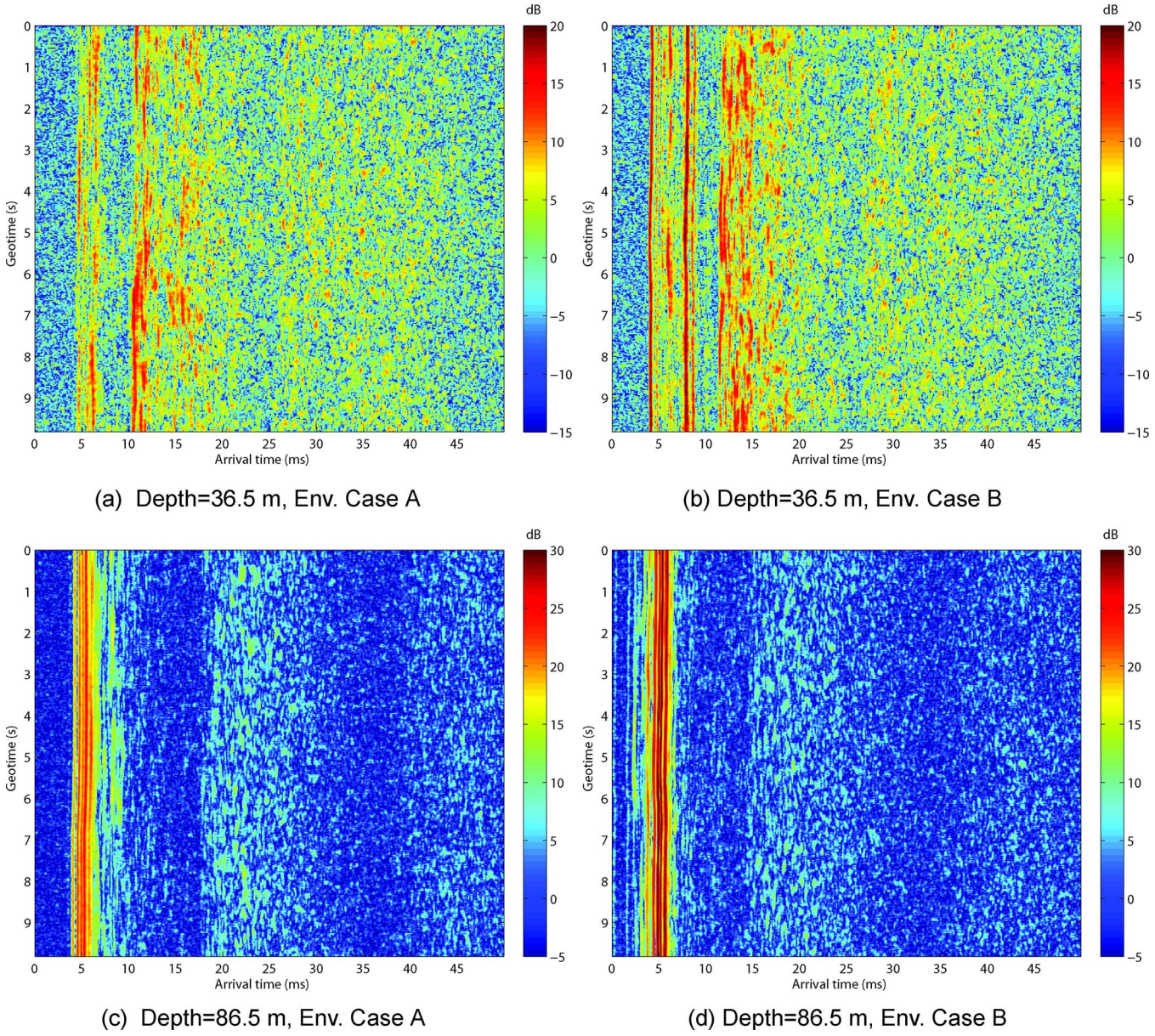


FIG. 6. (Color online) Estimated CIR functions at 36.5 m below the sea surface are shown (a) at 12:00 and (b) at 23:00 on July 2, 2003. Estimated CIR functions at 86.5 m below the sea surface are shown (c) at 12:00 and (d) at 23:00 on July 2, 2003. Note that the color scale of subplots (a) and (b) has a different dynamic range than that of subplots (c) and (d).

duration of a data packet. Note that time reversal combining alone with frequent channel updates previously has been discussed.[24] Joint time reversal combining and multichannel equalization also has been considered with emphasis on the use of low-complexity multichannel combining algorithms.[27]

In the multichannel DFEs developed by Stojanovic *et al.*, feedforward filters are applied to the individual channels and their outputs are combined prior to the feedback filter.[2,3] Phase synchronization at the individual channels is optimized jointly with the equalizer tap weights. The number of adaptive feedforward taps increases with the number of channels. Compared with the multichannel DFEs developed by Stojanovic *et al.*, the proposed receiver uses a single channel DFE after time reversal combining.

An advantage of the proposed receiver structure is its low complexity.[2] The complexity of a multichannel DFE increases at least as the square of the number of channels if RLS algorithms are used for a fast tracking capability.[22] Since time reversal combining collapses multiple channels into a single channel, the complexity of the successive DFE remains unchanged when the number of channels increases. The complexity of the channel estimators increases linearly with the number of channels. It also increases linearly with the number of taps estimated for each channel if a fast least squares algorithm, such as the LSQR algorithm, is employed.

To measure the receiver performance, the SNR at the soft output $\tilde{x}(n)$, denoted by ρ_{out}, is used. In the next section, the output SNR of the receiver is shown for the 27 h period during KauaiEx.

TABLE I. Receiver parameters.

Parameters	Description	Value
f_s	Sampling rate	50 kHz
f_c	Carrier frequency	12 kHz
R	Symbol rate	4 kHz
f_{EB}	Excess bandwidth of the square-root raised cosine filter	4 kHz
K	Oversampling factor	3
M	Total number of the channels	4 or 10
N_{preamble}	Size of the preamble	1000 symbols
L	Length of the CIR function	200 symbols
N_0	Channel estimation block size	400 symbols
N	Channel estimation update interval	100 symbols
N_Δ	Doppler observation block size	400 symbols
δf_0	Initial Doppler search range	10 Hz
δf	Doppler search range after the preamble	1.6 Hz
N_{ff}	Feedforward filter span in symbols	15 symbols
N_{fb}	Feedback filter tap number	8 symbols
K_{f_1}	Proportional tracking constant in PLL	0.0002
K_{f_2}	Integral tracking constant in PLL	0.0002
λ	RLS forgetting factor in the DEF	0.999

IV. THE RECEIVER PERFORMANCE IN KAUAIEX

A. Acoustic channels on the MPL array

Due to the large aperture and deployment range of the MPL array, the CIR functions at the top and at the bottom of the array show different characteristics. For example, Fig. 5(b) shows the CIR function across the MPL array at 12:00 on July 2, 2003. The CIR function is obtained from the received LFM signals. Based on the ray diagram generated by the BELLHOP model[28] with the downward refracting sound speed profile in Fig. 5(a), it can be seen that the CIR function at the top hydrophone is composed of direct (D), bottom (B), surface (S), B-S, S-B, and S-B-S paths, etc. Most ray paths have surface interaction. In contrast, the CIR function on the bottom hydrophone consists primarily of multiple bottom interacting arrivals.

As a consequence of the different propagation paths, the CIR functions in the upper water column have different energy levels and temporal coherence properties from those in the lower water column. Figure 6 shows a comparison of the CIR functions in the upper and lower water column for two data packets during two different environmental conditions. The CIR functions are obtained by the channel estimator in the receiver. Note that the color scale of subplots (a) and (b) represents a dynamic range from −15 to 20 dB, whereas that of subplots (c) and (d) represents a dynamic range from −5 to 30 dB. The energy of the CIR function in the lower water column is much higher than that in the upper water column. The input SNR is 12.4 dB at 36.5 m depth while it is 18.5 dB at 86.5 m depth. When environmental Case A changes to Case B, the CIR functions both in the upper and in the lower water column change. At 23:00 on July 2 during environmental Case B, the CIR functions in Figs. 6(b) and 6(d) show stronger ray paths that do not interact with the sea surface. The input SNR at 36.5 m depth increases to 16.0 dB and that at 86.5 m depth increases to 22.6 dB.

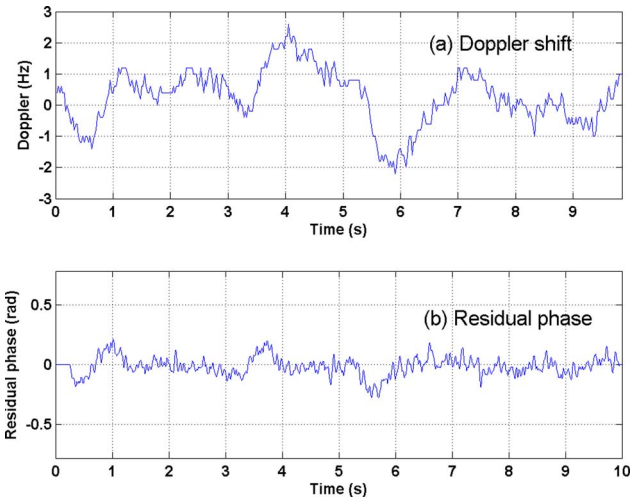


FIG. 7. (Color online) Doppler shift and phase fluctuations. (a) The Doppler estimate at 36.5 m below the sea surface and (b) the residual phase estimated by the PLL in the DFE on the MPL-TOP array at 23:00 on July 2, 2003.

To compare the performance of the communications data simultaneously recorded in the upper and the lower water column, the top 10 hydrophones of the MPL array (MPL-TOP) and the bottom 4 hydrophones of the MPL array (MPL-BTM) are considered as sub-arrays in the analysis because the CIR functions at these two sets of hydrophones show a similar arrival structure among themselves.

B. The receiver performance

To compare the receiver performance on the MPL-TOP and MPL-BTM arrays in different environments, a uniform set of receiver parameters are chosen as in Table I. As mentioned, the element data are oversampled in the receiver and the oversampling rate is $K=3$. The number of the feedforward taps is KN_{ff} for the fractionally spaced DFE[1] where N_{ff} is the feedforward filter span in symbols. The number of the feedback taps is N_{fb} because the feedback filter is applied to a symbol spaced sequence, i.e., previously detected symbols.

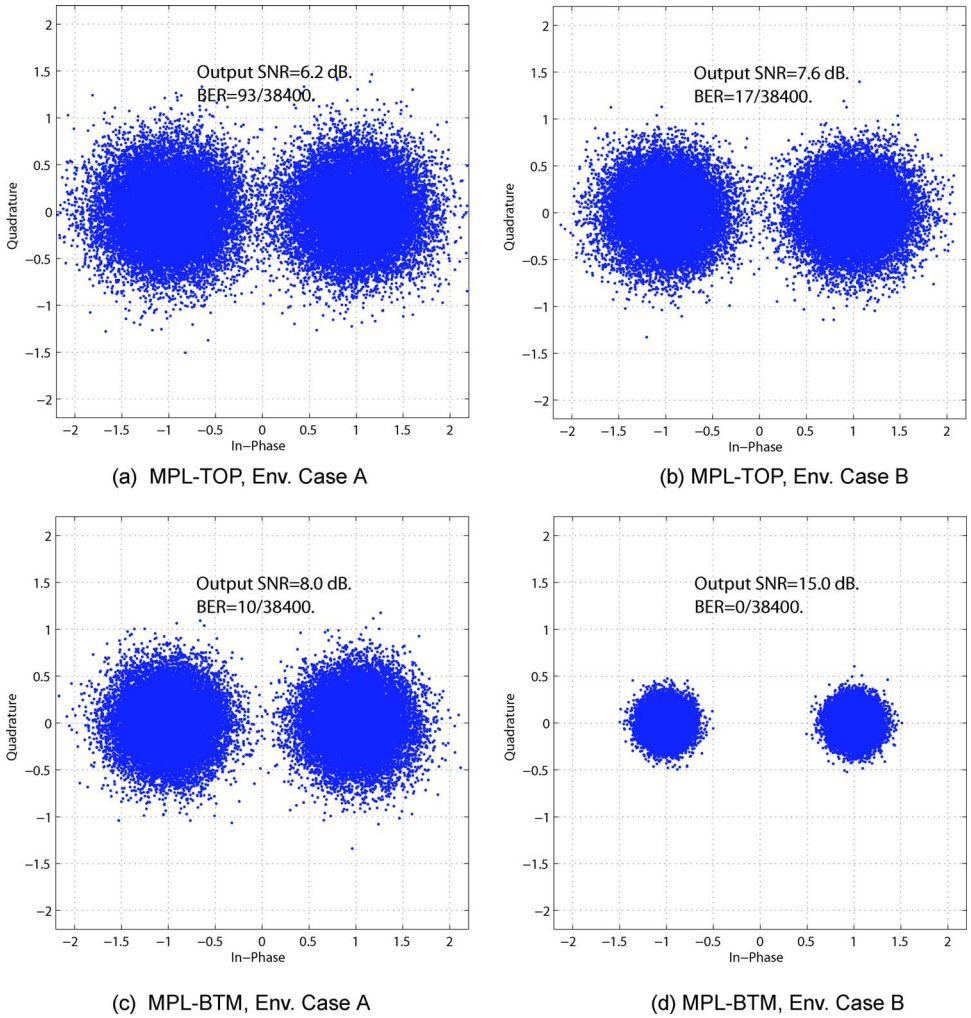


FIG. 8. (Color online) Scatter plots of the soft output $\tilde{x}(n)$ for the MPL-TOP array are shown (a) at 12:00 and (b) at 23:00 on July 2, 2003. Scatter plots of the soft output $\tilde{x}(n)$ for the MPL-BTM array are shown (c) at 12:00 and (d) at 23:00 on July 2, 2003.

The receiver parameters are optimized against fast channel fluctuations on the MPL-TOP array. For example, the channel estimation update interval is chosen as $N=100$ symbols, which is necessary for the MPL-TOP array receiver. In other words, channel and Doppler estimation is performed every 25 ms. As shown in Figs. 6(a) and 6(b), the CIR functions on the MPL-TOP array show significant fluctuations. Increasing N to 200 or 400 symbols deteriorates receiver performance for the MPL-TOP array while it does not affect receiver performance for the MPL-BTM array.

Frequent Doppler tracking and correction is also found necessary. As an example, Fig. 7(a) shows the observed time-varying Doppler shift (linear trend in the carrier phase offset) at a MPL-TOP array element during a single packet. Figure 7(b) shows the residual phase estimated by the PLL in the DFE when the Doppler tracking and correction is performed every 25 ms. Because of explicit Doppler tracking, the residual phase is small as shown. If the Doppler tracking and correction is only performed at the beginning of the packet, the receiver fails to track the channel and to demodulate the symbol sequence during 24 packets on the MPL-TOP array. If the Doppler tracking and correction is performed every 1 s, still there are 11 packets where the receiver fails to demodulate on the MPL-TOP array.

At the beginning of the 10 s BPSK packet, $N_{\text{preamble}}=1000$ symbols are used to carry out initial channel estimation, Doppler tracking, and DFE tap weight training. During the preamble, the initial Doppler search range is 10 Hz. As the Doppler shift is being tracked, the Doppler search range after the preamble is set relatively small, 1.6 Hz. The search step is 0.2 Hz. Therefore, the complexity introduced by Doppler tracking is very limited. The RLS forgetting factor λ in the DFE is chosen as 0.999.

Figure 8 shows the receiver performance for the MPL-TOP and MPL-BTM arrays during the two environmental cases. When the environment changed from Case A to Case B, the MPL-TOP array receiver performance improves slightly as shown in Figs. 8(a) and 8(b). The output SNR at 12:00 on July 2 during environmental Case A is 6.2 dB, whereas that at 23:00 on July 2 during environmental Case B is 8 dB. In contrast, the MPL-BTM array receiver performance improves significantly between environmental Case A and Case B with the output SNR increasing 7.0 dB.

Figure 9 shows the output SNR for the MPL-TOP and MPL-BTM arrays for the entire 27 h recording period. As shown in Fig. 2, the environmental characteristics changed significantly over this period as did receiver performance. For the MPL-BTM array, the average output SNR increases 5.8 dB from environmental Case A to environmental Case B. In contrast, for the MPL-TOP array, the average output SNR experiences a much smaller increase (1.8 dB) between these

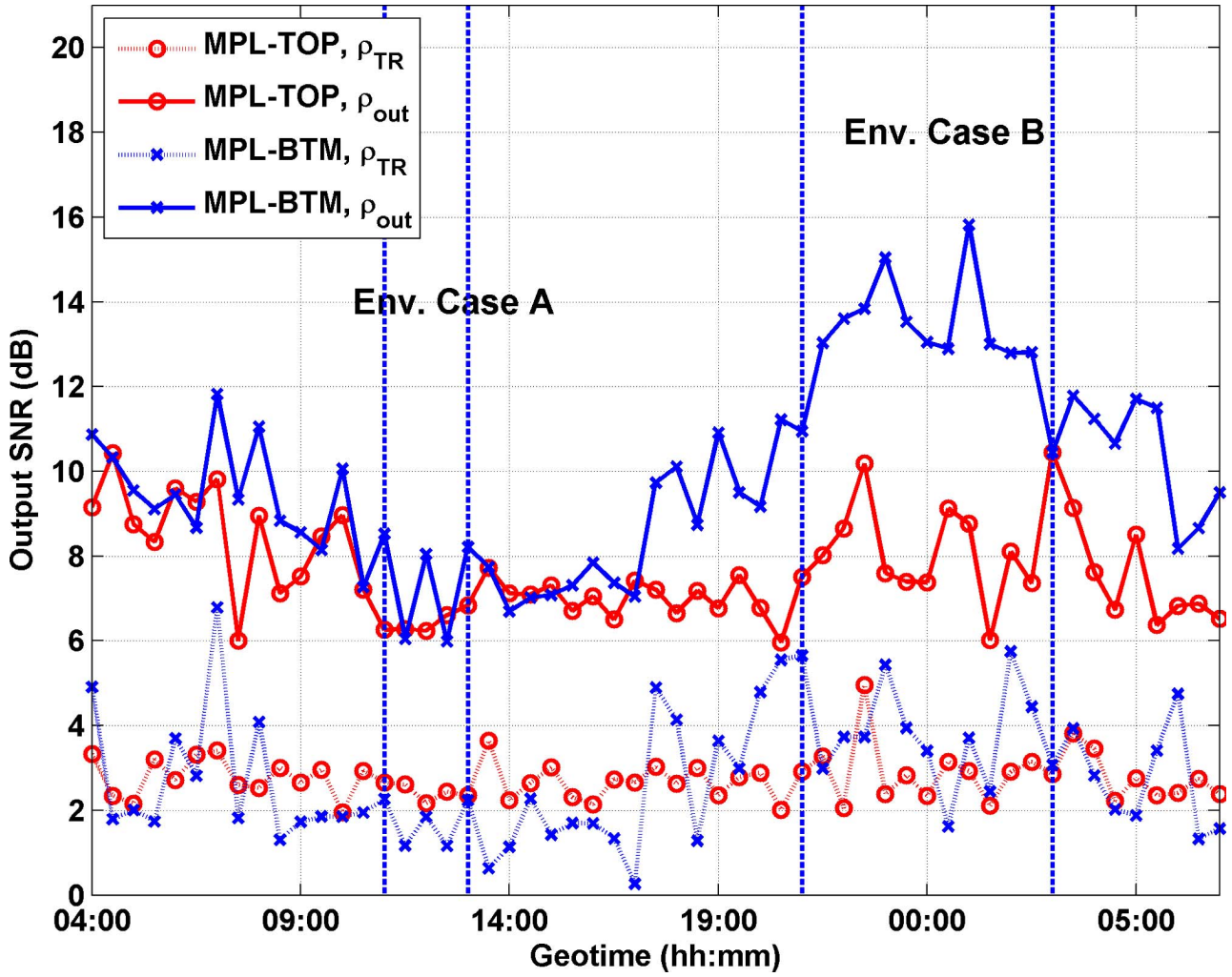


FIG. 9. (Color online) The output SNR for the MPL-TOP and MPL-BTM arrays from 04:00 on July 2 to 07:00 on July 3, 2003. For the MPL-BTM array, the average output SNR increases 5.8 dB during the change from environmental Case A to environmental Case B. In contrast, for the MPL-TOP array, the average output SNR experiences a much smaller increase (1.8 dB) when the environment changes from Case A to Case B. The intermediate SNR ρ_{TR} at the output of time reversal combining also is shown. Improvements using a DFE after time reversal combining are shown clearly.

cases. The primary distinction between these periods is the well-mixed water column during environmental Case A and the downward refracting sound speed profile during environmental Case B.

The output $r(n)$ after time reversal combining already provides an estimate of the transmitted symbols.[24] The SNR of $r(n)$ after phase correction[24] is defined as ρ_{TR}. Through a comparison between ρ_{TR} and ρ_{out}, the benefit of channel equalization after time reversal combining can be seen. As shown in Fig. 9, postprocessing $r(n)$ with a DFE on the MPL-TOP and MPL-BTM arrays can improve the performance for all BPSK packets. For the MPL-TOP array, the improvement is 4.9 dB for the entire MPL array recording period and the improvement is roughly uniform along geotime. In contrast, for the MPL-BTM array, the average improvement through use of a DFE is about 7.1 dB and the improvement varies during different environmental conditions. During environmental Case B, the improvement is as high as 9.3 dB.

It is worthwhile to note that although the water column and sea surface conditions vary during the 27 h period, the channel can be tracked and the bit error rate is below 10^{-2} for all 55 packets. For 9 packets on the MPL-TOP array and 29 packets on the MPL-BTM array, there are no demodulation errors.

As shown in Fig. 9, even with ten hydrophones, the receiver with the MPL-TOP array usually has inferior performance to that with the MPL-BTM array, which only has four hydrophones. As the water column sound speed environment changes from well-mixed to downward refracting, the MPL-BTM array is more strongly insonified.[29] In addition, on the MPL-TOP array, most acoustic arrivals have sea surface interaction so that these arrivals are weaker and have shorter temporal coherence.

V. CONCLUSIONS

During KauaiEx, 27 h of high frequency acoustic communications measurements centered at 12 kHz and environmental observations were made in 100 m depth water near Kauai, Hawaii. The acoustic channels in KauaiEx exhibited challenging features including severe multi-path spread and

fast fluctuations. To overcome these difficulties, the BPSK data with a symbol rate of 4 kilosymbols/s were demodulated by time reversal multichannel combining followed by a single channel DFE. Continuous channel updates along with Doppler tracking on the individual channels were required prior to time reversal combining.

The proposed receiver was applied to the communications data obtained for two subsets of array elements whose channel characteristics appear distinct: (1) top 10 and (2) bottom 4 elements of a 16 element array spanning the entire water column. With the aide of the 1000 symbol preamble, all BPSK packets can be demodulated successfully using both sets of array elements. The resulting communications performance in terms of output SNR exhibited significant change over the 27 h transmission duration. This is particularly evident as the water column changed from well-mixed to a downward refracting environment.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research (ONR) Code 321OA. The KauaiEx Group consists of Michael B. Porter, Paul Hursky, Martin Siderius (Heat, Light, & Sound Research Inc.), Mohsen Badiey (University of Delaware), Jerald Caruthers (University of Southern Mississippi), William S. Hodgkiss, Kaustubha Raghukumar (Scripps Institution of Oceanography), Daniel Rouseff, Warren Fox (University of Washington), Christian de Moustier, Brian Calder, Barbara J. Kraft (University of New Hampshire), Vincent McDonald (SPAWAR), Peter Stein, James K. Lewis, and Subramaniam Rajan (Scientific Solutions Inc.).

[1]M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Phase-coherent digital communications for underwater acoustic channels," IEEE J. Ocean. Eng. **19**, 100–111 (1994).

[2]M. Stojanovic, J. Catipovic, and J. G. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communications," J. Acoust. Soc. Am. **94**, 1621–1631 (1993).

[3]M. Stojanovic, L. Freitag, and M. Johnson, "Channel-estimation-based adaptive equalization of underwater acoustic signals," *Oceans* (IEEE, New York, 1999), Vol. **2**, pp. 590–595.

[4]A. Parvulescu and C. S. Clay, "Reproducibility of signal transmissions in the ocean," Radio Electron. Eng. **29**, 223–228 (1965).

[5]A. Parvulescu, "Matched-signal ("MESS") processing by the ocean," J. Acoust. Soc. Am. **98**, 943–960 (1995).

[6]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, P. Gerstoft, P. Roux, T. Akal, C. Ferla, and D. R. Jackson, "Ocean acoustic time reversal mirror," Proceedings of the Fourth European Conference Underwater Acoustics, 1998, pp. 493–498.

[7]G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, "An initial demonstration of underwater acoustic communications using time reversal," IEEE J. Ocean. Eng. **31**, 602–609 (2002).

[8]H. C. Song, P. Roux, W. S. Hodgkiss, W. A. Kuperman, T. Akal, and M. Stevenson, "Multiple-input/multiple-output coherent time reversal communications in a shallow water acoustic channel," IEEE J. Ocean. Eng. **31**, 170–178 (2006).

[9]D. Rouseff, D. R. Jackson, W. L. J. Fox, C. D. Jones, J. A. Ritcey, and D. R. Dowling, "Underwater acoustic communication by passive-phase conjugation: Theory and experimental results," IEEE J. Ocean. Eng. **26**, 821–831 (2001).

[10]T. C. Yang, "Temporal resolutions of time-reversal and passive-phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**, 229–245 (2003).

[11]T. C. Yang, "Differences between passive-phase conjugation and decision-feedback equalizer for underwater acoustic communications," IEEE J. Ocean. Eng. **29**, 472–487 (2004).

[12]G. F. Edelmann, H. C. Song, S. Kim, W. S. Hodgkiss, W. A. Kuperman, and T. Akal, "Underwater acoustic communications using time reversal," IEEE J. Ocean. Eng. **30**, 852–864 (2005).

[13]T. C. Yang, "Correlation-based decision-feedback equalizer for underwater acoustic communications," IEEE J. Ocean. Eng. **30**, 865–880 (2005).

[14]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, M. Stevenson, and T. Akal, "Improvement of time reversal communications using adaptive channel equalizers," IEEE J. Ocean. Eng. **31**, 487–496 (2006).

[15]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, W. J. Higley, K. Raghukumar, and T. Akal, "Spatial diversity in passive time reversal communications," J. Acoust. Soc. Am. **120**, 2067–2076 (2006).

[16]N. M. Carbone and W. S. Hodgkiss, "Effects of tidally driven temperature fluctuations on shallow-water acoustic communications at 18 kHz," IEEE J. Ocean. Eng. **25**, 84–94 (2000).

[17]M. Badiey, Y. Mu, J. A. Simmen, and S. E. Forsythe, "Signal variability in shallow-water sound channels," IEEE J. Ocean. Eng. **25**, 492–500 (2000).

[18]T. C. Yang, "Measurements of temporal coherence of sound transmissions through shallow water," J. Acoust. Soc. Am. **120**, 2595–2614 (2006).

[19]M. B. Porter, P. Hursky, M. Siderius, M. Badiey, J. Caruthers, W. S. Hodgkiss, K. Raghukumar, D. Rouseff, W. Fox, C. de Moustier, B. Calder, B. J. Kraft, K. McDonald, P. Stein, J. K. Lewis, and S. Rajan, "The Kauai experiment," *High Frequency Ocean Acoustics* (AIP, New York, 2004), pp. 307–321.

[20]M. Siderius, M. B. Porter, P. Hursky, V. McDonald, and the KauaiEx Group, "Effects of ocean thermocline variability on noncoherent underwater acoustic communications," J. Acoust. Soc. Am. **121**, 1895–1908 (2007).

[21]V. K. McDonald, P. Hursky, and the KauaiEx Group, "Telesonar testbed instrument provides a flexible platform for acoustic propagation and communication research in the 8–50 kHz band," High Frequency Ocean Acoustics (AIP, New York, 2004), pp. 336–349.

[22]J. G. Proakis, *Digital Communications*, 4th ed. (McGraw-Hill, New York, 2000).

[23]J. C. Preisig, "Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment," J. Acoust. Soc. Am. **118**, 263–278 (2005).

[24]J. A. Flynn, J. A. Ritcey, D. Rouseff, and W. L. J. Fox, "Multichannel equalization by decision-directed passive phase conjugation: Experimental results," IEEE J. Ocean. Eng. **29**, 824–836 (2004).

[25]M. Johnson, L. Freitag, and M. Stojanovic, "Improved Doppler tracking and correction for underwater acoustic communications," *ICASSP'97* (IEEE, New York, 1997), Vol. **1**, pp. 575–578.

[26]C. C. Paige, "Fast numerically stable computations for generalized linear least squares problems," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **16**, 165–171 (1979).

[27]J. Gomes, A. Silva, and S. Jesus, "Joint passive time reversal and multichannel equalization for underwater communications," *OCEANS'06* (IEEE, New York, 2006).

[28]M. B. Porter and Y. C. Liu, "Finite-element ray tracing," Proceedings of the International Conference on Theoretical Computational Acoustics, 1995, Vol. **2**, pp. 947–956.

[29]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 2000).

Green's function estimation in speckle using the decomposition of the time reversal operator: Application to aberration correction in medical imaging

Jean-Luc Robert
Philips Research North America, 345 Scarborough Road, Briarcliff Manor, New York 10510

Mathias Fink
Laboratoire Ondes et Acoustique, ESPCI, Université Paris 7, CNRS, 10 rue Vauquelin, 75005 Paris, France

(Received 15 May 2007; revised 12 October 2007; accepted 26 October 2007)

The FDORT method (French acronym for decomposition of the time reversal operator using focused beams) is a time reversal based method that can detect point scatterers in a heterogeneous medium and extract their Green's function. It is particularly useful when focusing in a heterogeneous medium. This paper generalizes the theory of the FDORT method to random media (speckle), and shows that it is possible to extract Green's functions from the speckle signal using this method. Therefore it is possible to achieve a good focusing even if no point scatterers are present. Moreover, a link is made between FDORT and the Van Cittert–Zernike theorem. It is deduced from this interpretation that the normalized first eigenvalue of the focused time reversal operator is a well-known focusing criterion. The concept of an equivalent virtual object is introduced that allows the random problem to be replaced by an equivalent deterministic problem and leads to an intuitive understanding of FDORT in speckle. Applications to aberration correction are presented. The reduction of the variance of the Green's function estimate is discussed. Finally, it is shown that the method works well in the presence of strong interfering scatterers.
 [DOI: 10.1121/1.2816562]

PACS number(s): 43.60.Fg, 43.80.Qf, 43.20.Fn [TDM] Pages: 866–877

I. INTRODUCTION

The resolution and signal-to-noise ratio of ultrasound images rely on the ability to focus the field at any point in the image. In order to focus properly at a point, its Green's function, expressed in the imaging array, must be known. This is straightforward to compute in a homogeneous medium if the speed of sound is known, but it can be challenging in heterogeneous media, where different layers can have different speeds of sound. For example, in medical ultrasound, the image can be distorted by the presence of a subcutaneous layer of fat where the speed is different from the other tissue. Adaptive methods are then required to estimate the Green's functions. Other than focusing, knowledge of Green's functions can be useful to extract some parameters from the medium, such as speed of sound. In this paper, the term Green's function of a point refers to the signal received by the array when a source located at the point transmits a pulse. In the time domain, it is a wave front, and in the frequency domain it is a complex function. To focus on the point, one has to transmit the time reversed version of its Green's function. For simplicity, this will also be referred to as the Green's function.

Iteration of the time reversal process[1,2] has been shown to lead to the Green's function of the brightest scatterer in the medium. The DORT (French acronym for decomposition of the time reversal operator) method[3,4] is a generalization of this process and has proven efficient to extract the Green's functions of point scatterers if they are well resolved. At a given temporal frequency ω, the so-called *transfer matrix* $K(\omega)$ is built, where $K_{ij}(\omega)$ is the signal received by array element j after a pulse has been transmitted by element i. Practically $K_{ij}(\omega)$ is obtained by taking the Fourier transform of the broadband signal received by array element i. We omit the index ω in the following. The time reversal operator is defined as KK^H, the product of the matrix by its Hermitian transpose. The DORT method consists of performing an eigenvalue decomposition of KK^H (in practice the singular value decomposition of K is computed, which is equivalent). The eigenvectors are the invariants of the time reversal operator, and in the case of a medium made of well-resolved point scatterers, where each eigenvector corresponds to the Green's function of a scatterer.[3] The corresponding eigenvalue is related to the reflectivity of the scatterer.

However, in a physical medium, like the human body, there are actually very few *well resolved* point scatterers. Most of the signal is speckle. Speckle is due to the presence of a large number of subresolution scatterers. The observed signal results from the interferences between the wave fronts from all the insonified scatterers and has a random aspect. Therefore the speckle has to be dealt with in terms of statistics. As speckle is more abundant than well resolved point scatterers, it is helpful to be able to extract Green's functions of the medium from speckle signals.

The focused DORT (FDORT) method is an implementation of the DORT method using focused transmits instead of single elements transmits.[5] For point scatterers, both methods give equivalent results. In addition FDORT can be used locally, as the focused transmits are localized in space, and has the ability to separate scatterers in range. FDORT is

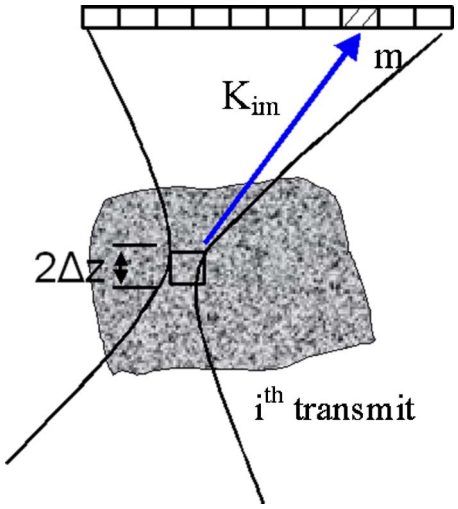


FIG. 1. (Color online) Acquisition of the transfer matrix in the FDORT method. The focused transmit i insonify the medium. The signal received by array element m is time gated to select the signal from depth $z-\Delta z$ to $z+\Delta z$, and its Fourier coefficient at frequence ω gives the matrix coefficient K_{im}. In speckle Δz is taken to be about the pulse length.

implemented in the following way (Fig. 1): the signal resulting from the insonification by the ith focused transmit is recorded by element m, the time gated to select the signal coming from the desired range, for example from depth $z-\Delta z$ to $z+\Delta z$, and Fourier transformed to give a coefficient $K_{\text{im}}(\omega)$ of the focused transfer matrix K. The size of K is N_{el} (number of elements)$^*I$ (number of transmits). Then the singular value decomposition of K is performed as in the DORT method. In the following K stands for the transfer matrix of the FDORT method.

In this paper, we show that FDORT can also be used to extract Green's functions from the pure speckle signal. In this case, we select a signal coming from around the focal depth. Although the original FDORT algorithm gave acceptable results in speckle,[6] a slight modification of the algorithm is better suited for estimation in speckle. This will be described and justified in Sec. II B.

In Sec. II, we review the physical interpretation of KK^H both for point-like scatterers and speckle. For point-like scatterers, it is interpreted as the time reversal operator between the physical array and a virtual array whose elements would be located at the transmit focal depth. For random speckle signal, KK^H can be interpreted as a spatial correlation matrix described by the Van Cittert–Zernike theorem.[7] This links the method to the work of others[8,9] and yields an interesting interpretation of the first eigenvalue as a focusing criterion. With speckle, a third interpretation is possible as a time reversal operator for an equivalent virtual object, which exhibits the physical meaning of the eigenvectors. The variance of the estimate of the Green's function is also discussed.

In Sec. III we will see that in the presence of aberration, the FDORT method in speckle needs to be iterated. It usually converges to the Green's function after a few iterations. Experimental results in a phantom are also presented.

Finally, in Sec. IV, it is shown that the method has the ability to separate the signal of interest from interference, and thus yields a better estimation of the Green's functions than a cross-correlation method.

As this paper makes a link between DORT and random signal processing (e.g., the VanCittert–Zernike theorem), notations of both fields are used, so that the reader can easily refer to the original papers. Thus, the signal received by array element m for the ith transmit (Fig. 1) is K_{im} in DORT notations[3,5] which highlights the fact that it is a coefficient of the transfer matrix K, and S_m^i in statistical notation, which highlights the fact that it is the ith realization of the signal received by element m.

II. INTERPRETATIONS OF KK^H

A. Deterministic objects: time reversal operator

With coherent objects, such as point scatterers, KK^H can be interpreted as a time reversal operator. Robert *et al.*[5] showed that mathematically KK^H was the time reversal operator between two arrays, the physical array corresponding to the per-element receive and a virtual array accounting for the fact that focused transmits are used. In fact, physical meaning can be given to this virtual array: each focused transmit can be replaced by a virtual element located at the focal spot of the transmit, and with a certain directivity. Indeed, at ranges deeper than the focal depth, the transmit wave front can be seen as a wave diverging from the virtual element, as if it had been emitted by the virtual element, and at depths shallower than the focus, the wave front is converging to the virtual element and can be considered as the time reversed version of a wave emitted by the element.

The time reversal operator interpretation of KK^H is fundamental, as it exhibits the physical meaning of the eigenvectors and eigenvalues using a thought experiment of time reversal iteration.[3,4] Indeed, the Green's functions of well resolved scatterers are invariants of the time reversal process. When one transmits a Green's function, only the corresponding scatterer is insonified, and the echo received is the time reversed version of the transmitted Green's function. One can time reverse the received signal and iterate the process, and still get the same signal. Mathematically, the eigenvectors of the time reversal operator are the invariants of time reversal and therefore the Green's function.

The difference between FDORT and DORT, though, is that with focused beams, the problem can be considered from the point of view of the actual array, or from the point of view of the virtual array. Practically, the singular value decomposition of K gives $K=USV^H$, where U would contains the Green functions expressed in the actual array (also termed *canonical basis*) and V would contains the Green functions expressed in the virtual array.[5]

Speckle is formed from a very large number of subresolved scatterers. In this case, the time reversal interpretation is still valid, but not helpful. Indeed a thought experiment of iterative time reversal no longer helps in predicting the behavior of the eigenvectors, nor their focusing properties, as many scatterers are insonified at the same time.

B. Spatial correlation matrix, or Van Cittert Zernike matrix

In speckle, many subresolution scatterers are insonified at the same time, and the resulting backscattered signal is the

sum of many coherent signals with a random phase (a random walk), which yields a random signal. The speckle can only be rigorously described using statistics, thus KK^H has to be interpreted statistically.

1. Interpretation of the transmits as realizations of the random backscattered signal

The volume of scatterers insonified by each transmit is limited in depth (from $z-\Delta z$ to $z+\Delta z$) by the time gating, as seen in Fig. 1. In speckle, we select volumes around the focal depth. Thus the beam pattern can be approximated by the beam pattern at the focal depth, where we write $p(x)$. For a homogeneous medium and a rectangular aperture, $p(x)$ is a sinc function. Each of the N transmits has the same beam shape, but insonifies a different set of scatterers and therefore gives a different realization of the random backscattered signal. This is equivalent to the problem where we fire N times the same transmit (instead of using N consecutive transmits) but with a different random medium each time. In our case however, the transmits are translated (or rotated, depending on the scheme used) versions of each other and consequently the insonified volumes are in slightly different locations. Better results are obtained if this is taken into account by a modification of the FDORT algorithm. Indeed, the echoes resulting from a focal spot at depth Z and azimuth $x=0$ typically have a geometrical delay (resulting from the propagation in an homogeneous medium with sound speed c) proportional to

$$\frac{1}{c}\sqrt{Z^2+X^2} \propto \left(Z+\frac{X^2}{2Z}\right)\frac{1}{c}, \tag{1}$$

where X is the coordinate in the array plane, and using the Fresnel approximation. For another transmit whose focal spot is at azimuth x_i, we have

$$\frac{1}{c}\sqrt{Z^2+(X+x_i)^2} \propto \left(Z+\frac{1}{2Z}(X^2+x_i^2+2X\cdot x_i)\right)\frac{1}{c}. \tag{2}$$

In order to have signals that really look like two realizations of the same signals, it is necessary to remove the part depending on x_i in the wave front curvature. This is done by time delaying the echo for the ith transmitted by

$$\left(\frac{1}{2}x_i^2+X\cdot x_i\right)\frac{1}{cZ}. \tag{3}$$

This is described in Fig. 2. As a result, it looks like the two realizations come from the same location, but that it is the medium that has moved between the two insonifications. Alternatively, the signal can be completely aligned. In this case, the complete geometrical curvature is removed using the time delays given by Eq. (2). This is interesting in the case where there is no interest in the whole Green's function, and only the perturbation due to heterogeneities is of interest, as in phase aberration correction.

2. Link between KK^H and the spatial correlation matrix

We have seen that each transmit i can be interpreted as a different realization of the random backscattered signal.

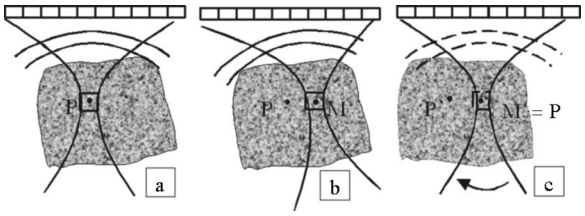


FIG. 2. (a) A transmit focusing at a point P and the corresponding received wave front. In speckle, ideally, to estimate the Green function at point P, we would like to fire several times the same transmit focusing at P, but with a different speckle distribution each time, to provide a good averaging of the randomness. (b) Instead, we use a neighbor beam focusing at M close to P (the distance between M and P is exaggerated here). (c) By properly shifting the wave front of the received signal, it looks like the whole medium is virtually translated and M', the new position of M, corresponds to P. Thus a new realization of the signal coming from P is obtained, with a different scatterer distribution. By virtually translating the phantom by a different distance, several realizations are obtained. This is valid only if the aberration seen by M is the same as the aberration seen by **P**.

Now, K_{im} is the ith realization of the backscattered signal received by the mth array element that we now denote S_m^i to highlight the statistical nature of the signal, and make the link with the VanCittert–Zernike paper.[7] We are now giving a statistical interpretation of the coefficients of KK^H as estimates of the cross correlation of the signals received by different elements.

Let I be the number of transmits. Developing the product KK^H yields for the coefficient (m,n) of KK^H:

$$(KK^H)_{m,n}=\sum_{i=1}^{I} S_m^i S_n^{i*} = I\left(\frac{1}{I}\sum_{m=1}^{M} S_m^i S_n^{i*}\right). \tag{4}$$

This is I times the average, over the I realizations of the speckle of the product of the signals received by element m and by element n. $(KK^H)_{m,n}$ is then an estimate of $\langle S_m S_n^* \rangle$, the cross correlation of the signals received on elements m and n, at frequency ω. Therefore KK^H is an estimate of the spatial correlation matrix, R_{SS} whose coefficients are $\langle S_m S_n^* \rangle$.

The bracket symbols stands for the expected value, which would be obtained by averaging on an infinite number of realizations. The term cross spectrum is sometimes used instead of cross correlation,[10] to highlight the fact that it is built in the temporal frequency domain. Rigorously, it is a cross spectrum regarding the temporal domain, but a cross correlation regarding to the spatial domain. Therefore, the term spatial cross correlation will be used in the following.

There is a difference between R_{SS} and KK^H, since the latter is only an estimate of R_{SS} derived from a limited number of realizations (the transmits). In particular, the estimate KK^H has a variance that depends on the number of transmits. It is only one possible estimate of R_{SS}, and most of the results derived here would apply to other estimates of R_{SS}.

In previously published work,[8,9] the first eigenvector of the correlation matrix R_{SS} was used to improve the focusing. In principle this is similar to our method. The novelty of this research is both practical and theoretical: we propose to use the FDORT matrix as a particular estimate of R_{SS}. This provides a very practical implementation of the method, which yields good experimental results. On the theoretical side,

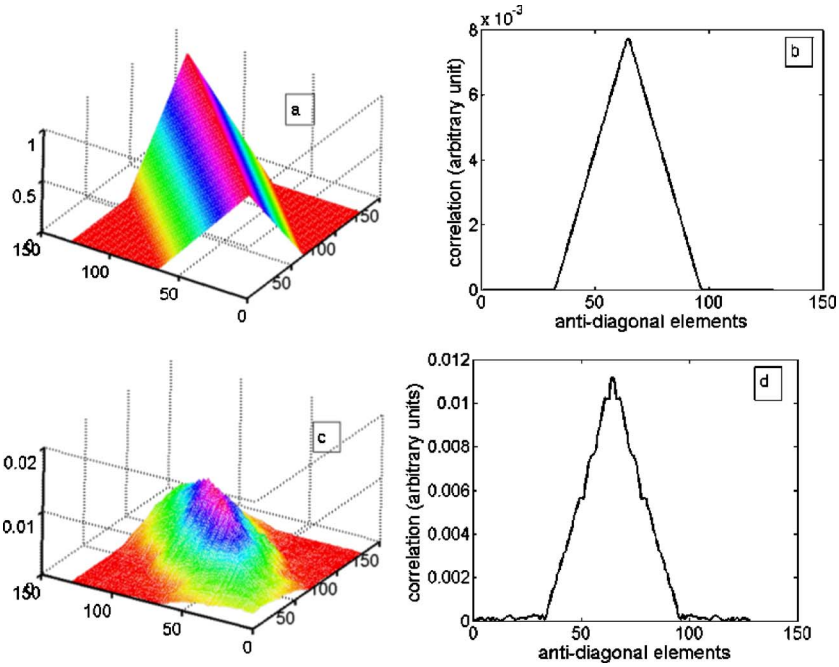


FIG. 3. (Color online) Theoretical amplitude of the spatial correlation matrix (a) and projection on the antidiagonal (b) in homogeneous media as predicted by the VanCittert–Zernike theorem. The triangular shape that arises from the Fourier transform of sinc^2 is easily identifiable. sixtyfour of the 128 elements were used in transmission, which is why the triangle base is 64 elements wide. (c), (d) Experimental amplitude of the FDORT matrix, for a 60 mm focus. The differences observed for the edge elements are mainly due to the transducers directivity, which was not taken into account in the VanCittert–Zernike prediction.

new insights are given; in particular the interpretation of the first eigenvalue as a focusing criterion is novel.

In previous work[10] an estimate of the coefficients of R_{SS} are computed in a similar fashion, but the coefficients are used in a different algorithm.

3. Link to the VanCittert–Zernike theorem

In speckle, the FDORT matrix is an estimate of the spatial correlation matrix, which contains the spatial cross correlation for every pair of array elements. Following the Van Cittert–Zernike theorem,[7] the spatial cross correlation for a pair of elements (m,n) only depends on the distance between the elements, $X_m - X_n$, and is proportional to the Fourier transform of the square of $p(x)$:

$$\langle S_m S_n^* \rangle = \exp\left(j\left(\phi^{Ab}(X_m) - \phi^{Ab}(X_n) + \frac{2\pi}{\lambda Z}(X_m^2 - X_n^2) \right) \right) \cdot \beta \cdot P\left(\frac{X_m - X_n}{\lambda Z} \right), \quad (5)$$

where

$$P(k) = FT[|p(x)|^2]. \quad (6)$$

β is a real constant, $2\pi/\lambda Z(X_m^2 - X_n^2)$ is the difference of geometric curvature between the two elements, that is the difference of propagation paths in a homogeneous medium, and $\phi^{Ab}(X_m) - \phi^{Ab}(X_n)$ is the differential aberration phase between elements m and n, that is due to the presence of inhomogeneities in the medium. In the usual near-field screen approximation, where the aberration introduces a mere time delay τ_m on the element m, $\phi^{Ab}(X_m) = \omega\tau_m$. However, the complex frequency-dependent representation allows more general aberrations to be handled.

The amplitude of the spatial correlation is particularly interesting. In a homogeneous medium, $p(x)$ is a sinc function, and therefore its Fourier transform is a triangle function. The amplitude of the cross correlation for a pair of elements decreases as a triangle function when the distance between elements increases. This is the pattern one can observe when the amplitude of the KK^H matrix is plotted (Fig. 3).

In inhomogeneous media, the aberration usually leads to a broader focus, and therefore the spatial correlation function decreases faster.

In conclusion, for speckle, KK^H is an estimate of the spatial correlation matrix—or Van Cittert–Zernike matrix—that is fully described by the Van Cittert–Zernike theorem. This interpretation of the matrix will be helpful when interpreting the eigenvalues in Sec. II E. It is also interesting because the correlation matrix is an important tool in random signal processing, and is the basis of numerous algorithms.[11]

This interpretation of the FDORT matrix as a correlation matrix is not surprising. For deterministic scatterers, it has already been shown[12,13] that the time reversal operator can be interpreted as a correlation matrix of the received signal.

C. Time reversal operator for an equivalent virtual object

As stated in the previous section, in speckle, it is not helpful, *a priori*, to interpret KK^H as a time reversal operator directly. However, we can play a trick and interpret it as the canonical time reversal operator for an equivalent deterministic object.

Everything happens as if we were performing a time reversal experiment on an object whose reflectivity is proportional to $p(x)^2$, the beam intensity. Indeed, it is shown in the following that for such an experiment, and under the Fraunhoffer approximation, the canonical transfer matrix that we can call K_{eq} has the same coefficients as the spatial correlation matrix described in Sec. II B. Intuitively, it is easy to see where this virtual object comes from. In fact, the speckle can be considered as a random mirror[7] that *reflects* an image of the transmit beam $p(x)$ times the random scatterers distribu-

tion. If only one transmit, or realization, is available, this image is very blurred. However, by averaging on several realizations, the random mirror is smoothed, and the object is revealed.

Now, a rigorous derivation is provided: Let us imagine that we have such an object with reflectivity $p(x)^2$ and let us derive the coefficient of the K_{eq} matrix in the Fraunhoffer approximation. A transmit by element m gives rise to a plane wave $\exp[j(2\pi/\lambda Z)x\cdot X_m]$ (in the Frauhoffer aproximation, spherical waves are approximated by plane waves). The field reflected by the object is then, at depth Z,

$$p_{\mathrm{ref}}(x) = p(x)^2 \exp\left(j\frac{2\pi}{\lambda Z}x\cdot X_m\right)$$

and the field received by array element n is

$$K_{\mathrm{eq}_{m,n}} = \int p(x)^2 \exp\left(j\frac{2\pi}{\lambda Z}x\cdot (X_m + X_n)\right)dx$$
$$= P\left(\frac{X_m + X_n}{\lambda Z}\right).$$

The last equality is obtained using the definition of the Fourier transform and introducing $P(k)=FT[|p(x)|^2]$. Noting that $X_{Nel+1-n}=-X_n$ which results from the fact that the array is centered on abscissa 0, and thus the element $N_{\mathrm{el}}+1-n$ is symmetrical to the element n, and using Eq. (5) yields $Rss_{m,n}=K_{\mathrm{eq}_{m,N_{\mathrm{el}}+1-n}}$. K_{eq} is deduced from R_{SS} by a column flip. It can be shown that in this case the eigenvectors of $K_{\mathrm{eq}}K_{\mathrm{eq}}^H$ are identical to the eigenvectors of R_{SS}. Thus the eigenvectors for the focused time reversal operator in speckle are the same as the eigenvectors for a time reversal operator for a deterministic object that has the shape of the focused transmit.

Thus from a mathematical standpoint, everything happens as if the focused transmit was creating a virtual scatterer at the focal spot and the DORT method was performed on the virtual point scatterer. The first eigenvector is then the Green's function of the focal spot. As the focal spot is not a perfect point, but has a finite size given by the resolution of the array, there is not a single eigenvector, but several, as observed with extended objects.[14] If the focal spot is sufficiently small, the first eigenvector can still be considered to be the Green's function of the focal point. It is therefore important to use the whole aperture in transmit to keep the focal spot as small as possible.

D. Variance and standard deviation of the estimation

We have determined asymptotic expressions and properties for the FDORT matrix in speckle. However, in a random medium like speckle, one can only talk in statistical terms. The asymptotic properties have to be considered as mean values, obtained only if one could average on an infinite number of realizations of the signal. The estimated coefficients of the matrix, and in turn the eigenvectors, and the estimated Green's functions, will in practice fluctuate from this mean value. The random fluctuation is characterized by a standard deviation, defined as the square root of the mean square of the fluctuation. The term variance is usually used in the literature[15,16] and refers to the square of the standard deviation. The standard deviation seems to have more physical meaning though, as the quality of focusing depends on the standard deviation of the phase of the Green's function rather than its variance. The term standard deviation is adopted in the following. The results given here apply to the standard deviation of most estimated quantities in this paper (amplitude and phase of the estimated Green's functions, coefficients of $KK^H\ldots$).

Expressions of the variance for the coefficients of the spatial correlation matrix can be found in Masoy *et al.*[9,15] The main result is that the standard deviation is proportional to the square root of the number of independent realizations (and also to the coherence of the signals). Experimentally, this result seems to also hold for the eigenvectors of KK^H, although in this case no theoretical expression is known by the authors.

In our case, the different realizations are given by the different transmits. The homogeneous transmit beam pattern is $p(x)=\mathrm{sinc}(\pi xD/\lambda Z)$, where Z is the focal depth, λ is the wavelength, and D is the aperture size. It can be shown that two transmits yield independent (or noncorrelated) realizations if they are separated by a distance equal to the width of the transmit resolution cell $\lambda Z/D$.[17] Therefore the number of realizations that one can extract from a part of the medium of the width L is $LD/(\lambda Z)$. L is usually limited because we assume (Sec. II B) that the curvature of the wave fronts coming from all the locations that we use differs only by a linear term that we remove before averaging [Eqs. (1)–(3)]. It makes sense that if we want to estimate the Green's function at a specific location, only the region of the medium with the same Green's function can be used. Removing the linear term enables us to use the region where the Fresnel approximation holds. However, in most cases, we are interested in estimating the Green's function in the inhomogeneous medium. In this case, the Green's function can vary quickly from one location to another. The region where we can still make the assumption that the Green's function differs only by a linear term is called the isoplanatic patch. In medical ultrasound, a typical isoplanatic patch in the breast is 1 mm laterally and 2 mm axially.[18] Therefore, the number of independent realizations that one can take laterally is given by the lateral size of the isoplanatic patch divided by $\lambda Z/D$, which is typically 0.4 mm. Consequently, only three or four independent realizations can be taken laterally.

As a limited number of realizations can be taken laterally, the standard deviation can be further reduced by taking realizations from different depths. As shown in Fig. 4, signals from a few depths surrounding the focal depths can be selected by time gating. Two different windows in depth yield independent realizations if they are separated by about $pulselength\cdot c$, where c is the sound speed, and *pulse length* is the pulse duration, in seconds.

In our simulations, four realizations were selected laterally, and six axially which yielded 24 realizations, and a standard deviation reduction by a factor 5. This was enough to obtain a good estimate of the Green's functions in both simulations and experimental results. With a two-dimensional (2D) array,[10,19] and 3D imaging, the number of

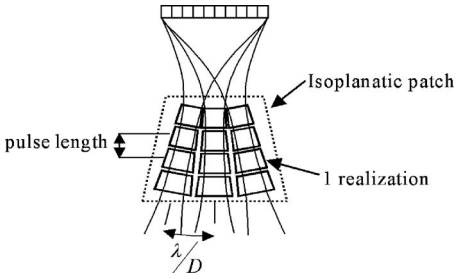


FIG. 4. In speckle, the standard deviation of the estimation is inversely proportional to the square root of the number of realizations used in the estimation. However, the number of lateral realizations is limited by the size of the isoplanatic patch, which is the size of the region where all the points *see* the same aberration, and thus where the Green functions are similar enough to be averaged. Therefore, to increase the number of realizations, additional windows can be selected using the depth dimension.

realizations can be increased by taking additional realizations in the elevation dimension (the third dimension). In Refs. 10, 16, and 20, 70 realizations are typically taken in a 3D volume.

To summarize, for 2D imaging in speckle the standard deviation of the estimate decreases with the square root of the number of realizations, and the number of realizations is given by the area of the isoplanatic patch divided by the area of the resolution cell $(\lambda Z/D)\cdot pulse\ length\cdot c$. In 3D, the number of realizations would be given by the volume of the isoplanatic patch divided by the volume of the resolution cell. In Fig. 4, where the scale has been exaggerated, it is obvious that the area of the resolution cell increases with depth. In practice, the isoplanatic patch is much smaller than the focal depth Z, and to a very good approximation we can consider that all the resolution cells in the patch have a width equal to $\lambda Z/D$.

The number of resolution cells in a volume is a good measure of the quantity of information one can extract from the volume. This is also equivalent to the *space-bandwidth* product, a well-known measure of information quantity in engineering, which is equal to the interrogated area times the area of the spectrum (in k-space[20]) of the system. The area in k-space can be approximated by the product of the axial bandwidth B/c, where B is the temporal bandwidth, and the lateral bandwidth $D/\lambda Z$. Therefore, the space-bandwidth product is $\text{Area}_{\text{Isoplanatic\_patch}}\cdot(B/c)\cdot(D/\lambda Z)$.

One could be tempted to increase the transmit density (by reducing the distance between consecutive transmits) in order to have more realizations and decrease the standard deviation further. However, the additional realizations would not be independent from the initial realizations. Therefore, they do not bring additional information. They only bring redundancy. Doing this would slow the acquisition process, as more transmits are required, but would not reduce further the standard deviation. In simulations, the standard deviation using a high density of transmits separated by 0.01 mm has been compared to the standard deviation in the case where the transmits were separated by $\lambda Z/D=0.4$ mm. The results are plotted in Fig. 5. Despite the fact that the transmits are 40 times as numerous in the first case, resulting in an acquisition time 40 times slower, the standard deviation reduction is as good in the second case.

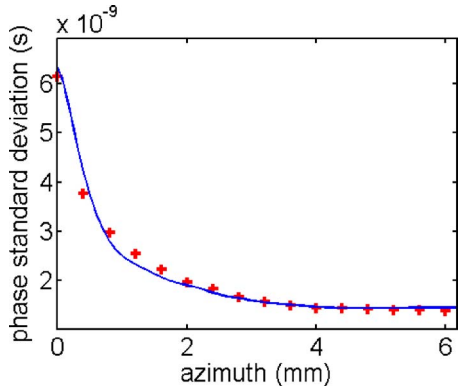


FIG. 5. (Color online) Standard deviation of the phase in function of the lateral size of the speckle region. In the first numerical experiment (solid blue plot), the transmits are very close to each other (0.01 mm) while in a second one (red cross), the transmits used in the estimation were separated by the beam width (0.4 mm). Taking more transmits than required by the condition to have independent realizations does not improve the estimation.

E. Interpretation of the first eigenvalue in speckle

1. Interest of the focusing criterion

We use the statistical interpretation developed in Sec. II B. to show that the first eigenvalue in speckle has an important interpretation: if it is properly normalized, it is the focusing criterion C introduced by Mallart and Fink.[7] C is equal to the ratio of the coherent intensity and incoherent intensity. The focusing criterion is an objective measure of the quality of focusing. It is a number that can vary from 0 to 1, the upper bound being reached only for signals from point scatterers. In speckle, its value depends on how well the signals received are correlated,[19] which in turn (from the VanCittert–Zernike theorem) depends on the quality of focusing. In a homogeneous medium (perfect focusing), the spatial correlation function is a triangle whose base is equal to the array length, and C is equal to 2/3 in this case.[7] However this is not exactly the maximum value that C can reach in speckle, as it is for a square apodization. With other apodizations higher values of C can be reached, and values close to 0.75 have been observed during simulations. Other parameters like the transducer's directivity can influence C. C drops quickly when the focusing degrades, and a good rules of thumb is to consider that the focusing is acceptable if C is above 0.5.

An explicit relationship between C and the coherence length of the signal can be derived. The spatial correlation function for a backscattered signal of coherence length L can be roughly approximated by a triangle of base L. It results that C would be $2/3(L/D)$ in this case, where D is the full array width. When the base of the triangle was equal to D, C was 2/3. The coefficient 2/3 in this case is not really important as it depends on the shape of the correlation function (2/3 is for a triangle), but the main result is the proportionality between C and L/D. The focusing criterion is basically inversely proportional to the number of coherent cells in the array.

As noted earlier, the focusing criterion is equal to the ratio of the coherent intensity and incoherent intensity. Now we need to express this quantity in term of the spatial correlation matrix.

2. Link between first eigenvalue of R_{SS} and coherent intensity

Varslot *et al.*[8] demonstrated that the first eigenvector of the spatial correlation matrix gives the receive focal law and apodization that maximizes the speckle brightness and can therefore be used to correct an aberration. Indeed, let V be the $N_{el}\times 1$ complex vector used to beamform the signal in receive. The phase term of V is a focal law, and the amplitude term is an apodization law. V is usually the Green's function of the desired focal point. The beamformed amplitude, for the ith realization (ith transmit) is then

$$A^i = \sum_{n=1}^{N_{el}} S_n^i V_n^* = V^H S^i,$$

where S^i is a $N_{el}\times 1$ vector whose coefficients are the S_n^i. Now, the intensity is

$$I^i = A^i A^{i*} = V^H S^i (V^H S^i)^H = V^H S^i S^{iH} V,$$

where $S^i S^{iH}$forms a matrix whose element (m,n) is $S_m^i S_n^{i*}$. Finally the speckle brightness, or expected coherent intensity, is

$$\langle I_c \rangle = V^H \langle SS^H \rangle V.$$

$\langle SS^H \rangle$ is a matrix whose element (m,n) is $\langle S_m S_n^* \rangle$, and therefore it is the spatial correlation matrix R_{SS}. We have therefore an expression of the expected coherent intensity in terms of the spatial correlation matrix and of a vector V used to beamform the signal. $\langle I_c \rangle = V^H R_{SS} V$. An estimate of the coherent intensity averaged on the few transmits used in FDORT is therefore $V^H K K^H V$.

It is well known[21,22] that the coherent intensity or speckle brightness increases with the quality of focusing. Therefore the best focusing is obtained by maximizing the coherent intensity. Now, let us show that the normalized vector $\mathbf{V}$ that maximizes the coherent intensity is the first eigenvector of the spatial correlation matrix.

Indeed let $\mathbf{e_i}$ be the ith eigenvector of R_{SS}. The eigenvectors form an orthonormal basis, where one can decompose a normalized vector $\mathbf{V}$ in that basis $\mathbf{V} = \Sigma \langle \mathbf{V} | \mathbf{e_i} \rangle \mathbf{e_i}$ and therefore

$$\mathbf{V}^H R_{SS} \mathbf{V} = \sum \langle \mathbf{V} | \mathbf{e} \rangle_\mathbf{i}^2 \lambda_i \leq \sum \langle \mathbf{V} | \mathbf{e_i} \rangle^2 \lambda_1 = \lambda_1,$$

where λ_i is the ith eigenvalue. We used the fact that the norm of $\mathbf{V}$ is 1. The inequality is reached when $\mathbf{V} = \mathbf{e_1}$.

Consequently, the first eigenvector of the spatial correlation matrix maximizes the coherent intensity, and hence the focusing (this confirms the result we derived using time reversal on the equivalent object in Sec. II C) and the first eigenvalue is the corresponding coherent intensity. In practice, estimates of these parameters are given by the first eigenvector and eigenvalue of KK^H.

The coherent intensity is proportional to the focusing quality. However, it is only a relative criterion. Looking at the first eigenvalue by itself does not give information about the focusing quality, as it depends on many parameters like the transmitted power, the scatterer reflectivity, etc. The focusing criterion C is a more interesting since it is an absolute measure.

3. Link between the incoherent intensity and the sum of the eigenvalues

The expected incoherent intensity is the sum of the intensity received by each element

$$\langle I_{Inc} \rangle = \sum_{i=1}^{N_{el}} \langle |S_i|^2 \rangle = \sum_{i=1}^{N_{el}} R_{SS}(i,i) = \mathrm{Tr}(R_{SS}),$$

where Tr is the trace of a matrix, that is the sum of its diagonal elements. The Trace is conserved when the matrix is expressed in another basis. In the eigenvector basis, the matrix is diagonal, and the diagonal elements are the eigenvalues. Thus the incoherent intensity is also equal to the sum of all eigenvalues

$$\langle I_{Inc} \rangle = \sum_{i=1}^{N_{el}} \lambda_i.$$

The incoherent intensity can therefore be estimated by the sum of the eigenvalues of KK^H.

4. Link between C and the normalized first eigenvalue

The focusing criterion C is the ratio of the coherent intensity over the incoherent intensity. Thus, an estimate of C is given by

$$C_1 = \frac{\lambda_1}{\mathrm{Tr}(KK^H)} = \frac{\lambda_1}{\sum_{n=1}^{N_{el}} \lambda_i}.$$

The notation C_1 is used to highlight the fact that it is the focusing criterion obtained when the first eigenvector $\mathbf{e_1}$ is used to beamform the received signal.

Alternatively, a normalized version of the spatial correlation matrix can be built

$$\overline{\overline{KK^H}} = \frac{KK^H}{\mathrm{Tr}(KK^H)}.$$

This matrix has the same eigenvector as $\mathbf{KK^H}$ as it just differs by a scaling factor, but its first eigenvalue is directly C_1.

Thus, the decomposition of the FDORT matrix not only provides an estimate of the Green's function, $\mathbf{e_1}$, but also provides a direct measure of the quality of the focusing associated with this estimate, or in other words, a measure of the quality of the image if it is used for aberration correction. For example, a first normalized eigenvalue above 0.5 means that the estimate of the Green's function is good. A poor value of the first normalized eigenvalue means that an iteration is needed (this will be elaborated on in Sec. III). With other aberration correction methods, such a criterion has to be computed separately.[19]

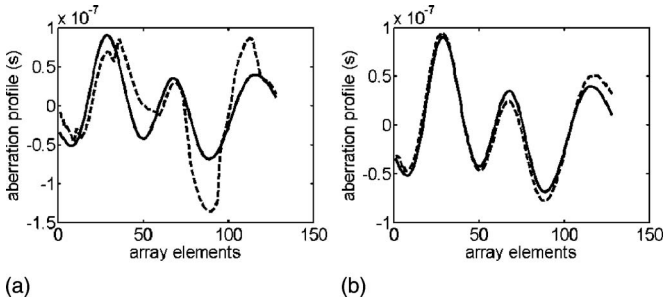


FIG. 6. The phase of the estimated Green functions is unwrapped and the geometrical delay removed to show an estimate of the aberrator delay profile. The estimate (dash line) is compared with the true profile (solid line), after the first iteration (a) and after five iterations (b). The estimate is not very good at the first iteration but converges after a few iterations.

III. APPLICATION TO FOCUSING IN HETEROGENEOUS MEDIA

A. Equivalent virtual object and iteration of the method

It has been shown in Sec. II C that performing FDORT in speckle is equivalent to performing conventional DORT on an equivalent virtual object which has the shape of the transmit beam pattern. In homogeneous media, a virtual scatterer is created at the focal spot, and the first eigenvector of the FDORT matrix is the Green's function of the focal spot.

However, in the presence of phase aberration, the transmit is no longer well focused and the virtual object is more complex. In order to illustrate the effects of phase aberration, an FDORT experiment in speckle has been simulated using Field II.[23] A 1D linear array of 128 elements at 7.3 MHz central frequency is simulated. It focuses at 60 mm depth through a near-field phase screen. All the elements are used in transmit in order to get the best possible estimation of the phase screen, as has been explained in Sec. II C. The statistic of the phase screen is 45 ns average delay variation, 4 mm spatial correlation length of the variation. The delay profile is shown in Fig. 6, (along with its estimate after one and five iterations of the FDORT method). The speckle phantom is generated with 15 scatterers per resolution cell. Twenty four realizations are used (four realizations laterally and six axially) for the measurement of the transfer matrix.

In Fig. 7(a) we see that the beampattern through the aberrator appears as a collection of points that are not well resolved. In the case of such an object, the first eigenvector is mainly the Green's function of the brightest spot but as it is not well separated from the other points, it is perturbed by the signal from these points.[4] As a result, the focusing obtained by backpropagation of the first eigenvector is not very good [Fig. 7(b)], but it is better than the original transmit. The estimate of the aberrator profile is obtained from the phase of the 1st eigenvector. It is not very good [Fig. 6(a)]. The process can therefore be iterated: the transmits are partially corrected using the first estimate of the Green's function and a new KK^H matrix is built. The virtual object now corresponds to Fig. 7(b). One point is now clearly brighter than the other and the new first eigenvector will be mainly the Green's function from this point, with less interferences from the other points than during the first iteration. As seen

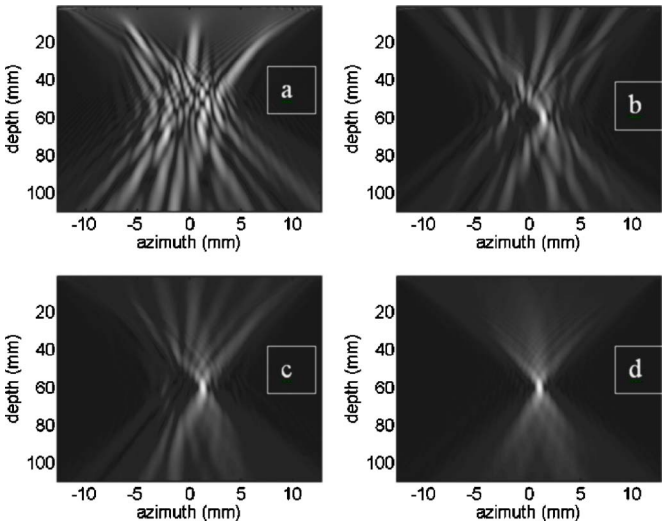


FIG. 7. (Color online) Simulated transmit fields at 7.3 MHz in presence of the near field phase aberrator for a different number of iterations of the algorithm. (a) Initial transmit: it is based on the Green function in homogeneous medium, and therefore the focusing is very poor. It is the equivalent virtual object for the first FDORT iteration (b) The first eigenvector obtained in the first iteration is back-propagated. It focuses mainly on the brightest spot of the first transmit, but with significant interferences. The focusing criterion C is only 0.3. It is used to correct the transmit for the second iteration of FDORT. (c) First eigenvector from the second iteration. The focusing has improved. It is used to correct the transmit for the third iteration. (d) Fifth iteration: the first eigenvector now yields a very good focusing. It is an accurate estimate of the Green function in presence of the phase aberrator. C is now equal to 0.7.

in Fig. 7(c), the focusing quality improved. By iterating the process a few times, the interferences decrease to zero and the first eigenvector converges to the Green's function of the brightest spot of the initial virtual object. This yields a very good focusing through the inhomogeneous medium [Fig. 7(d)]. The phase of the 1st eigenvector yield now a very good estimate of the aberration profile [Fig. 7(b)] after unwrapping and removing the geometric delay.

Our interpretation of the first eigenvalue as the focusing criterion C is very helpful here. It is an objective assessment of the quality of focusing and indicates when the iteration should be stopped. Indeed, as soon as C reaches a certain threshold, we know that we have a good focusing, and a good estimate of the Green's function. The evolution of C as a function of the number of iterations is shown in Fig. 8.

To get a good estimation at the last iteration, it is important to use the largest number of elements in transmit, as explained in Sec. II C.

B. Focusing through a far-field phase screen

In the previous example (Figs. 8 and 7), the heterogeneity was modeled by a *near-field phase screen*. This is a good model if the heterogeneity is localized in a thin layer close to the transducer. This model is also the easiest to correct, as it can be corrected by simple time delays. In the frequency domain, that we consider here, this translates to a phase correction $\phi_m = \omega\tau_m$. Another advantage of the *near-field phase screen*, is that the isoplanatic patch discussed in Sec. II D. is large, as in the homogeneous case. Thus, more realizations can be taken.

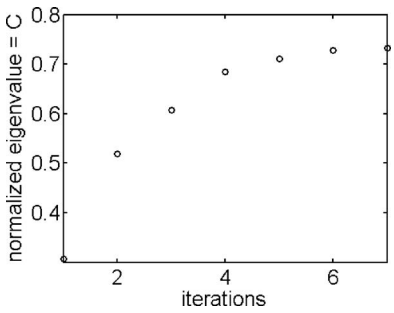


FIG. 8. Evolution of the normalized first eigenvalue in speckle with a simulated near-field phase screen. The normalized first eigenvalue is equal to the focusing criterion C. At the first iteration, C is about 0.3, which is the sign of a bad focusing related to the phase screen. After a few iterations, the algorithm converges and C passes 0.7, which means that the focusing is excellent. The corresponding transmit fields are displayed in Fig. 7.

A more difficult case is obtained if the phase screen is no longer close to the array, but deeper in the medium. Indeed, in this case, the wave fronts are diffracted by the phase screen.[24] This results in deformation of the wave fronts and interference that causes amplitude variation across the array. Therefore, the effect in the array plane can no longer be modeled by simple time delays, but by phase and amplitude (because of the interferences) variations at each frequency.[25] The phase is *a priori* no longer linear as a function of frequency. An additional difficulty results from the limited size of the isoplanatic patch in this case.

The same phase screen as in Sec. III A is simulated, but it is now located at a depth of 20 mm. The focal depth was chosen to be 80 mm. The results are shown in Fig. 9.

As in the case of the near-field phase screen, the convergence of the iteration is reached after a few iterations, with a

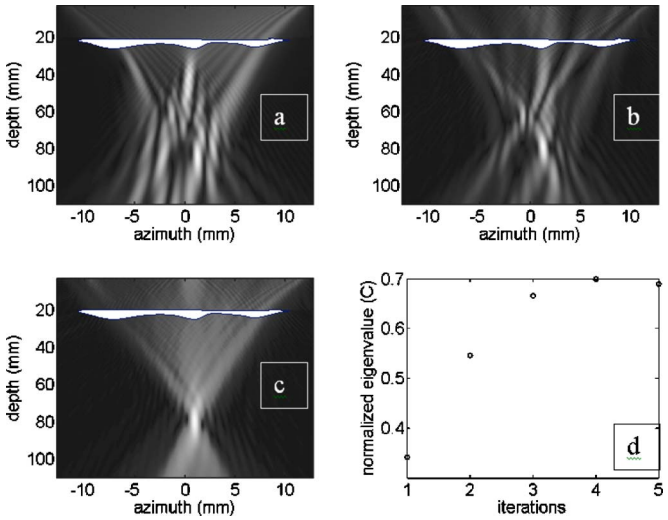


FIG. 9. (Color online) Simulated transmit fields at 7.3 MHz in the presence of the far field phase aberrator (drawn on each transmit in white) located at 20 mm depth, for a different number of iteration of the algorithm. (a) Initial transmit: it is based on the Green function in homogeneous medium, and therefore the focusing is very poor. It is the equivalent virtual object for the first FDORT iteration (b) Field obtained after back-propagation of the first eigenvector obtained in the first iteration. It focuses mainly on the brightest spot of the first transmit, but with interferences. C is 0.35. (c) First eigenvector after the fourth iteration. C is 0.7. (d) Evolution of the normalized eigenvalue, or C factor. Even if the screen is not close to the array, the convergence is reached quickly.

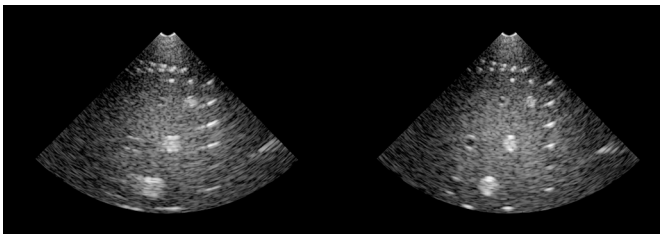


FIG. 10. (Color online) Image of the phantom with a rubber aberrator before (left) and after correction (right) using the FDORT method to improve the focusing in transmit and receive.

focusing criterion close to 0.7. [Fig. 9(d)]. Now, the first eigenvector has an amplitude variation, as expected with the screen in the far-field. Backpropagation of the eigenvector (amplitude plus phase) yields a good focusing [Fig. 9(d)].

To summarize, FDORT has the ability to estimate accurate Green's functions in speckle in the more general model of a far-field phase screen.

C. Medical phantom results

Medical phantom data were obtained using a phased array at 2.7 MHz, and the Philips HDI-5000 scanner. A rubber aberrator was positioned between the probe and the phantom. The thickness of the rubber varies, thus introducing different delays on different elements.

The rubber aberrator was characterized by a 30 ns average delay variation, 3 mm spatial correlation length of the variation (30 ns rms, 3 mm full width half maximum). The aberrator was one dimensional: the delays vary only along the azimuthal dimension of the probe. The aberrated image can be seen in Fig. 10 (left). The region of speckle used for FDORT was 2 mm wide and 2 mm deep, and was located in the center right part of the phantom. The region was carefully selected so that no bright scatterers lie within. About 25 realizations were selected in the region.

1. Results

The convergence of the algorithm was reached after only one iteration. The focusing criterion then reaches a plateau, with a value of 0.7. The phase of the first eigenvector was unwrapped to obtain an estimate of the delay profile introduced by the aberrator, and the delay estimate was used to correct the beamforming process, both in transmit and in receive. An additional difficulty arises here: although it has been shown that the first eigenvector yields a good focusing, the position of the focus is not known. This can yield to an additional linear term in the delay estimate, and unwanted steering. In our results, the delay estimates were detrended to remove any linear term. The corrected image is shown in Fig. 10 (right). A very good image quality is achieved.

2. Discussion

The fact that the convergence occurs faster than in simulation is possibly due to the nature of the speckle. In simulation, a fully developed speckle model was used. The speckle in the phantom may not be fully developed.

The method has also been tried with a two-dimensional rubber aberrator, where the delay variations occur both in azimuth and elevation. The algorithm did not succeed in this case with the simple 1D array. This seems to justify the need for 2D arrays in aberration correction.[10,26,27]

IV. GREEN FUNCTION ESTIMATION AND FOCUSING IN THE PRESENCE OF STRONG INTERFERING SIGNALS

The FDORT method is quite complex to implement as it involves a singular value decomposition (SVD) of the spatial correlation matrix. Other methods exist to estimate Green's functions from speckle signals, for example using cross correlation between neighbor elements to find the differential delays (or phase). Then the differential delays can be integrated to yield the wave-front profile. Implementations in the time[28] and temporal frequency domain[10,16] exist. This would be equivalent to integrating the phase of the coefficients of the first diagonal $(KK^H)_{m,m+1}$ of the FDORT matrix, corresponding to correlation between neighbors. So what is the advantage of the FDORT method? The first advantage has already been discussed: unlike the other methods, FDORT directly provides a measure of the focusing quality through the first eigenvalue. However, such a measure could also be estimated independently.

There is a unique advantage of using FDORT, which is similar to the advantage of using the DORT (or FDORT) method with point scatterers. The fundamental advantage of the DORT method is that it is able to separate the wave fronts from different scatterers. Each eigenvector corresponds to a different scatterer.[4,6] It is therefore able to separate the signal of the target of interest from other signals (other scatterers, interference, etc.). This property is fundamentally linked to the interpretation as a covariance matrix[12] and principal component analysis.[29,30]

In the same fashion, the fact that the FDORT method in speckle involves the decomposition of the spatial correlation matrix enables it to separate signals. Let us consider a simple example where one tries to estimate a Green's function from speckle signals in the presence of a strong interference, coming from another direction than the region of interest. In our case, the interference is a signal coming from "infinite." In practice, the interference can be due to an active source (e.g. a boat in underwater acoustic for example). In medical ultrasound, the interference could be a bright off-axis scatterer. A simple aberration estimation method does not separate the signals, so the estimated phase will be an average of the interference Green's function and the desired Green's function that one wants to estimate.[31] If the interference is much stronger than the speckle signal, then the estimate gives the phase of the interference and it is impossible to know the desired Green's function. However FDORT has the ability to separate the two signals as long as they are orthogonal. This means that the location of the interference source has to be well resolved compared to the region of interest. The first eigenvector will typically be the interference Green's function, while the second eigenvector will be the desired speckle Green's function. It does not matter how much stronger the interference signal is.

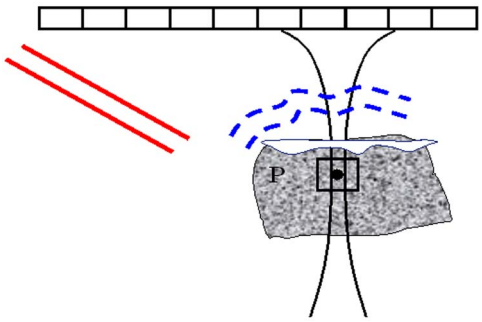


FIG. 11. (Color online) Setup for the simulation: the goal is to estimate the Green's function of point P in speckle (dash blue wave front) using focused transmit, in the presence of a strong interferer signal (solid red wave front). The signal from point P travels through a heterogeneity, but the interferer does not. In order to estimate the parameters of the heterogeneity it is necessary to isolate the wave front of P from the interferers. In the simulation, the interferer is 1000 times stronger (60 dB) than the signal.

A. Simulation

The setup is illustrated in Fig. 11. The speckle is placed behind a phase screen. The interference signal is 1000 times stronger (60 dB) than the speckle signal of interest. The magnitude of the interference is exaggerated to demonstrate the efficiency of the FDORT algorithm.

The phases for the first and second eigenvector of FDORT are displayed in Fig. 12. The geometrical delay law has been subtracted from the second eigenvector for clarity. It is clear that the first eigenvector corresponds to the interference and the second eigenvector to the speckle target. The distortion due to the heterogeneity is seen in the phase of the

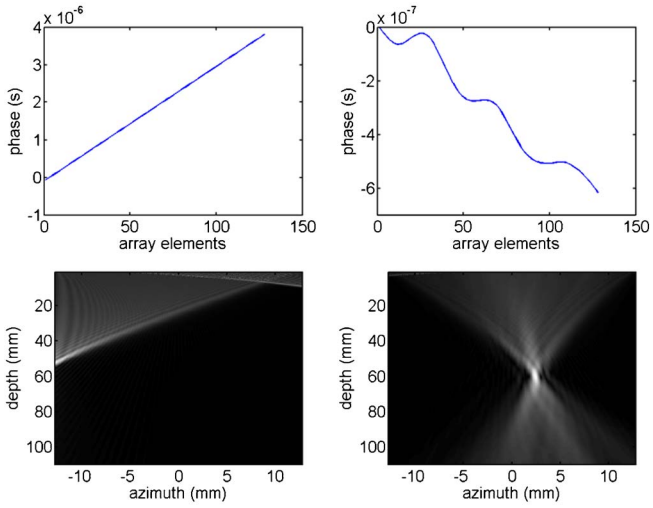


FIG. 12. (Color online) (Top) Phase for the first and second eigenvector of FDORT, corresponding to the setup of Fig. 11. The first eigenvector (left) corresponds clearly to the interference, while the second corresponds to the speckle, and carries information on the heterogeneity. For clarity, a geometrical delay has been removed from the second eigenvector phase. (Bottom) Fields after numerical backpropagation of the first (left) and second eigenvectors. This confirms that the first eigenvector corresponds to the interference, while the second eigenvector corresponds to the desired speckle signal. The eigenvector corresponding to the speckle is a very good estimate of the Green's function and leads to a good focusing through the heterogeneity despite the presence of a strong interfering signal.

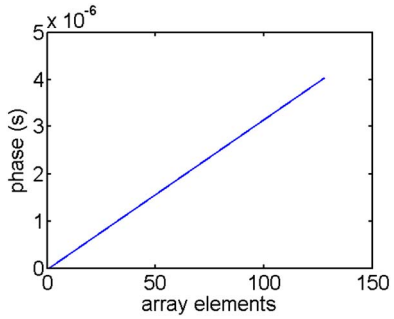


FIG. 13. (Color online) Phase estimate obtained with the 1-lag cross-correlation method. The estimate is completely dominated by the interference, and it is not possible to extract any information about the heterogeneity.

speckle Green's function, and not in the phase of the interference as the interference signal did not go through the heterogeneity.

Backpropagation of the eigenvectors, shown in Fig. 12, confirms that the two signals are very well separated. In addition, it shows that the estimate of the speckle Green's function is excellent as one is able to achieve very good focusing through the heterogeneity. The eigenvectors shown here were obtained after four iterations, where the second eigenvector was used to correct the transmit.

It is remarkable that such an accurate estimate of the speckle Green's function is possible in the presence of an interfering signal 1000 times stronger (60 dB). This is an important property of FDORT. Note that the method does not require the interference to be much stronger than the signal of interest.

With another estimation method, such as the neighboring element cross-correlation method, this is not possible. The phase profile estimated with this method is shown in Fig. 13. It is completely dominated by the strong interference. It is difficult to extract any information about the heterogeneity from this curve, and focusing through the aberrator cannot be achieved using this method.

Some application of this signal separation property also arises when there are a few bright scatterers in a speckle medium. If one tries to estimate the Green's function in a region of pure speckle, there are chances that one receives not only echoes from the speckle region, which is the region insonified by the transmits, but also from bright scatterers outside the region of interest.[31] This happens even if the transmit is far from the bright scatterers because there is some energy in the side lobes of the transmit. In this case we talk about off-axis scatterers. This can be an important problem in aberration correction if the off-axis scatterer is not in the isoplanatic patch of the speckle. In this case the wave fronts of the speckle and of the off-axis scaterrers are distorted by different aberrations, and a classic estimation method will yield an average of the two aberrations, which will not lead to very good focusing on the speckle region, nor on the off-axis scatterer. FDORT can separate the aberration profiles.

V. CONCLUSION

It has been shown that FDORT could be used to extract Green's functions from speckle signals as well as from resolved scatterers. With resolved scatterers, the operator KK^H, built with the FDORT method, is connected to time reversal. In speckle, it has been shown that KK^H is connected to another fundamental theorem of acoustics, the Van Cittert–Zernike theorem. This yields an interpretation of the first eigenvalue of the FDORT matrix as the focusing criterion C, which is a measure of the quality of the focusing, and has been used in numerous publications. In speckle KK^H can alternatively be interpreted as a canonical time reversal operator for an equivalent virtual object that has the shape of the transmit. This leads to an intuitive understanding of the eigenvectors. Thus, the first eigenvector focuses on and is the Green's function of the brightest point of the transmit beam pattern, which, in a homogeneous medium is the focal spot of the transmit. In a heterogeneous medium, a few iterations of the method are often needed to converge on the Green's function of a point. The interpretation of the first eigenvalue of KK^H is of particular interest, as it indicates objectively when the focusing is good enough and the iteration can be stopped. The focusing properties of the first eigenvector of KK^H have been demonstrated for different models of heterogeneities. Finally, a main feature of the FDORT method is its ability to separate wave fronts. In particular, it is able to separate the signal of interest from interferences and off-axis scatterers.

The method in speckle is in principle very similar to previously published work,[8,9] where the first eigenvector of a correlation matrix was considered. The novelty of this work is both theoretical, in particular the interpretation of the first eigenvalue as a focusing criterion and experimental: the link with the FDORT method provides a very practical implementation of the method, which yields good experimental results.

ACKNOWLEDGMENTS

The authors would like to thank Michael Burcher for his help editing the manuscript.

[1]C. Prada, F. Wu, and M. Fink, "The iterative time reversal mirror: A solution to self-focusing in the pulse echo mode," J. Acoust. Soc. Am. **90**, 1119–1129 (1991).
[2]M. Fink, "Time reversal of ultrasonic fields. I. Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[3]C. Prada and M. Fink, "Eigenmodes of the time reversal operator: a solution to selective focusing in multiple-target media," Wave Motion **20**, 151–163 (1994).
[4]C. Prada, S. Manneville, D. Spoliansky, and M. Fink, "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[5]J.-L. Robert, M. Burcher, C. Cohen-Bacrie, and M. Fink, "Time reversal operator decomposition with focused transmission and robustness to speckle noise: Application to microcalcification detection," J. Acoust. Soc. Am. **119**, 3848–3859 (2006).
[6]M. R. Burcher, A. T. Fernandez, and C. Cohen-Bacrie, "A novel phase aberration measurement technique derived from the DORT method: Comparison with correlation-based method on simulated and in-vivo data," *Proceeding of the IEEE Ultrasonics Symposium*, Vol. **2**, 860–865 (2004).
[7]R. Mallart and M. Fink, "Adaptive focusing in scattering media through sound-speed inhomogeneities: The Van Cittert Zernike approach and focusing criterion," J. Acoust. Soc. Am. **96**, 3721–3732 (1994).
[8]T. Varslot, H. Krogstadt, E. Mo, and B. A. Angelsen, "Eigenfunction analysis of stochastic backscatter for characterization of acoustic aberration in medical ultrasound imaging," J. Acoust. Soc. Am. **115**, 3068–3076 (2004).

[9]S. -E. Masoy, T. Varslot, and B. A. Angelsen, "Iteration of transmit-beam aberration correction in medical ultrasound imaging," J. Acoust. Soc. Am. **117**, 450–461 (2005).
[10]R. C. Waag and J. P. Astheimer, "Statistical estimation of ultrasonic propagation path parameters for aberration correction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 851–869 (2005).
[11]H. V. Trees, *Optimum Array Processing - Detection, Estimation, and Modulation Theory* (Wiley-Interscience, New York, 2002), Vol. **IV**.
[12]F. K. Gruber, E. A. Marengo, and A. J. Devaney, "Time-reversal imaging with multiple signal classification considering multiple scattering between the targets," J. Acoust. Soc. Am. **115**, 3042–3047 (2004).
[13]C. Prada and J-L. Thomas, "Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix," J. Acoust. Soc. Am. **114**, 235–243 (2003).
[14]A. Aubry, J. d. Rosny, J.-G. Minonzio, C. Prada, and M. Fink, "Gaussian beams and Legendre polynomials as invariants of the time reversal operator for a large rigid cylinder," J. Acoust. Soc. Am. **120**, 2746–2754 (2006).
[15]S.-E. Masoy, B. Angelsen, and T. Varslot, "Variance analysis of arrival time and amplitude estimates from random speckle signal," J. Acoust. Soc. Am. **121**, 286–297 (2007).
[16]J. P. Astheimer, W. C. Pilkington, and R. C. Waag, "Reduction of variance in spectral estimates for correction of ultrasonic aberration," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 79–89 (2006).
[17]R. F. Wagner, M. F. Insana, and S. W. Smith, "Fundamental correlation lengths of coherent speckle in medical ultrasonic images," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 34–44 (1988).
[18]J. J. Dahl, M. S. Soo, and G. E. Trahey, "Spatial and temporal aberrator stability for real-time adaptive imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1504–1517 (2005).
[19]J. Lacefield and R. C. Waag, "Spatial coherence analysis applied to aberration correction using a two-dimensional array system," J. Acoust. Soc. Am. **112**, 2558–2566 (2002).
[20]W. F. Walker and G. E. Trahey, "The application of k-space in pulse echo ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 541–558 (1998).
[21]M. O. Donnell, "Quantitative ultrasonic backscatter measurements in the presence of phase distortion," J. Acoust. Soc. Am. **72**, 1719–1725 (1982).
[22]N. Levin, E. T. Gregg, and W. S. Stephen, "Phase aberration correction in medical ultrasound using speckle brightness as a quality factor," J. Acoust. Soc. Am. **85**, 1819–1833 (1989).
[23]J. A. Jensen, "Field: A program for simulating ultrasound systems," Med. Biol. Eng. Comput. **34**, 351–353 (1996).
[24]C. Dorme and M. A. Fink, "Ultrasonic beam steering through inhomogeneous layers with a time reversal mirror," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 167–175 (1996).
[25]S.-E. Masoy, B. Angelsen, and T. Varslot, "Estimation of ultrasound wave aberration with signals from random scatterers," J. Acoust. Soc. Am. **115**, 2998–3009 (2004).
[26]A. T. Fernandez, K. L. Gammelmark, J. J. Dahl, C. G. Keen, R. C. Gauss, and G. E. Trahey, "Synthetic elevation beamforming and image acquisition capabilities using an 8 /spl times/ 128 1.75D array," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 40–57 (2003).
[27]A. T. Fernandez and G. E. Trahey, "Two-dimensional phase aberration correction using an ultrasonic 1.75D array: case study on breast microcalcifications," Proc.-IEEE Ultrason. Symp., **1**, 348–353 (2003).
[28]S. W. Flax and M. O'Donnell, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 758–767 (1988).
[29]C. R. Rao, "The use and interpretation of principal component analysis in applied research," Sankhya, Ser. A **26**, 329–358 (1964).
[30]W. W. Cooley and P. R. Lohnes, *Multivariate Data Analysis* (Wiley, New York 1971).
[31]J. J. Dahl and G. E. Trahey, "Off-axis scatterer filters for improved aberration measurements," Proc.-IEEE Ultrason. Symp., **2**, 1094–1098 (2003).

Determining tomographic arrival times based on matched filter processing: Considering the impact of ocean waves

James K. Lewis[a)]
Scientific Solutions, Inc., 4875 Kikala Road, Kalaheo, Hawaii 96741

(Received 2 August 2007; accepted 13 November 2007)

Reflection of high-frequency acoustic signals from an air-sea interface with waves is considered in terms of determining travel times for acoustic tomography. Wave-induced, multi-path rays are investigated to determine how they influence the assumption that the time of the largest matched filter magnitude between the source and receiver signals is the best estimate of the arrival time of the flat-surface specular ray path. A simple reflection model is developed to consider the impact of in-plane, multi-path arrivals on the signal detected by a receiver. It is found that the number of multi-path rays between a source and receiver increases significantly with the number of times the ray paths strike the ocean surface. In test cases, there was always one of the multi-path rays that closely followed the flat-surface specular ray path. But all the multi-path rays arrive at the receiver almost simultaneously, resulting in interference with the signal from the flat-surface specular ray path. As a result, multi-path arrivals due to open ocean surface waves often distort the received signal such that maxima of matched filtering magnitudes will not always be a reliable indicator of the arrival time of flat-surface specular ray paths. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2821974]

PACS number(s): 43.60.Rw, 43.30.Re [AIT] Pages: 878–886

I. INTRODUCTION

Determining precise arrival times of acoustic signals in the open ocean can be complicated by numerous factors. One of these is the impact of multiple ray paths that can be generated as an acoustic wave front interacts with the air-sea interface. The two-dimensional wave spectra of a typical ocean surface can result in numerous locations at the sea surface at which the sound from an omni-directional source can be directed toward a receiver.[1]

Our interest in these multi-paths is a result of the recent work of Lewis *et al.*[2] (hereafter referred to as L05). In that study, we considered travel times for tomographic purposes for single- and double-surface bounce ray paths between bottom-mounted sources and receivers. Acoustic information collected by a receiver includes not only that of a ray path of interest but also all the acoustic information from the additional multi-paths that can be generated by each individual source ping as a result of interactions of the acoustic signal with the ocean surface. Care must be taken to develop appropriate processing schemes to best determine the arrival time for a specific ray path of interest among all the acoustic information that is contained within the receiver data. In this paper we address some critical factors that must be understood before developing a scheme for determining arrival times for acoustic tomography.

This work assumes that we are dealing with high-frequency acoustic signals for which the air-sea interface acts as a perfectly reflective pressure-release surface. We also assume that any portion of a receiver signal due to any small scale roughness of the air-sea interface consists of incoherent scattering.[3] In L05, the signals were at 8–11 kHz. At these frequencies, acoustic wavelengths ($\lambda = 14–19$ cm) are considerably shorter than the spatial variations of the open ocean surface due to surface gravity waves. Even for short, 1 s, open-ocean waves, the wavelength of such gravity waves is an order of magnitude larger than the acoustic wavelengths at 8–11 kHz. Thus, an air-sea interface consisting of gravity waves acts as a smooth but undulating boundary reflecting the 8–11 kHz sound. For typical open-ocean, gravity wave conditions, we would expect our assumption of a reflective surface to be valid for frequencies greater than ~3 kHz.

The work of Dahl[1] provides an interesting set of results dealing with single-surface bounce ray paths between a source and receiver. Using approximations of the two-dimensional surface wave spectra, Dahl developed an expression for the variable σ, the cross sectional area (m^2) of surfaces per ocean surface area at which the acoustic wave front would be directed from the ocean surface toward the receiver. Dahl used this air-sea interface variable to study the spatial distribution of locations where such single-surface bounce ray paths would strike the ocean surface above a source-receiver pair. Dahl dealt with a frequency of 30 kHz (~5 cm acoustic wavelength), with nominal grazing angles of 14° and 20°.

Following Dahl's conventions, in-plane ray paths refer to those within the plane defined by the source-receiver pair and the point at the ocean surface at which the ray path would be reflected to the receiver if the ocean surface were flat and the reflection purely specular. We refer to such a ray path as the flat-surface ray path (FSRP). Dahl found that the number of in-plane ray paths increased with distance from the source (and from the receiver) toward the flat-surface specular point. For the out-of-plane direction, the number of multi-path rays between the source and receiver decreased

a)Electronic mail: jlewis@scisol.com

 0001-4966/2008/123(2)/878/9/$23.00

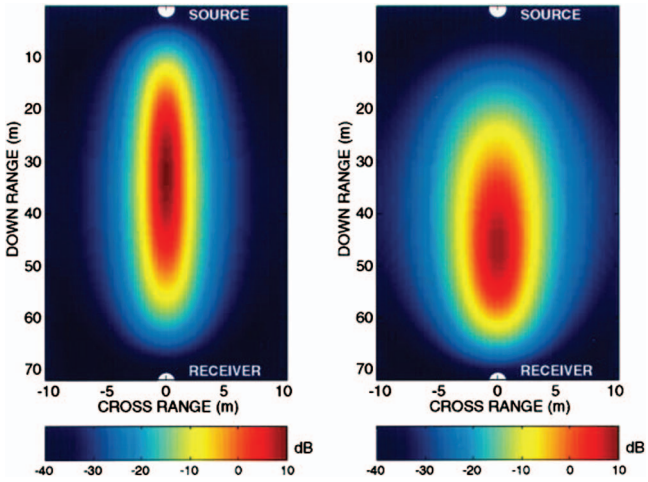


FIG. 1. Dahl's model (see ref. 1) results for the bistatic cross section σ for single-surface bounce acoustic rays as a function of position on the sea surface for a source depth of 7.7 m and a receiver depth of 9.5 m (left) and a source depth of 17.7 m and receiver depth of 9.5 m (right).

dramatically with distance away from the in-plane region. Examples of Dahl's results are shown in Fig. 1.

For this study, we interpret σ as providing a guideline as to the distribution of locations on the sea surface that can give rise to single-surface bounce ray paths between a source and a receiver. And although Dahl's results provide very useful insights, they do not address critical factors that must be understood before developing a scheme for determining arrival times for acoustic tomography. This is the topic we pursue in this paper. In the next section we will discuss the details of the acoustic tomography approach being used (L05) and how surface reflection and scattering of an acoustic signal might be expected to impact the technology. The approach of L05 utilized magnitudes of matched filtering the data from a source and a receiver, assuming that the time of the largest matched filter magnitude was the best estimate of the arrival time of the desired ray path between the source and receiver. In Sec. III, we put forward a simple model for the impact of wave-induced, multi-path arrivals on the signal detected by a receiver. Results of this model are discussed in terms of their relationships to matched filter results. In Sec. IV we present and discuss observed matched filter magnitudes for ray paths that reflect once or twice from the air-sea interface. In Sec. V, a hypothesis is developed to explain differences between the structures of the matched filter peaks for single- and double-surface bounce multi-paths as would be indicated by the model and as seen in the observations. In the final section, our discussion relates the results of our model, observed matched filter peaks, and some calculations intended to test our hypothesis.

We find that our model provides a good explanation of the structure of observed matched filtered magnitudes for both single- and double-surface bounce ray paths. And although it is not unreasonable to assume that the largest matched filter magnitude is the best estimate of a ray path arrival time, we find that it is true only for the single-surface bounce ray path. As it turns out, the number of multi-path arrivals due to surface waves increases considerably with the number of reflections from the ocean surface. Yet, the actual range of arrival times of the various multi-path rays varies by very little, usually 10's of microseconds. As a result, the signals from the multi-path rays overlap one another in time at the receiver. With a small number of multi-paths (like what you might get with single-surface bounce ray paths), the net signal made up of the overlapping multi-path signals detected at the receiver is still relatively well correlated to the source signal. As a result, the time of the largest matched filter magnitude is the best estimate of the travel time of the desired ray path. But as the receiver is inundated with more and more acoustic information from multi-path signals arriving over longer periods of time (as with double-surface bounce multi-paths), the received signal becomes so distorted that the time of the largest matched filter magnitude is no longer the best estimate of the travel time of the desired ray path.

II. TOMOGRAPHIC SCHEME: ASSIMILATING TRAVEL TIME DATA INTO AN OCEAN CIRCULATION MODEL

In L05, variations in acoustic travel times for a specific ray path between a source and a receiver are assimilated into a dynamic ocean circulation model in order to adjust the model temperatures and salinities. The adjustments are such that the model-predicted travel times better match the observed travel times. With this technology, observed acoustic information for a specific ray path is associated with an established set of grid cells within the domain of the ocean model. As such, we work with acoustic rays whose paths have been established *a priori*.

At times we work with ray paths that travel from a bottom-mounted source, up to the ocean surface, and then directly to a bottom-mounted receiver. Such a single-surface bounce ray path is determined *a priori* utilizing an acoustic propagation model with a flat ocean surface (specular reflection). *Thus*, *when we speak of the "desired ray path" in this paper*, *we are referring to the flat-surface*, *specular ray path*. It is the arrival time of this ray path that we wish to determine from the acoustic signal detected at the receiver. In some instances, we have considered double-surface bounce ray paths.[2] So we will also discuss the impact of surface waves on determining the arrival of the flat-surface specular ray for such ray paths.

In L05, we worked with receiver data covering a window of time during which we would expect the multi-path signals for specific ray paths to arrive. The receiver data were correlated for all possible arrival times with the source signal using the technique of matched filtering.[2] The assumption was that the time of the largest matched filter magnitude (a measure of the correlation between the source and receiver signals) would be the best estimate of the arrival time of the specular FSRP, whether we are dealing with single- or double-surface bounce ray paths. Although this is not an unreasonable assumption, it has not been proven for the conditions of a received signal that is comprised of many multi-path reflections resulting from the waves at the air-sea interface.

III. A MODEL

To obtain additional insights into the impact of an air-sea interface with waves, we consider a simple reflection model that is somewhat similar to the approach of Dahl[1] and Heitsenrether and Badiey.[3] In this case, we focus on a model that can be utilized to relate multi-path arrivals to matched filter results for both single- and double-surface bounce reflections.

The depths of a bottom-mounted, omni-directional source and a bottom-mounted receiver are specified in our model, along with the lateral distance between the source and the receiver. The slope of the ocean bottom between the source and receiver is taken as constant. The density of the medium within the model is set as a constant based on the assumption that refraction plays only a minor role in determining the differences in arrival times for various ray paths. Moreover, perfect reflection of an acoustic ray can occur at both the ocean surface and the ocean bottom between the source and receiver.

For double-surface arrivals with steeper grazing angles across the ocean bottom, we would expect bottom loss to have an impact on the intensity of the acoustic signal as it arrives at the receiver. However, we found that the distribution of locations at which double-surface bounce rays strike the ocean bottom is fairly limited: i.e., the range of grazing angles is limited. As a result, we ignore the impact of bottom loss on acoustic signal intensities for this study.

Although the ocean bottom has a constant slope, we allow the ocean surface to be sinusoidal. We can specify the wave amplitude and the wave period in the model, and the wavelength is determined by the deep-water condition for ocean surface waves.[4] Thus, the sinusoidal surface provides a means of simplifying surface wave conditions in terms of average amplitude, average period, and average wavelength. We can also specify the phase of the sinusoidal pattern of the wave field across the ocean surface between the source and receiver. We consider the sinusoidal surface to be fixed as the acoustic front strikes the different locations at the ocean surface between the source and the receiver (we will show later that this is a valid assumption). As such, the curvature of the ocean surface is known at all points. Thus, we can always determine the direction that an impacting ray will be reflected from any point on the ocean surface.

The sinusoidal surface represents the idealized situation in which the wave field is traveling parallel to the line joining the source and receiver. Thus, our model only results in in-plane reflections between the source and receiver. In reality, out-of-the-plane reflections due to components of a wave field traveling perpendicular to the source-receiver plane will also exist. If the length of the out-of-plane surface wave is small compared to the distance between the source and receiver, out-of-plane reflections will have characteristics similar to that of in-plane reflections, just longer path lengths (and, thus, longer travel times between the source and receiver). As pointed out by Dahl[1] (and seen in Fig. 1), the number of out-of-plane reflections decreases dramatically with distance from the in-plane location.

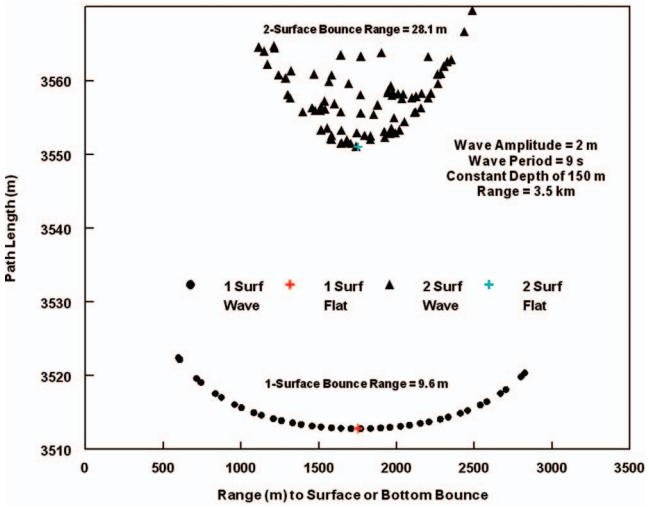


FIG. 2. Path length versus distance to where the ray path strikes the ocean surface (circles) or the ocean bottom (triangles). The red (blue) plus indicates where the single-surface (double-surface) bounce would occur with a flat ocean surface.

A. Model simulations

The acoustic wave front from the omni-directional source will strike all points of the ocean surface between the source and receiver. The model simulates ray paths from the source striking the ocean surface every 0.1 m between the source and receiver, and each ray path is followed to ascertain whether it travels to the location of the receiver. We present two examples from the reflection model. Each example shows the results of single- and double-surface bounce ray paths that travel between the source and receiver.

Our graphical representations (Figs. 2 and 3) are in terms of the total length of a ray path from the source to the receiver (ordinate) vs. the location where the ray path (a) hits the ocean surface for single-surface bounce rays or (b) where the ray hits the ocean bottom for double-surface bounce rays. These results are for a surface wave with an amplitude of

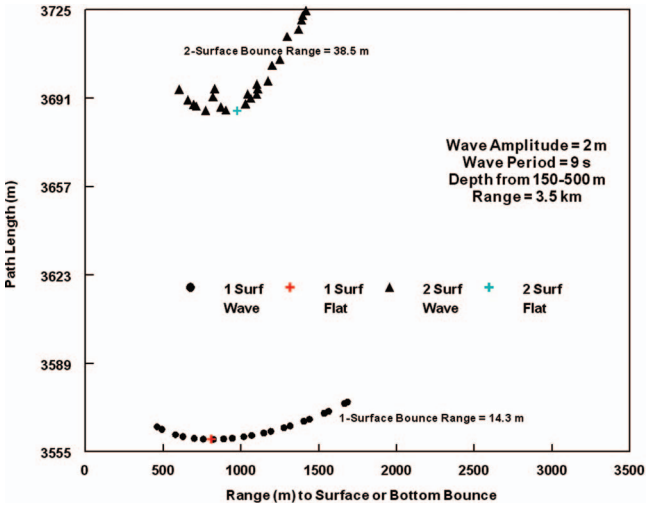


FIG. 3. Path length versus distance to where the ray path strikes the ocean surface (circles) or the ocean bottom (triangles). The red (blue) plus indicates where the single-surface (double-surface) bounce would occur with a flat ocean surface.

2 m and a period of 9 s (corresponding to a wavelength of ~126 m). We use a 9 s wave period because average wave periods in the region of our previous work[2] vary from 7 to 11 s.

The first results (Fig. 2) are for a source and receiver at a depth of 150 m, separated by 3.5 km. The model indicates that reflections from the ocean surface will result in a number of multi-path arrivals at the receiver for both the single- and double-surface bounce rays. Our specified ocean surface reflects 37 single-surface bounce rays from the acoustic front emanating from the source. There are 215 reflected double-surface bounce ray paths. Geometrically, it can be shown that the shortest ray path will always be that which strikes the ocean surface (or bottom) exactly halfway between the source and receiver (red and blue +'s), these representing the specular FSRPs. We see that there are ray paths reflecting from the sinusoidal surface that are very close to the FSRPs.

The second example (Fig. 3) is for a source at a depth of 150 m and a receiver at a depth of 500 m. The lateral distance between the two is again 3.5 km. There are only certain regions of the sinusoidal ocean surface that result in multi-path reflections of the source's acoustic wave front to the receiver: those in shallower water. Once again, we see that there are ray paths for our sinusoidal surface that are very close to the flat-surface ray paths. But in this case, there are some double-surface bounce arrivals that actually have path lengths that are almost equal to the flat-surface ray path but have noticeable variations in their paths (e.g., where they strike the ocean bottom relative to where the flat-surface ray path would strike).

The results shown in Figs. 2 and 3 are analogous to the findings of Dahl (Fig. 1). The density of source-to-receiver reflection points tends to increase to some maximum between the source and receiver, with this maximum shifting toward shallower water when the source-receive depths are different. Simple geometry shows that any out-of-plane reflections will always have path lengths greater than the in-plane reflections, so out-of-plane path lengths would be found above any corresponding in-plane points in Figs. 2 and 3. Moreover, geometry shows that there will be a decline of out-of-plane points of reflections with distance in the out-of-plane direction, similar to what is seen in Fig. 1.

We now consider our assumption of the sinusoidal surface being constant in structure as the various rays strike the ocean surface. The maximum difference between the path lengths of any of the rays was 38.5 m. If the average sound speed is 1500 m/s, then at the very most the time span between when the various rays strike the ocean surface would be 0.026 s. The spatial structure of the 9 s waves varies minutely over a 0.026 s time period, changing by only 0.3%. Thus, the surface structure can be assumed stationary during the period that the various ray paths would strike the ocean surface.

B. Model results as related to matched filter results

The model results clearly show that a typical open ocean surface can result in a number of multi-path arrivals from each signal sent from an omni-directional source. The number of multi-path arrivals is related to the direction of wave propagation, approaching only one reflection as the propagation direction of the sinusoidal surface wave approaches 90° relative to the source-receiver plane (making the in-plane sea surface flat for all locations between the source and receiver). However, in the open ocean we can expect a spectrum of waves in both the in-plane and out-of-plane directions, so we can almost always expect more than one multi-path arrival for each signal sent from a source.

The model shows that the ray paths required in the approach of L05 for tomography (the specular FSRPs) have the shortest (or very close to the shortest) path lengths of all the multi-path rays. Thus, near-FSRPs are the first (or, one of the first for double-surface bounce ray paths) of all the multi-paths to reach the receiver. Individually, each reflection indicated in Figs. 2 and 3 would result in a matched filter peak as a function of arrival time. Thus, for a series of matched filter peaks associated with, say, the single-surface bounce multi-paths, the arrival time for the near-FSRP should be the very first matched filter peak (shortest ray path). This would be true for double-surface bounce multi-paths also.

However, there is a complicating factor of signal overlap in the receiver data. Considering the small differences in path lengths, the range of arrival times (again using an average sound speed of 1500 m/s) would only span at most 26 ms. Considering that the source signals were 0.1 s, our model results would imply that all the in-plane rays from a single ping would effectively arrive at the receiver at the same time. We need to quantify how these overlapping arrivals will impact the ability of matched filter calculations to extract the correct arrival time of the near-FSRPs.

Finally, we note the range of path lengths predicted for the single- and double-surface bounce ray paths. From the reflection model range of path lengths (Figs. 2 and 3), we would expect that the duration at the receiver of the single-bounce rays would be up to T_S=0.1 s (the duration of the source signal) plus Δt_1=9.5 ms (using the 14.3 m range from Fig. 3 and an average sound speed of 1500 m/s). The expected duration of the double-surface bounce paths would be up to T_S plus Δt_2=25.7 ms (using the 38.5 m range from Fig. 3 and an average sound speed of 1500 m/s). We see that Δt_2 is ~2.5 times longer than that of the Δt_1 of the single-surface bounce paths. This will be shown to be a critical factor for properly determining the specular path arrival time for double-surface bounce paths using matched filter processing.

IV. OBSERVED MATCHED FILTER MAGNITUDES

In the L05 approach, 12–16 signals over ~30 s period are transmitted from each source for each tomographic measurement. All these are utilized in determining arrival times using matched filtering. The propagation of the surface wave field over the 30 s period will result in slightly different arrival times of the reflected rays for each signal. As a result, we can almost always expect that one of the reflected ray paths out of the 12–16 signals will be close to that of the FSRP.

In addition, overlaying the matched filter peaks calculated from 12 to 16 signals provides a reliable means of

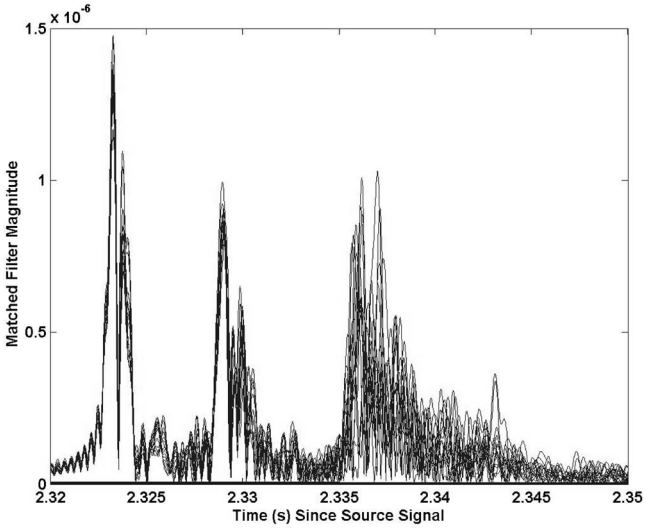


FIG. 4. Matched filter magnitudes calculated from 12 individual signals collected over a 30 s period. The single-surface bounce arrival is 2.331 s as calculated using the Bellhop acoustic propagation model with sound speed structure from an ocean circulation model.

distinguishing those arrivals that do not interact with the ocean surface from the single- and double-surface bounce ray paths. Over a 30 s period, we do not expect much variation in those ray paths traveling through the water column and not interacting with the ocean surface. Consider Fig. 4, which shows overlaid matched filter magnitudes calculated from 12 signals collected over a 30 s period. An acoustic propagation model using model-predicted sound speed structure indicated an arrival time for the single-surface bounce ray as 2.331 s. Although there is a spike in magnitudes starting at ~2.329 s, the variations of the matched filter magnitudes from all 12 signals are relatively small, producing a series of "hollow" spikes of magnitudes. This is not a signature of 12 sets of multi-path arrivals that have interacted with an open ocean surface over a 30 s time interval. The actual single-surface bounce arrivals occur from 2.335 to 2.34 s, a confusion of matched filter peaks due to the variations of the surface wave field during the 30 s that the acoustic data were collected.

We now consider several matched filter results (Fig. 5) from data collected during 2003 from a number of source-

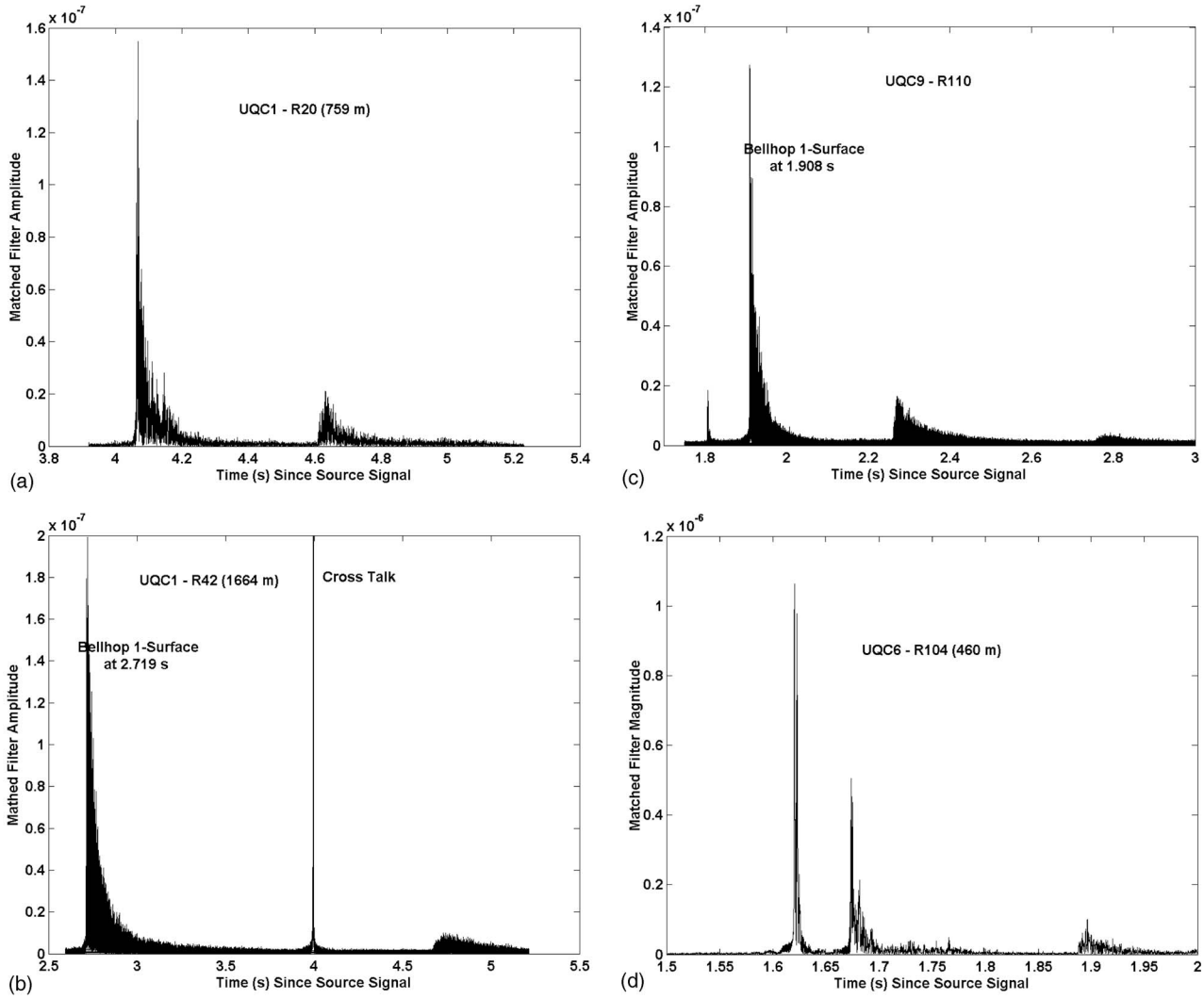


FIG. 5. Examples of matched filter results from data collected offshore of Kauai, Hawaii, in June 2003. A—source depth of 1623 m, receiver depth of 759 m. B—source depth of 1623 m, receiver depth of 1664 m. C—source depth of 575 m, receiver depth of 490 m. D—source depth of 306 m, receiver depth of 460 m.

receiver pairs[2] with different depth configurations. Although we can expect that the sound speed profiles and ray paths would vary for these different source-receiver pairs, we see that the structures of the matched filter results are quite similar. These and other matched filter results[5] indicate that the single-surface bounce matched filter magnitudes (larger peaks) have a Rayleigh-type of distribution (in the bottom right of Fig. 5(d), this arrival is at 1.68 s). These matched filter magnitudes almost instantly increase from a background level to a maximum, and then quickly decrease in an exponential fashion. The decrease is referred to as *time spreading*, a phenomenon related to the interactions of the acoustic signal with multiple locations at the air-sea interface.[6]

The double-surface bounce rays are represented in Fig. 5 by the smaller increases in matched filter magnitudes that occur well after the single-surface bounce matched filter magnitudes. We found that, in almost every data set we have examined, the maximum in the double-surface bounce matched filter peaks occurs after a more gradual increase from the background level when compared to the increases of the single-surface bounce peaks. In addition, the following decrease in double-surface bounce matched filter magnitudes (time spreading) is more gradual and occurs over a longer period of time than for single-surface bounce matched filter results.

The observations show that matched filter processing of those parts of receiver data with the overlapping, single-surface bounce, multi-path signals results in a maximum magnitude at the very beginning of the detected arrivals. From our reflection model, the first arrivals should correspond to the signal associated with the near-FSRPs. Thus, for single-surface ray paths, it appears that the assumption that the arrival time of the desired ray path (the specular FSRP) is indeed associated with the maximum matched filter magnitude.

However, for those parts of receiver data that consist of the overlapping, double-surface bounce, multi-path signals, observations show that matched filter processing results in maximum magnitudes that consistently occur after the first double-surface bounce arrivals are detected. The reflection model indicates that the near-FSRP arrivals for the double-surface bounce ray paths should be the found at the first of the matched filter results. So here we have the situation in which maximum matched filter magnitudes apparently do not correspond with the arrivals of the ray paths that we wish to use for tomography.

V. HYPOTHESIS

We now consider a hypothesis concerning the differences between the structures of the matched filter peaks for single- and double-surface bounce ray paths. We use the reflection model results in terms of number of reflections and total path lengths. First, we note that the variations in path lengths are small for either set of ray paths. As such, we assume that the intensities at the receiver of the various multi-path signals do not differ from one another to any discernable degree. The source signal used by L05 is the chirp

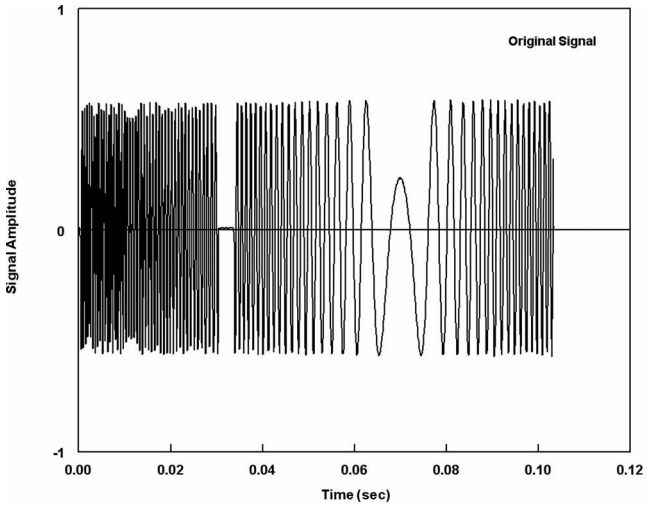


FIG. 6. The source signal used in L05.

shown in Fig. 6. This chirp has a very distinct structure for correlating with a receiver signal, and this signal will be present for each multi-path ray indicated in Figs. 2 and 3.

We generate synthetic receiver data for the various multi-paths in Fig. 2 using the following procedure. First, the path lengths for the single- and double-surface bounce paths are converted to relative arrival times T assuming that the average sound speed along each path is 1500 m/s:

$$T = (\text{path length} - L_{\min})/1500 \text{ m/s},$$

where $L_{\min}$ is either the minimum single- or double-surface bounce path length, depending on which set of multi-paths rays being worked with. Thus, the sound from the near-FSRP (minimum path length) provides the initial variation in acoustic levels. Based on intensity as a function of time (Fig. 6) for each reflection, we can determine a net sound level as a function of time as the sum of the intensity levels for all the multi-path rays (37 for the single-surface bounce rays, 215 for the double-surface bounce rays), each adjusted for the time when it arrives at the receiver.

First, consider the single-surface bounce rays indicated in Fig. 2. The signal seen at the receiver due to all the multi-path rays is the sum of intensity levels seen in Fig. 6 but adjusted for each path by arrival times. Constructive and destructive interference between individual rays results in the receiver signal shown in Fig. 7(a). We see that the signals in Figs. 6 and 7(a) should be reasonably well correlated, resulting in a larger matched filter peak. Moreover, the time difference between the most notable feature of the signals in Figs. 6 and 7(a) (the spreading of the oscillation of the signal at 0.07 s in Fig. 6) is quite small. This should result in a larger matched filter magnitude at about the time of the near-FSRP arrival. From this we would infer that the matched filter peak for the near-FSRP, single-surface bounce ray path should be the first major peak of the matched filter results.

We now consider the double-surface bounce ray paths shown in Fig. 2. Again, the path lengths were converted to relative arrival times, and sound levels were summed over time for all 215 double-surface bounce rays. The results are shown in Fig. 7(b). The major consequence of having far

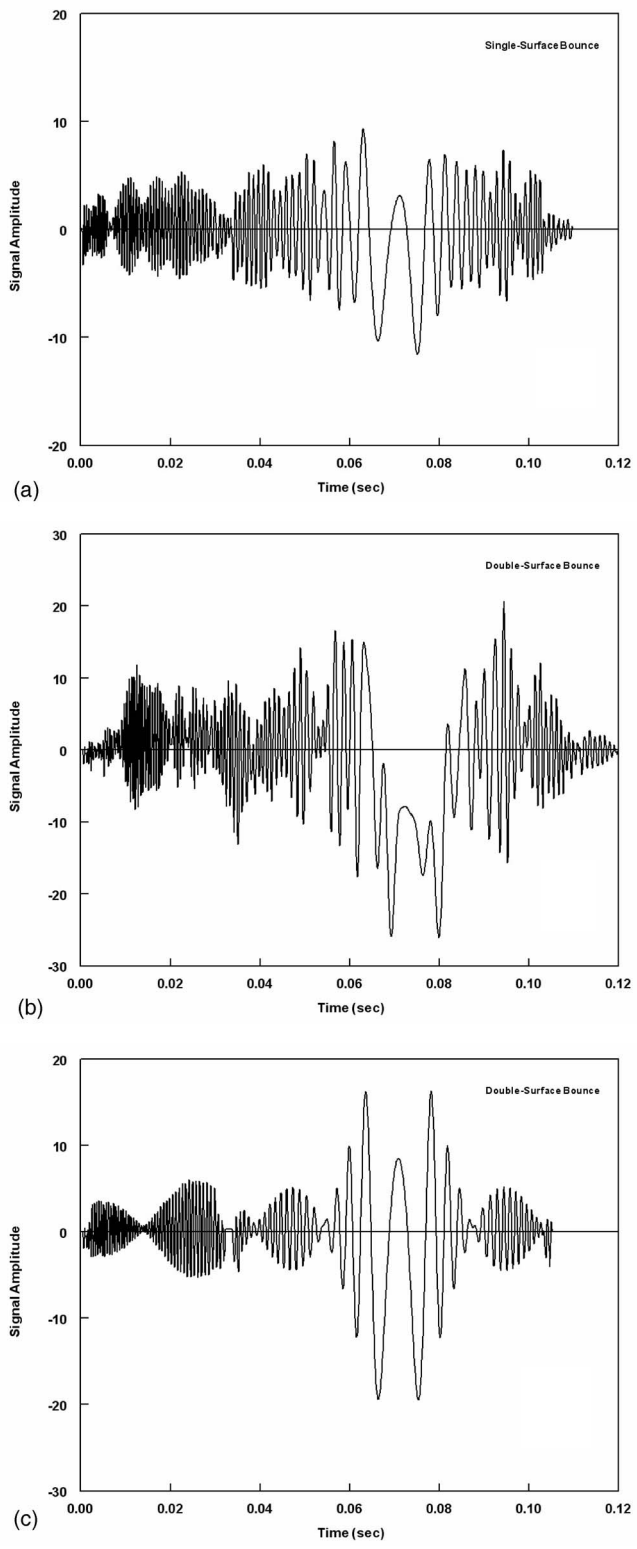


FIG. 7. A: accumulated sound levels due to the 37 reflected rays for single-surface bounce acoustic paths indicated in Fig. 2. B: accumulated sound levels due to the reflected rays for all 215 double-surface bounce acoustic paths indicated in Fig. 2. C: accumulated sound levels due to the reflected rays for the 37 shortest double-surface bounce acoustic paths indicated in Fig. 2.

more double-surface bounce reflections than single-surface bounce reflections is that the received double-surface bounce signal appears to be far less correlated with the original signal than the received single-surface bounce signal. The latter is simply a result of the superimposing at the location of the receiver of more and more signals with larger and larger shifts in time from that of the near-FSRP arrival. We note that a distorted version of the spreading of the oscillation of the source signal is seen in Fig. 7(b). As this is a major correlation feature in the signals, the structure of the signal Fig. 7(b) indicates that the largest matched filter peak will be calculated at a time of +0.075 s relative to the near-FSRP that started at 0.0 s.

If we only utilize the shortest 37 double-surface bounce ray paths in Fig. 2, we get the signal shown in Fig. 7(c). With just these limited number of signals, the correlation between the signals in Figs. 6 and 7(c) would indicate a maximum matched filter peak at the near-FSRP arrival time. From this we could conclude that signal distortion due to many multi-path arrivals over time for double-surface bounce rays can result in maximum matched filter peaks occurring after the first arrivals of the double-surface bounce rays.

VI. DISCUSSION

For tomography purposes, our reflection model provides important information related to the impact of ocean surface waves on acoustic signals between a source and receiver at frequencies greater than ~3 kHz:

(1) For a single ping, the number of multi-path rays between a source and receiver increases significantly with n, where n is the number of times the ray paths strike the ocean surface.

(2) In our test cases, there was always one of the multi-path rays that very closely followed the specular flat-surface ray path, the ray whose arrival time is required for the tomography scheme of L05. And certainly, if a technology utilizes 12–16 pings over a 30 s interval, one of the multi-paths will be almost identical to that of the FSRP. In all cases, the near-FSRP will be the first (or one of the first for double-surface bounce multi-paths) to arrive at the receiver.

(3) Because of the additional multi-path rays reflected from the ocean surface, the time frame beyond T_S over which acoustic energy from double-surface bounce rays arrive at a receiver is considerably longer than the time for single-surface bounce rays ($\Delta t_2 > \Delta t_1$).

(4) For standard source signal durations of 0.1–0.2 s and the configuration of source-receiver pairs used in L05, the recording of single- and double-surface bounce multi-paths does not exceed the source signal duration by a significant amount (i.e., the multi-path rays arrive at the receiver almost simultaneously). The receiver will first record the sound from the near-FSRP, but it will quickly start receiving sound from the other ray paths, resulting in interference with the signal from the near-FSRP.

As we have shown, maximum matched filter magnitudes for observations for double-surface bounce ray paths consistently occur after the first double-surface bounce arrivals are detected. This does not agree with the model implication that the near-FSRP (which we assume would have the maximum matched filter magnitude) should be one of the first double-surface bounce arrivals. The arguments in the previous sec-

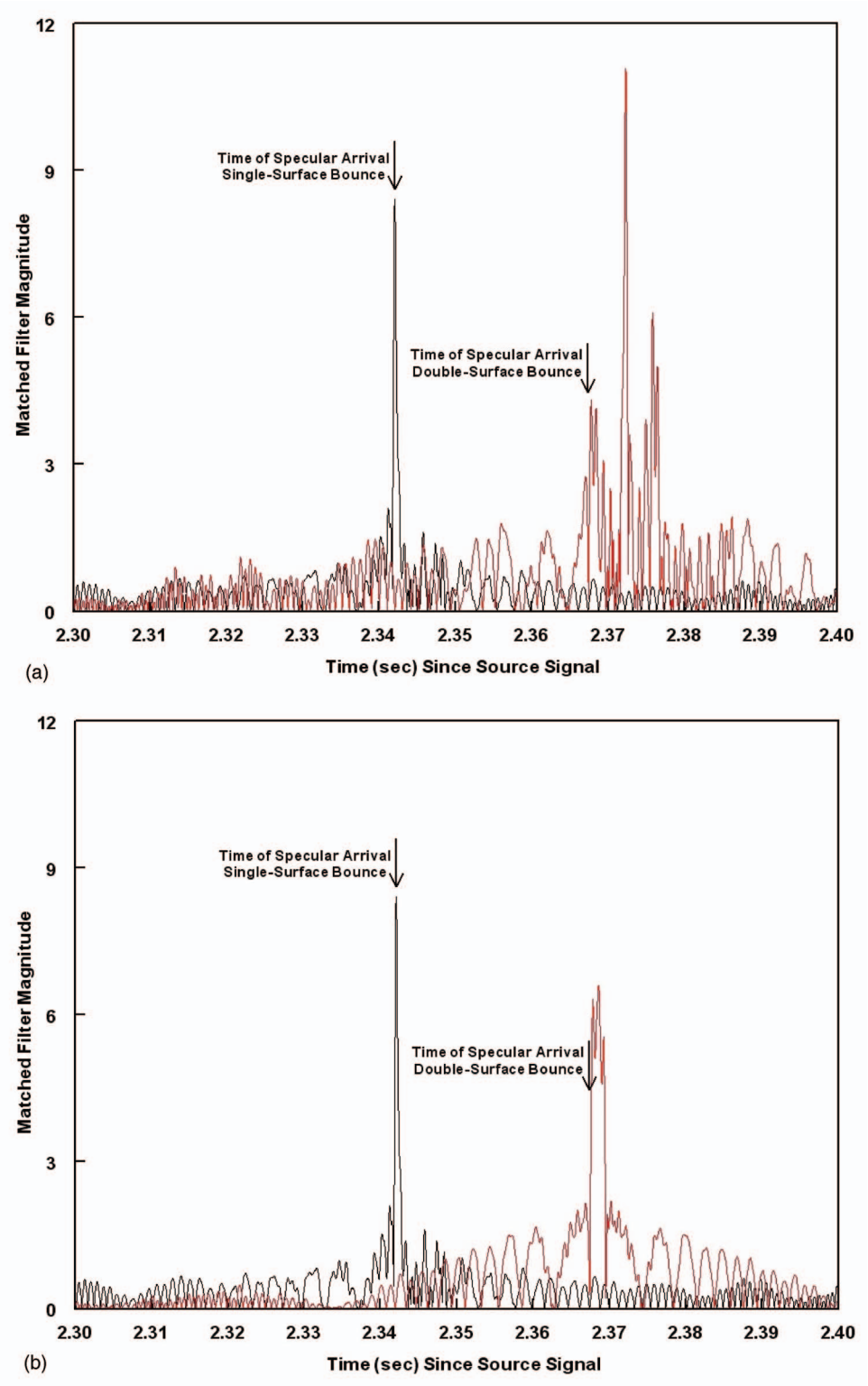


FIG. 8. Matched filter results between the source signal in Fig. 6 and the three synthetic received signals in Fig. 7. A: black—matched filter magnitudes for the single-surface bounce rays from Fig. 2; red—matched filter magnitudes for the double-surface bounce rays from Fig. 2. B: black—matched filter magnitudes for the single-surface bounce rays from Fig. 2; red—matched filter magnitudes for the shortest 37 double-surface bounce ray paths in Fig. 2.

tion provide a conceptual framework to explain this disagreement and help better interpret matched filter results in general. These concepts need to be verified, and this is done by match filtering the signals in Figs. 6 and 7.

The results are shown in Fig. 8. For single-surface bounce ray paths (Fig. 8(a)), our results indicate that the first matched filter peak (which should correspond to the near-FSRP according to the reflection model) will indeed be the maximum of all the matched filter peaks for the single-surface bounce multi-paths. Any following matched filter peaks will be for combined sound due to reflections that occur elsewhere along the ocean surface, for acoustic rays that travel different paths. It appears that the data we should utilize to calculate an arrival time and its variance for the single-surface bounce, flat-surface ray path are the maximum matched filter peak for all pings made for the ~30 s period used by L05 methodology. The cluster analysis technique used in L05 could be eliminated in favor of working with the

times of the matched filter maxima for each ping. With 12–16 pings over ~30 s, we would have 12–16 arrival times. However, the calculated arrival time should be weighted toward the first of the 12–16 arrivals, since the shortest ray path can be shown to be the near-FSRP.

Figure 8(a) also shows the matched filter magnitudes (red) between the signal in Fig. 6 and the double-surface bounce signal Fig. 7(b). As in the observations, the maximum matched filter magnitude occurs well after the time when the double-surface bounce near-flat-surface ray would arrive as well as other multi-path rays. In our hypothesis, this offset between the times of the maximum matched filter magnitude and the arrival of the double-surface bounce near-flat-surface ray would be a result of the overlapping of the many multi-path signals arriving at the receiver. Figure 8(b) shows the matched filter magnitudes (red) when we only consider the shortest 37 of all the double-surface bounce ray paths (Fig. 7(c)). As implied by our hypothesis, the elimination of the later double-surface bounce arrivals results in the maximum matched filter magnitude occurring much closer in time to the arrival of the double-surface bounce near-flat-surface ray.

Although our conceptual framework provides an explanation for the observed structure of double-surface bounce matched filter results, it does not suggest a means for precisely determining the double-surface bounce near-FSRP arrival. The near-FSRP arrival should be the first peak in the matched filter arrivals, but that peak will be small relative to the following peaks (see Fig. 5). The distortion of the receiver data is due to the sound from the large number of reflections arriving after the near-FSRP reflection, but there are no formulations that could be employed to remove the distortion that would not require prior knowledge of the individual signals causing the problem. In effect, we would need to utilize some *ad hoc* scheme that looks for a matched filter peak that occurs before the largest peak and is some set percentage of the largest peak.

Thus, for determining arrivals of FSRPs for tomographic uses, multi-path arrivals due to surface waves in the open ocean will often distort the received signal to the point that maxima of matched filtering magnitudes are not always a reliable indicator of the arrival time of the FSRPs. Here we found that, for source-receiver pairs separated by the order of kilometers, a correspondence between maximum matched filter magnitudes and the time of the arrival of the near-flat-surface ray works well for single-surface bounce rays but not for double-surface bounce rays. Additional simulations with the reflection model with the source-receiver pairs separated by 30 km resulted in approximately 440 wave-induced, single-surface bounce, in-plane multi-paths. The Δt_1 values ranged from 9.5 ms (when both source and receiver were at 150 m) to 29.6 ms (when one hydrophone was at 150 m while the other was at 500 m). Based on our results here, the correspondence between maximum matched filter magnitude and the arrival of the FSRP will likely exist for the former (shallower water) source-receiver configuration but not for the latter (deeper water) configuration.

ACKNOWLEDGMENTS

This work was supported by the Scientific Solutions, Inc., the Office of Naval Research, and the National Defense Center of Excellence for Research in Ocean Sciences. The author would like to thank I. Lucifrida for her help with some of the computations.

[1]P. H. Dahl, "On bistatic sea surface scattering: Field measurements and modeling," J. Acoust. Soc. Am. **105**(4), 2155–2169 (1999).
[2]J. K. Lewis, J. Rudzinsky, S. Rajan, P. J. Stein, and A. Vandiver, "Model-oriented ocean tomography using higher frequency, bottom-mounted hydrophones," J. Acoust. Soc. Am. **117**(6), 3539–3554 (2005).
[3]R. M. Heitsenrether and M. Badiey, "Modeling acoustic signal fluctuations induced by sea surface roughness," In *High Frequency Ocean Acoustics*, edited by M. Porter, M. Siderius, and W. Kuperman, AIP Conference Proceedings **728**, 214–221 (2004).
[4]P. H. LaBlond and L. A. Mysak, *Waves in the Ocean* (Elsevier, New York, 1978), p. 602.
[5]L. Culver and D. Bradley, "On the relationship between signal bandwidth and frequency correlation for surface forward scattered signals," In *High Frequency Ocean Acoustics*, edited by M. Porter, M. Siderius, and W. Kuperman, AIP Conference Proceedings **728**, 204–213 (2004).
[6]J. C. Reeves, "Distortion of acoustic pulses reflected from the sea surface," Ph.D. Dissertation, University of California, Los Angeles (1974).

Effects of low-frequency biasing on spontaneous otoacoustic emissions: Amplitude modulation

Lin Bian[a] and Kelly L. Watts
Auditory Physiology Laboratory, 3430 Coor Hall, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 25 July 2007; accepted 14 November 2007)

The dynamic effects of low-frequency biasing on spontaneous otoacoustic emissions (SOAEs) were studied in human subjects under various signal conditions. Results showed a combined suppression and modulation of the SOAE amplitudes at high bias tone levels. Ear-canal acoustic spectra demonstrated a reduction in SOAE amplitude and growths of sidebands while increasing the bias tone level. These effects varied depending on the relative strength of the bias tone to a particular SOAE. The SOAE magnitudes were suppressed when the cochlear partition was biased in both directions. This quasi-static modulation pattern showed a shape consistent with the first derivative of a sigmoid-shaped nonlinear function. In the time domain, the SOAE amplitudes were modulated with the instantaneous phase of the bias tone. For each biasing cycle, the SOAE envelope showed two peaks each corresponded to a zero crossing of the bias tone. The temporal modulation patterns varied systematically with the level and frequency of the bias tone. These dynamic behaviors of the SOAEs are consistent with the shifting of the operating point along the nonlinear transducer function of the cochlea. The results suggest that the nonlinearity in cochlear hair cell transduction may be involved in the generation of SOAEs. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2821983]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Bt, 43.64.Ld [BLM] Pages: 887–898

I. INTRODUCTION

An intriguing feature of our inner ear is that it can produce sounds even without acoustic stimulation. These self-generated tonal noises from the inner ear, called spontaneous otoacoustic emissions (SOAEs), have long been thought to reflect the active feedback from the outer hair cell (OHC) activities (Gold, 1948). It is speculated that these cells are on the verge of self-sustained oscillation to maximize their sensitivity. This indicates that at the auditory threshold the cochlear partition is under-damped and the cellular structures undergo autonomous resonance. Due to sporadically irregular arrangements of these cells, e.g., an extra row or a few missing ones (Lonsbury-Martin *et al.*, 1988), or perhaps a reduced efferent control from the upper auditory system (Braun, 2000), these oscillations from some particular locations in the cochlea become large enough and eventually escape the boundary of the inner ear, so that they can be recordable in the ear canal. When these oscillatory waves propagate out of the inner ear, part of the mechanical energy may be reflected at the stapes footplate in the middle ear back to their generating sites in the inner ear thus forming a standing wave that could be amplified after multiple roundtrips (Shera, 2003). Despite the complex mechanisms of SOAE generation, the presence of low-level SOAEs is naturally related to the normal function of cochlear OHCs.

Since the discovery of otoacoustic emissions (OAEs) in humans (Kemp, 1978), SOAEs are found in many species, e.g., frogs (Palmer and Wilson, 1982), monkeys (Martin *et al.*, 1985), guinea pigs (Ohyama *et al.*, 1991), and birds (Manley and Taschenberger, 1993). The drastic differences in the anatomical structures of inner ears between these species indicate that the modes of sound propagation in these ears are different. However, the functions of these different ears are the same, i.e., transferring mechanical energy into neural signals. A common element in these inner ears is the hair cell where acoustical vibrations are converted into electrical voltages, or mechano-electrical transduction. It has been found that the stereocilia of the reptile hair cells are capable of spontaneous movement (Martin *et al.*, 2003). Additional evidence, such as abolishment of SOAEs in humans (McFadden and Plattsmier, 1984) and lizards (Stewart and Hudspeth, 2000) by aspirin known for altering hair cell transduction, also points to a common process involved in the generation of SOAEs in different species. It seems likely that there may be a universal mechanism, i.e., hair cell transduction, responsible for the production of SOAEs. Studying the characteristics of SOAEs can provide a window for looking into normal auditory hair cell functions and their relation to the "active process" in the inner ear.

It is well known that the cochlear transduction is highly nonlinear as it is widely observed that suppression and distortion are typical when two tones with different frequencies ($f_1, f_2, f_1 < f_2$) are presented to the ear. Thus, using an external tone to influence the internally generated tonal sounds becomes a rather interesting approach to study the features of SOAEs (Zurek, 1981; Schloth and Zwicker, 1983; Rabinowitz and Widin, 1984). It has been shown (Frick and Matthies, 1988, Norrix and Glattke, 1996) that an external tone presented near the SOAE frequency can suppress the SOAE and generate distortion products (DPs). The effects of very low frequency external tones on SOAEs have not been studied, but it is known that these tones can produce suppression of other types of OAEs, such as transient evoked OAEs

[a]Author to whom correspondence should be addressed. Electronic mail: lin.bian@asu.edu.

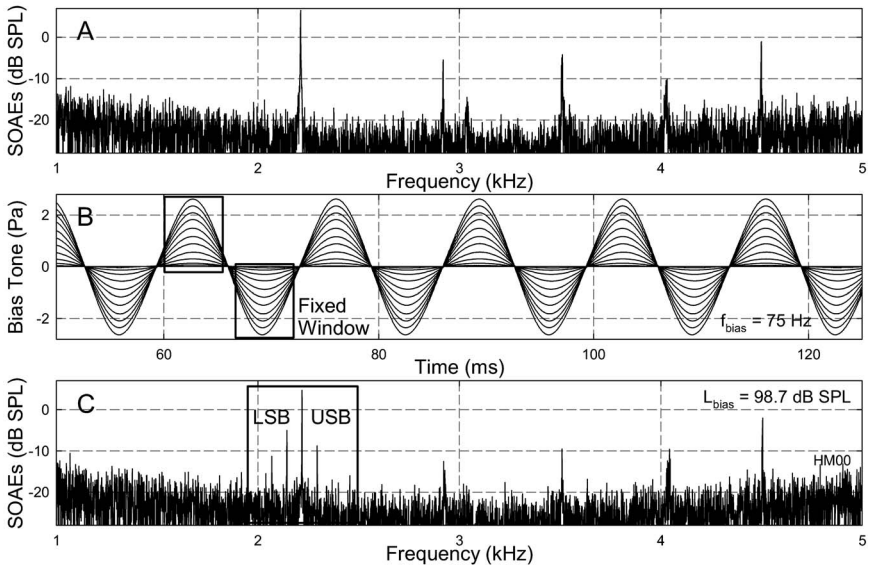


FIG. 1. Ear-canal acoustic spectra and bias tones. A: Spectrum of the ear canal acoustics without bias tone showing five major SOAE components. B: Samples of a 75 Hz bias tone with series of amplitudes. Rectangles indicate the fixed FFT windows centered at the peaks and troughs of the bias tones to obtain SOAE amplitudes for the quasistatic modulation pattern. C: Spectrum of an ear-canal acoustic signal with the presence of a bias tone showing modulation sidebands. A rectangular window centered at the SOAE frequency covering multiple sideband components around the SOAE was used to obtain a temporal SOAE envelope via IFFT. USB: upper sideband; LSB: lower sideband.

(Zwicker, 1981) and distortion products (Frank and Kössl, 1996; Scholz *et al.*, 1999; Bian *et al.*, 2002). Systematic investigations with a low-frequency tone revealed that it can produce an amplitude modulation (AM) of the distortion product otoacoustic emissions (DPOAEs) due to biasing of the cochlear partition (Bian, 2004 and 2006; Bian *et al.*, 2004). These manipulations of cochlear mechanics are consistent with the consequences of modulating a saturating nonlinear system which could be attributed to the OHC transduction. Quantifying the cochlear transducer function (F_{Tr}) using DPOAEs can provide a noninvasive means for acquiring information of cochlear mechanics which is critical for possible clinical applications (Bian and Scherrer, 2007). Since the SOAEs may reflect the inner ear mechanical activities below or at hearing threshold where the cochlear transducer gain is highest, it is hypothesized that biasing the cochlear partition could reduce the gain and thus modulate the SOAE magnitudes. If this was true, the relation between the AM of SOAE and the bias tone could allow a cochlear F_{Tr} to be derived. Therefore, to test the hypothesis, experiments were carried out on human subjects to study the effects of low-frequency biasing on SOAEs. Here, the dynamical effects on the SOAE amplitudes are reported.

II. METHODS

A. Experimental procedures

A total of 57 healthy students attending Arizona State University were screened for possessing large synchronized SOAEs (ILO92, Otodynamics) to participate in the study. The subject selection criteria were set as the following: (1) at least one SOAE in one ear was 20 dB above noise floor, and (2) the spacing between two adjacent SOAEs was greater than 300 Hz to avoid possible interference. All the participating subjects were verified to have normal hearing thresholds from 250 to 8000 Hz. Their middle- and outer-ear functions were normal based on a middle ear impedance test and an otoscopy. All these screening and pre-experimental tests were performed in the audiology clinic of the Department of Speech and Hearing Science. The experimental procedures were approved by the Institutional Review Board on human research subjects at the ASU. Data were collected from 11 ears from a total of ten qualified subjects. The ear-canal acoustical signal was measured with a calibrated probe microphone (Etymotic Research, ER-10B+) which was inserted into the ear canal with an immittance probe tip (ER-34). One port on the ER-10B+ system was coupled with a silicon front tube (ER1-21) and the other was plugged. The bias tone was produced from an 8.5 mm insertion earphone (HA-FX55, JVC) which was connected to the ER-10B via the front tube. The ear canal acoustic signal was recorded while presenting the bias tone at frequencies ranging from 25 to 100 Hz. At each bias tone frequency, the peak level of the bias tone was attenuated automatically from about 2 or 4 Pa to 0 in 41 steps. The whole recording session was repeated after a 20 min break. The purpose of the second trial was to check for variation in the SOAEs over time and the repeatability of the effects of the low-frequency modulation.

B. Signal processing and data acquisition

The ear-canal acoustic signal was recorded using a software developed in LabVIEW (v.8, National Instruments, NI) similar to previously described (Bian and Scherrer, 2007). Briefly, a 1-s-long bias tone with 10 ms onset/offset ramps and a 0.1 s flat tail (Fig. 1(b)) was output to a channel on a 24 bit dynamic signal acquisition (DAQ) and generation cards (PXI-4461, NI). The initial bias tone level (L_{bias}) was set by observing a complete suppression of the largest SOAE. The L_{bias} was attenuated automatically from the initial level to 0 Pa in 41 steps with smaller steps at higher intensities. For most subjects, the bias tone frequency (f_{bias}) was set at 25, 32, 50, 75, and 100 Hz in random order to modulate the SOAEs. Data were collected via an input channel on the same DAQ card. Both the input and output channels were synchronized with an onboard clock and triggered with a digital pulse. For each biasing step, the ear-canal acoustic signal was amplified 20 dB by the build-in preamplifier on the ER-10B+, averaged 8–12 times depending on

TABLE I. Characteristics of qualified subjects.

Subject	Gender	Sides	Side(s) used	Age	Num of SOAEs
HM00	F	Both	R	27	5
HM03	F	Both	R	22	5
HM16	F	Single	R	23	1
HM17	F	Both	R	24	2
HM25	M	Both	L, R	34	3, 5
HM36	F	Both	L	18	4
HM46	F	Both	L	19	2
HM47	M	Single	R	26	1
HM48	M	Both	L	19	2
HM54	M	Both	R	28	5

the size of the SOAE, and digitized at 204.8 kHz for offline analysis in MATLAB (MathWorks).

C. Data analysis

The spectral feature of the SOAEs was extracted from a fast Fourier transform (FFT) of the high-pass filtered (at 400 Hz) ear-canal acoustic signal (Figs. 1(a) and 1(c)). Since the low-frequency modulation of SOAEs is reflected by the presence of lower and upper sidebands (LSBs and USBs) around the emission components, the amplitudes of two LSBs and two USBs were measured for each OAE component. System distortions could be ruled out, because no sidebands could be observed in a 2-cc coupler when a 2 kHz probe tone was presented below 50 dB sound pressure level (SPL) with the presence of a bias tone at 106 dB SPL. The quasi-static modulation patterns (Bian and Chertoff, 2006) of SOAEs were derived from short-time FFTs with windows fixed at the peaks and troughs of the bias tone (Fig. 1(b)) assuming static cochlear responses at these moments. The length of the FFT window was so determined that it was shorter than 1/2 biasing cycle and the SOAE frequency was centered in a frequency bin. The SOAE amplitudes for two adjacent bias tone peak or trough sequences were paired to form the positive and negative parts of the quasi-static modulation pattern. These patterns then were averaged across 22–88 pairs of peaks and troughs for different biasing frequencies. The dynamic or temporal features of the SOAEs were obtained from a spectral windowing method (Fig. 1(c)). For a specific SOAE, the complex spectrum of the ear canal signal was rectangular windowed at the positive SOAE frequency including 3–4 sidebands on each side and inverse fast Fourier transformed (IFFT) to recover the amplitude envelope of the SOAE. The instantaneous SOAE amplitude was examined with the bias tone waveform obtained from low-pass filtering at 400 Hz.

III. RESULTS

A. General characteristics of SOAEs

From the 114 ears of 57 subjects who were screened for SOAEs, 11 ears from ten subjects were qualified for participating in the experiments. As can be noted from Table I, there were six females and four males with ages ranging from 18 to 34 (mean age: 24). Most of them, except two, had SOAEs in both ears. Except subject HM25, low-frequency modulation of SOAEs was measured from only one side. The data were collected from the right ears of seven subjects and the left ears of four subjects. Since the characteristics of SOAEs were highly dependent on each individual ear, the features of SOAEs are described on a case by case basis. Among the 11 ears (Table I), the number of SOAEs in each ear ranged from 1 to 5 components with an average of 3 SOAEs/ear. Averaged across ears, the absolute differences between the SOAE amplitudes obtained at the two trials were less than 1.75 dB. If we consider the change as an increase or a decrease for each component, the average change in SOAE amplitudes was a negligible decrease (−0.02 dB). Therefore, the data obtained at the two trials were combined in reporting. There was also a frequency shift during the low-frequency biasing which will be reported in a separate paper. In the present report, only the effects on the amplitudes of the SOAEs are considered.

The amplitude and frequency distributions of these SOAEs averaged across the two trials for each ear are illustrated as a form of "SOAE-gram" (Fig. 2). As can be seen, the amplitudes of SOAEs ranged from −7 to 13 dB SPL across subjects. Within a single subject, the variation of SOAEs amplitude was smaller, and the amplitude difference between two adjacent SOAEs was less than 15 dB. The amplitude of SOAEs averaged across all components was greater than −3 dB SPL. The frequency range of the SOAEs covered from about 900 to 5500 Hz. For subjects with multiple SOAE components, the averaged frequency spacing was about 764 Hz with a minimal spacing of as small as 290 Hz in the left ear of subject HM25. From Fig. 2, there is a trend that can be observed, i.e., the SOAE amplitudes were higher in the low frequencies and lower in the higher frequencies showing a form of a low-pass filter. The slope of the low-pass filter was roughly −6 dB/octave.

B. Spectral effects

1. Suppression of SOAE and generation of sidebands

In the spectral domain, the effects of low-frequency biasing were expressed as reductions in SOAE amplitudes and generations of sidebands (Fig. 3). When the L_{bias} was high enough, the SOAE could be completely suppressed (top panels). While the L_{bias} was decreased, the SOAE amplitude started to recover and multiple sidebands appeared on both

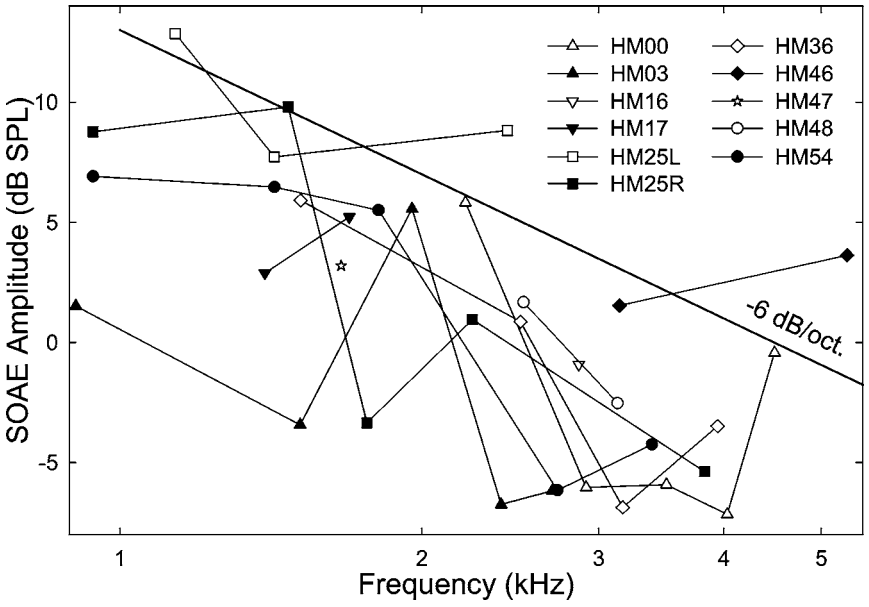


FIG. 2. SOAE-grams. The amplitudes and frequencies of the SOAEs for each subject are displayed with a line marked by symbols. Data represent an average of the two trials. A trend can be observed from the data: SOAE amplitudes are inversely related to their frequencies. The dark line indicates the slope of the trend: −6 dB/octive.

sides. The number and sizes of the sidebands varied depending on the signal conditions, such as, the L_{bias}, the f_{bias}, and the original SOAE level. There could be single or double sidebands on each side of the SOAE component (Fig. 3 middle row) indicating different temporal modulation characteristics. The sidebands presented at the spectral positions of integer multiples of the f_{bias} from the SOAE frequency. In most of the observations, the sidebands that had a spectral spacing of f_{bias} from the SOAE component (sidebands I) were the largest in magnitude. The sidebands $2f_{bias}$ from the SOAE in either LSB or USB were smaller than sidebands I, but larger than other more distant sidebands. Thus, the analysis was focused on these two sidebands. As the L_{bias} further decreased, the SOAE amplitudes fully recovered, and sidebands also diminished (bottom).

These opposite changes in the amplitudes of SOAE and its sidebands as functions of increasing L_{bias} were better observed in Fig. 4. As the example showed, the SOAE amplitude remained relatively stable with less than 5 dB fluctuations when the bias tone was below 100 dB SPL. Above this L_{bias}, the SOAE amplitude became suppressed (top). The rate of decline in the SOAE depended on the f_{bias}, i.e., faster for higher biasing frequencies than lower frequencies. For example, the SOAE amplitude reduced at a 3 dB/dB rate at 100 Hz f_{bias} compared to 1 dB/dB at 25 Hz. The growths of sidebands started at much lower biasing levels (lower four panels). The LSB I began to increase at about 80 dB SPL, and reached a peak at 102 dB SPL L_{bias} while USB I peaked at 98 dB SPL. The peak amplitudes of sidebands I (usually −5 to −10 dB SPL) were higher than those of sidebands II

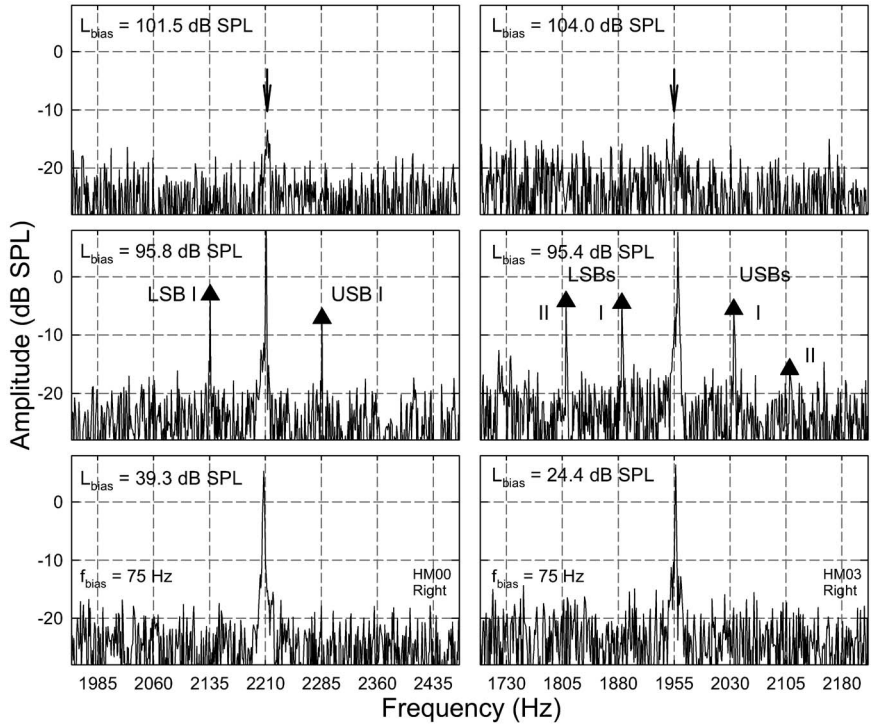


FIG. 3. Spectral representation of low-frequency biasing on SOAEs: suppression and modulation. Each column corresponds to a subject. Panels at the bottom show that there is no effect on the SOAE when the L_{bias} is very low. Middle row indicates the presence of AM of the SOAEs at moderately high L_{bias} indicated by the generation of spectral sidebands around the SOAE components. Sidebands are labeled as I and II based on their frequency spacing with the SOAE. Note: the difference in the number of sidebands between the two subjects. Top row: suppression of the SOAEs when the L_{bias} is very high. Arrows point to the suppressed SOAEs.

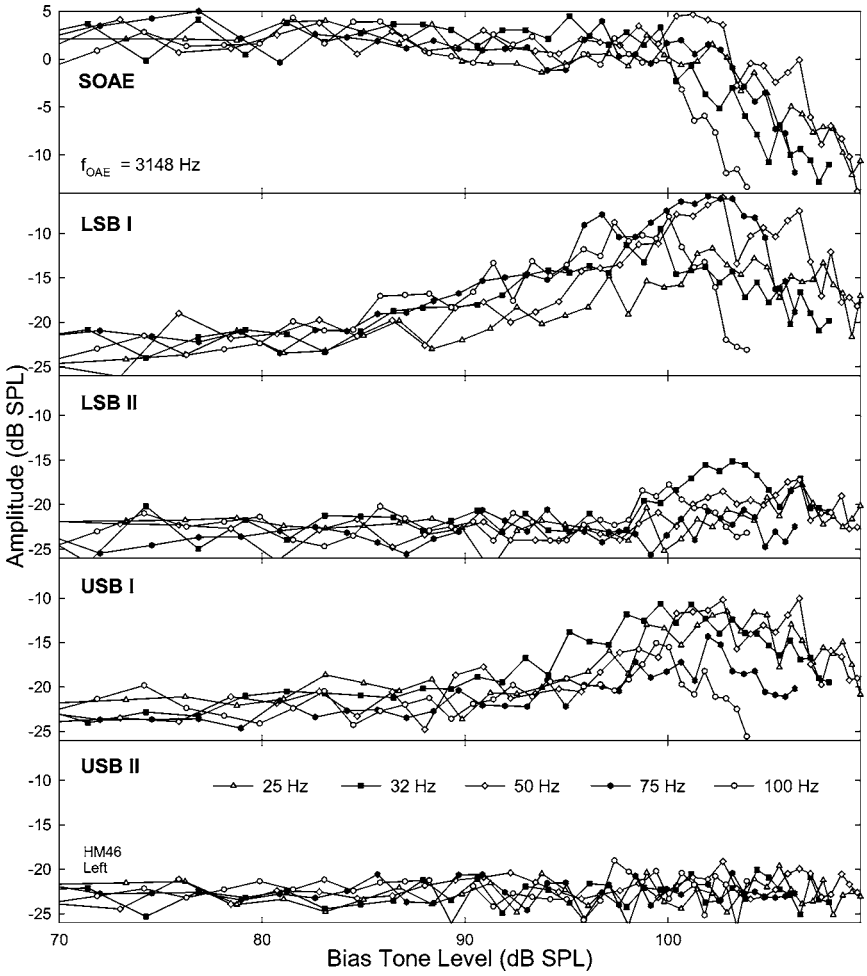


FIG. 4. Opposite biasing effects on the magnitudes of an SOAE and its sidebands. Top panel: suppression of SOAEs as L_{bias} exceeds about 100 dB SPL. Each line represents an average across the two trials. Lower four panels show the effect of L_{bias} on the SOAE sidebands. The sideband amplitudes increase with L_{bias} and reach their maximal value above 100 dB SPL biasing level where the SOAEs are significantly suppressed. Note the differences between various bias tone frequencies.

(<−10 dB SPL), and the LSBs were larger than their counterparts in USB. For about a 10–15 dB range of L_{bias} above 80 dB SPL, the sidebands increased at rates slower than 1 dB/dB without suppressing the SOAE noticeably. After the sidebands reached their peaks, it was evident that the SOAE often reduced more than 3–5 dB. When the SOAE was drastically suppressed and diminished, the sidebands showed a roll-over and eventually fell back to the noise floor. The biasing levels corresponding to the onsets and the peaks of the sidebands were lower for higher biasing frequencies, suggesting that these bias tones were more effective in suppressing and modulating the emissions.

2. Effects of f_{bias} and SOAE frequencies

Since the SOAEs were larger and less variable compared to their sidebands, the effects of the f_{bias} were examined by measuring the suppression of the SOAEs. Because the peaks of the sidebands corresponded to a 3–5 dB reduction in the SOAE amplitude, a 5 dB suppression of SOAE was used as a criterion to evaluate these frequency effects. For each SOAE, the L_{bias} required to produce a 5 dB reduction in the SOAE magnitude was obtained from the SOAE suppression curve like those in the top panel of Fig. 4. For each subject, the values derived from the two trials were averaged to form a 5 dB iso-suppression contour representing the effectiveness of the bias tone in suppressing and modulating the SOAEs at different frequencies (Fig. 5). A general trend could be observed that it required higher sound pressures to induce the same extent suppression of SOAEs at higher frequencies. This trend held for all biasing frequencies regardless the sizes of SOAEs. The slope of the iso-suppression curve seemed shallower for larger SOAEs. However, for most subjects with modest SOAE magnitudes the iso-suppression curves approached a slope of about 8 dB/octave with increasing SOAE frequency. If observing each subject closely, it could be found that the SOAE suppression curves shifted downwards as the f_{bias} increased. This indicated that lower frequency bias tones were less efficient in suppressing the SOAEs, thus higher biasing levels were used.

Because the extent of temporal AM in SOAE can be represented by the magnitudes of the sidebands, the peak levels of sidebands I were measured for all SOAE components to form a sideband vs. SOAE frequency function (Fig. 6). The maximal sideband I amplitudes for different SOAEs decreased with increasing SOAE frequencies. Except the upper end of the SOAE frequency range, the slope of these sideband-frequency functions approached a slope of −6 dB/octave, similar to the SOAE-grams (Fig. 2). For frequencies greater than about 4 kHz, the slope became shallower, because of the smaller SOAE amplitudes and limited power of the low-frequency bias tone in this region. Compared to Fig. 2, where the amplitudes of the SOAEs were more random for a single subject, the maximal sidebands for

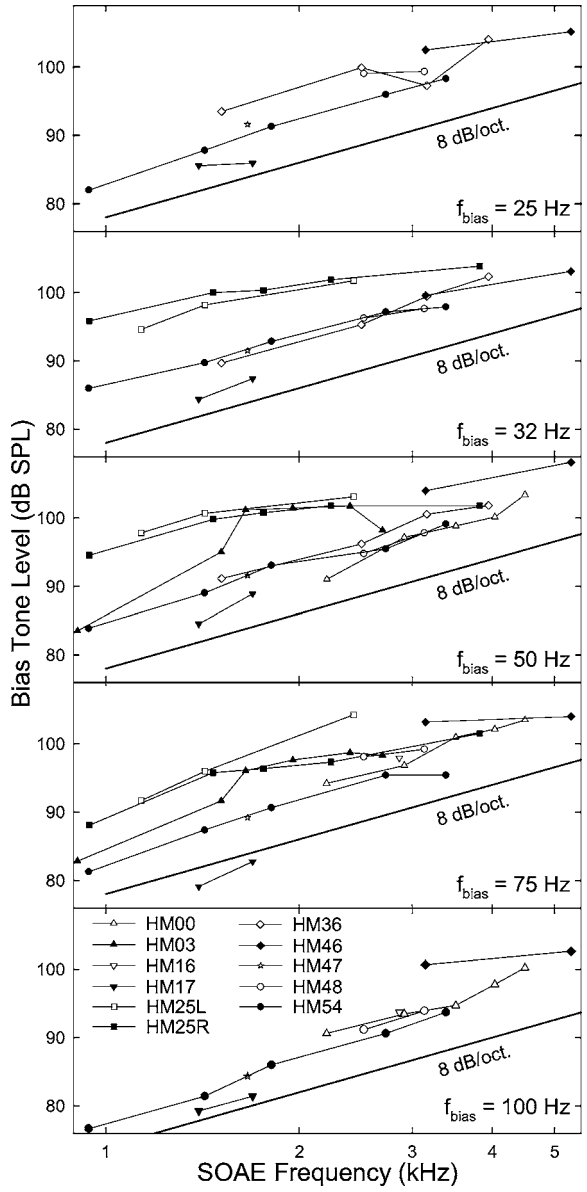


FIG. 5. 5 dB iso-suppression curves. Each data point represents the L_{bias} required to produce a 5 dB reduction in the SOAE amplitude. The curves imply the efficiencies of the bias tone on SOAEs at different frequencies. For a bias tone with fixed frequency, it is more effective in suppressing SOAEs with lower frequencies. The change in the ability of the bias tone to suppress the SOAEs is predicted by the darker line in each panel which has a slope of 8 dB/octave. Data reflect an average across the two trials.

these SOAEs tended to decline more monotonically towards high frequencies. In other words, these maximal sidebands were not quite parallel to their corresponding SOAE-grams. This seemed to suggest that the sizes of the sidebands were more related to the influence of the bias tone than the SOAE amplitude.

C. Mechanical effects

1. Quasi-static modulation patterns

The influence of mechanical biasing of the cochlear partition on the SOAE amplitudes was examined by measuring the modulation patterns. Two types of SOAE modulation patterns were derived: quasi-static and temporal modulation patterns. The quasi-static modulation pattern represented the effect of biasing the cochlear partition in different directions as the SOAE magnitude was measured at opposing extremes of the bias tone. This modulation patterns generally demonstrated a bell shape which varied with both SOAE and bias tone frequencies (Fig. 7). The SOAE magnitude showed a maximum near 0 Pa biasing pressure and reduced with increasing the pressure in either direction. At the peak region of the modulation pattern, the SOAE was unstable and showed an up to 6 dB fluctuation. When the SOAE became suppressed, the amplitude fluctuation was also limited. The maximal reduction in SOAE amplitude or modulation depth, represented by the difference between the peak and trough of the pattern, was often more than 10 dB. For a single bias tone of a fixed frequency, the modulation pattern varied with the frequencies of SOAEs (Fig. 7 top panels). The modulation pattern became flatter and wider towards higher SOAE frequencies. This is consistent with the limited power of the low-frequency bias tone in a high-frequency region where the suppression only occurred at very high biasing levels. Focusing on a single SOAE component (Fig. 7 bottom), there was an effect of the f_{bias} on the quasi-static modulation pattern. The reduction of SOAE amplitude became more abrupt and the modulation patterns were narrowed for higher f_{bias}. Such a variation in the modulation pattern was due to the more effective suppression by higher-frequency bias tones.

Suppression of SOAE when the cochlear partition was displaced in either direction was consistent with a saturating nonlinearity where the gain is reduced at the extremes of the input. The bias-tone induced motion of the cochlear partition could most likely affect the hair cell transduction channels. Thus, the shape of the modulation pattern was self-explanatory for the cause of SOAE suppression, i.e., the modulation of the cochlear transducer gain by shifting the operating point. Therefore, the modulation pattern was modeled and fit with the first derivative of the hair cell F_{Tr} in the form of a second-order Boltzmann function (Bian *et al.*, 2002):

$$G = -A \cdot M \frac{b + (b+d)N}{[1 + M(1+N)]^2},$$
$$M = e^{bx-c} \text{ and } N = e^{dx-e}, \quad (1)$$

where G is the first derivative of a Boltzmann function $F_{Tr} = A/[1+M(1+N)]$, x is the cochlear partition displacement represented by the biasing pressure at the peaks and troughs, A is a scaling factor, b and d are parameters relating to the slope of the transducer curve, c and e are constants setting the transducer operating point. Representations of the model fit and the F_{Tr} parameters for each f_{bias} are illustrated in Fig. 8. The correlation coefficients (r^2) for the fits ranged from about 0.7 to 0.9. The derived Boltzmann functions (right panels) all consisted of a sigmoid shape with slight variations in slope and symmetry.

2. Temporal (period) modulation pattern

Another aspect of mechanical alteration of SOAE was the temporal modulation of emission amplitudes. Because the SOAE magnitudes were quite small (<5 dB SPL in most cases), the SOAE envelopes obtained from IFFT were subject to contaminations from random noise and acoustic emissions from multiple reflections in the cochlear capsule. As

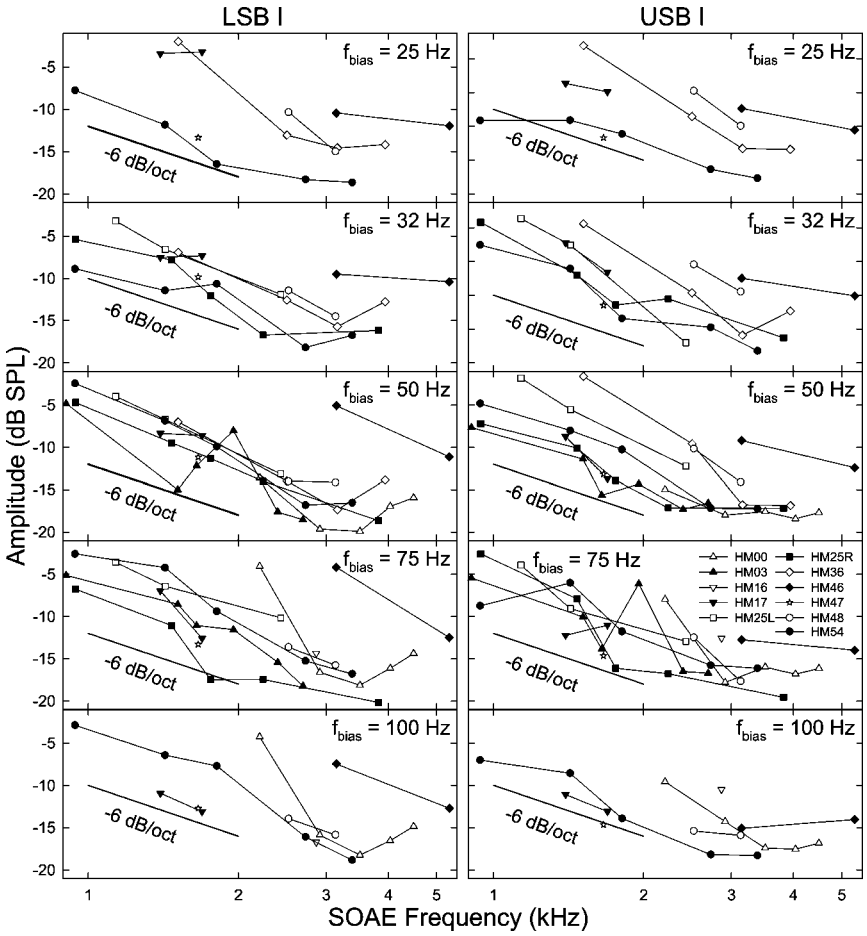


FIG. 6. Maximal sidebands I amplitudes. Each data point reflects the highest sideband amplitude a bias tone can produce. SOAEs with lower frequencies show larger sidebands. This correlates with the distribution of the original SOAE amplitudes (Fig. 2). The dark line in each panel indicates the slope of this trend (−6 dB/octave). Data reflect an average across the two trials.

shown in Fig. 9(a), the SOAE appeared to be noisy and irregular. With careful comparison with the bias tone (lower trace), it was not difficult to notice that the SOAE envelope peaked twice within each biasing cycle starting with a trough. Such a periodicity was confirmed by a spectral analysis of the SOAE envelope (Fig. 9(b)). Two large peaks appeared at the f_{bias} (50 Hz) and $2f_{\text{bias}}$ (100 Hz) suggesting that there was a regularity repeating in one and one half basing cycle. To reduce the noise and other contaminations, segments of the SOAE envelope corresponding to 20–86 biasing periods were averaged to produce a period modulation pattern (Fig. 9(c)). Since only the bias tones with high levels can effectively modulate the emissions, the period modulation patterns obtained from the top 20 biasing levels (a 15 dB range) were averaged across the two trials and examined. The most typical period modulation pattern consisted of two

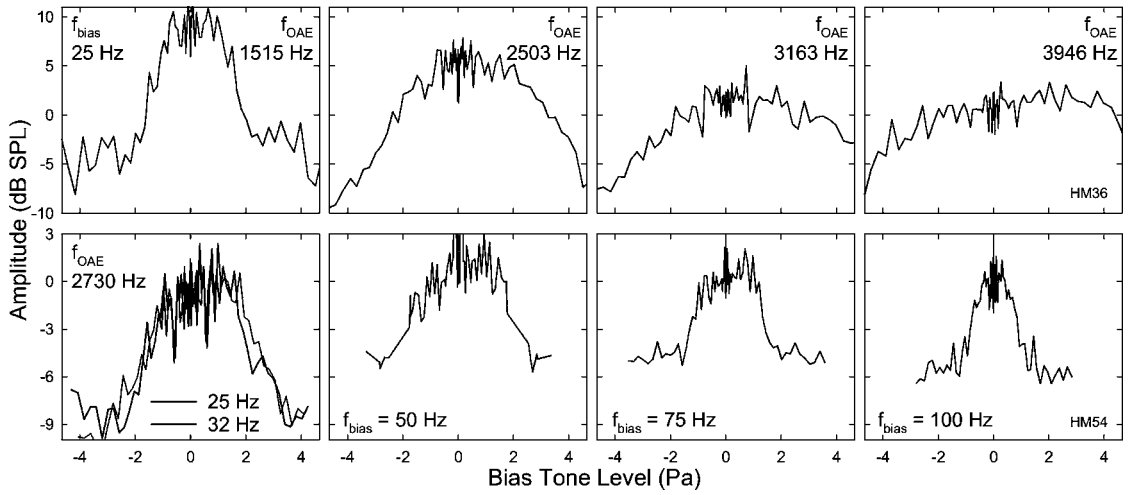


FIG. 7. Quasi-static modulation patterns. Top row: effects of a fixed bias tone on SOAEs with different frequencies. The amplitudes of SOAEs show a bell shape. The modulation depth (height of the pattern) reduces and the width of the pattern increases with the SOAE frequency. Bottom row: effect of f_{bias} on the modulation pattern. The modulation width becomes narrower with the increase in f_{bias} because of the increased relative strength of the bias tone with respect to the SOAE. Data reflect an average across the two trials. Note: the fluctuations of SOAE amplitudes at the peak regions of the modulation patterns.

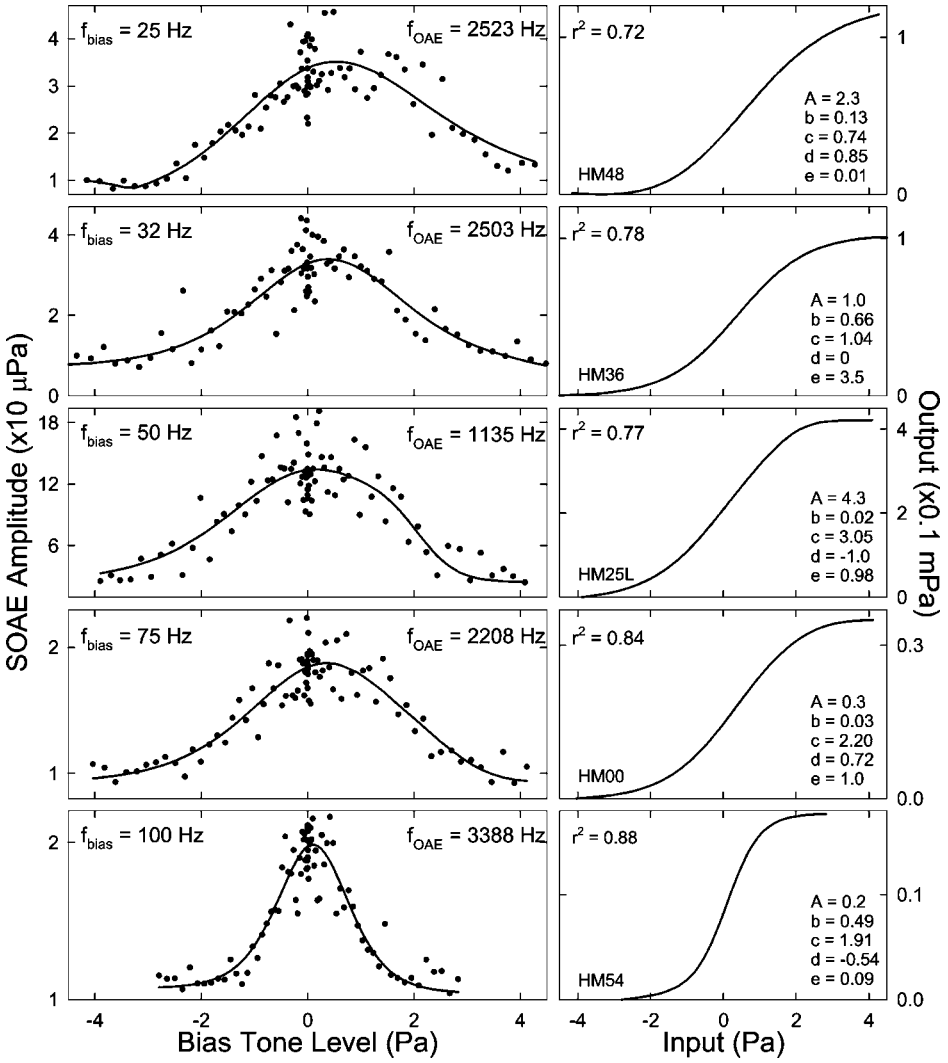


FIG. 8. Quasi-static modulation patterns and cochlear F_{Tr}. The SOAE amplitudes as a function of the peak level of the bias tones at each frequency (scattered points) are fit with the first derivative of a second-order Boltzmann function (Eq. (1)). The correlation coefficients (r^2) of the fits (solid lines) range from 0.72 to 0.88. The shapes of the derived cochlear F_{Tr} are similar despite small variations. The obtained Boltzmann parameters are indicated in the right panels. Data reflect an average across the two trials.

peaks each corresponding to the zero crossings of the bias tone with a short delay. Each half of the period modulation pattern marked by a large peak was similar to the bell-shaped first derivative of the cochlear F_{Tr} (Fig. 8 left).

As the L_{bias} reduced, the SOAE period modulation patterns showed progressive variations (Fig. 10). The most noticeable change was the decrease in the depth of the notch between the two SOAE peaks which was induced by the positive biasing extreme (lower right panel). As a result of the decreased suppression of SOAE at the positive biasing peak, the two SOAE peaks began to merge (top curves). Moreover, the SOAE still remained to be suppressed at the

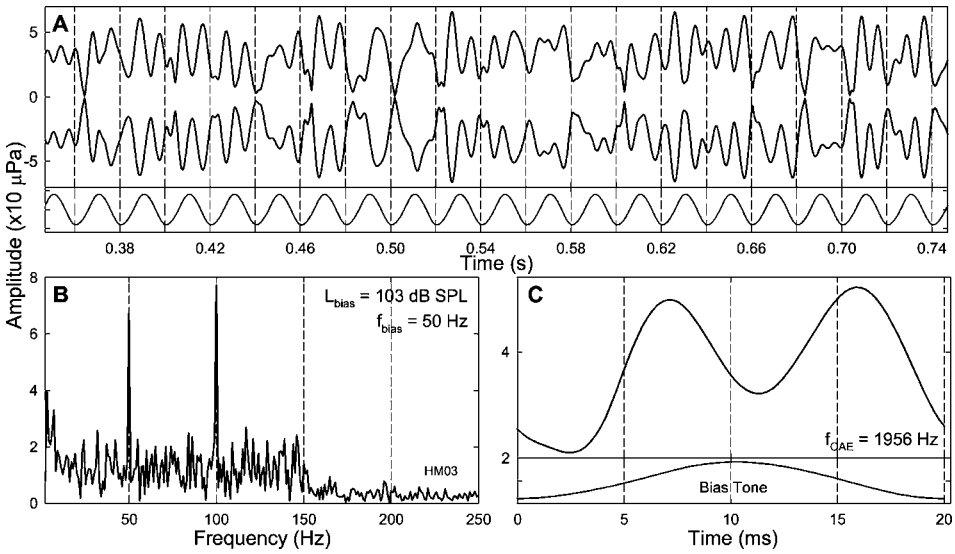


FIG. 9. Temporal modulation of SOAE. A: Temporal envelope obtained from IFFT of the spectral contents around the SOAE. Referenced to the bias tone (lower trace), it is noticeable that two SOAE peaks present within one biasing cycle (between two vertical dashed grid lines). B: Spectrum of the SOAE envelope in panel A. Note: two large peaks present at f_{bias} (50 Hz) and $2f_{bias}$ (100 Hz) indicating that a temporal pattern repeats once and twice in every biasing cycle. C: Period modulation pattern derived from averaging the SOAE envelope over 44 biasing cycles. The SOAE amplitude shows two peaks each correlating to a zero crossing of the bias tone (lower trace).

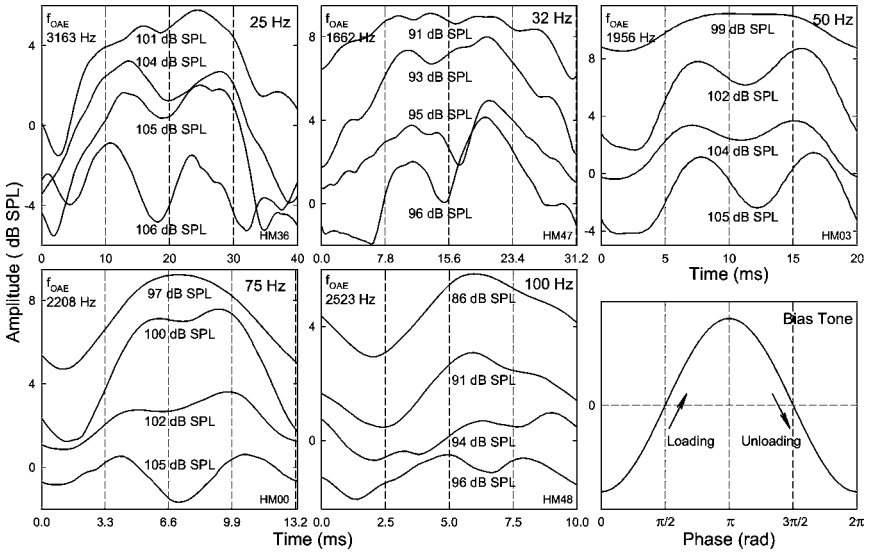


FIG. 10. Period modulation patterns. Note: the systematic change in the period modulation pattern of SOAE as the L_{bias} varies in each panel. As L_{bias} decreases (from bottom trace to top), the SOAE amplitude increases with merging of the two peaks. Each half of the bottom traces is similar to the first derivative of the Boltzmann function. The bias tone phase is shown in the lower right panel for a reference. Loading: a monotonic increase in biasing pressure; unloading: a monotonic decrease in the instantaneous biasing pressure.

negative biasing extremes, but with smaller modulation depths. These different modulation patterns could be a source for the variation in the number and sizes of the sidebands around the SOAE component (Fig. 3). The single-peaked period modulation pattern was similar to a simple sinusoidal AM signal which contained only one sideband on each side of the SOAE. Using the bias tone as a reference, it was noted that the delay of the second SOAE peak appeared smaller and even lead the zero crossing of the falling biasing pressure.

The effect of the f_{bias} on the SOAE period modulation pattern was a transition from a double peaked pattern to a single peaked one (Fig. 11). At high bias tone levels, as the f_{bias} increased, the first SOAE peak corresponding to the loading of cochlear partition became considerably delayed and merged with the second peak (Fig. 11 middle panels and Fig. 10 lower middle panel). At the highest f_{bias} (100 Hz), the first SOAE peak was suppressed so much that it never recovered. In this case, the period modulation pattern only contained a single SOAE peak that was related to the unloading of the cochlear partition. Thus, for fast biasing of the cochlear structures, increasing in the biasing pressure or loading would produce more suppression than decreasing the pressure or unloading. Consider the sequence of events within one biasing cycle, the delay of the first peak was due to the prolonged suppression from the negative maximal biasing pressure, and biasing in the positive direction was less effective. This unevenness in SOAE suppression and modulation depending on the directions of cochlear partition motion was a consequence of the asymmetry in cochlear transduction (Figs. 7 and 8) where biasing in the negative pressure direction yielded smaller SOAE amplitudes which could be translated into more compression in the cochlear F_{Tr}.

IV. DISCUSSION

A. Spectral features of SOAEs

The general properties of the SOAEs measured in the present study include (1) 1/10 of the subjects, esp., females, have large SOAEs, (2) if present in an ear, usually there are on average about three SOAE components, (3) for majority (64%) of the subjects, SOAEs are found in both ears, with a higher occurrence in the right ears, and (4) the SOAEs frequencies are distributed in the range from 0.9 to 5 kHz in a spectral shape of a low-pass filter with a −6 dB/octave slope (Fig. 2). These descriptive features are in line with many observations (see Probst *et al.*, 1991 for a review). The observed spectral characteristic of a low-pass filter in SOAE amplitudes is consistent with the studies in infants, young, and older adults (Lonsbury-Martin *et al.*, 1991; Braun, 2006). The 6 dB/octive reduction rate of SOAE amplitude

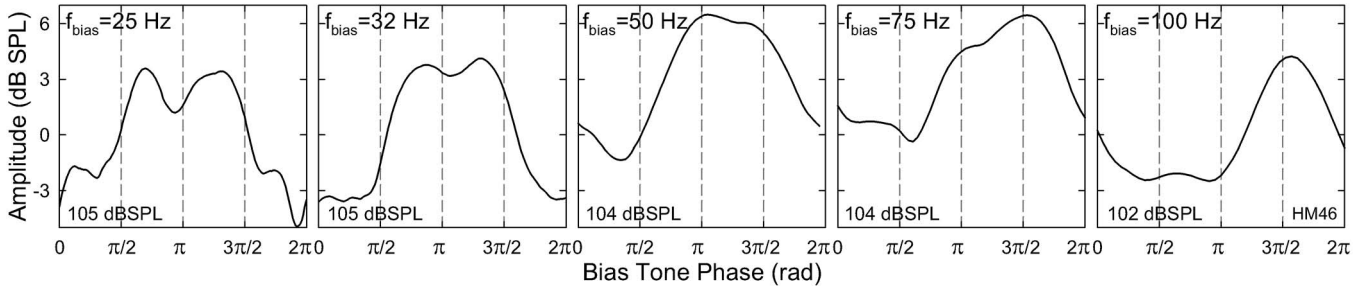


FIG. 11. Effect of the f_{bias} on the period modulation pattern. As the f_{bias} increases, the period modulation pattern demonstrates a merging of the two peaks, further delay and suppression of the first peak. This represents a transition from the double peaked pattern to a single peaked one which implies a reduction in the number of sidebands in the frequency domain. Note: at 100 Hz f_{bias}, the SOAE remains to be suppressed throughout the loading process and released during unloading. The first suppression phase is a delayed effect of displacing the cochlear partition in the negative pressure direction.

with increasing frequency is in close agreement with the results of some large scale investigations (Moulin *et al.*, 1993; Penner *et al.*, 1993). It is naturally supposed that this filtering effect is a result of the reverse transmission of the middle and outer ears. However, the middle and outer ear are bandpass filters with center frequency located around 2–3 kHz (Aibara *et al.*, 2001; Whittemore *et al.*, 2004). Anatomic evidences showed that the cochlear hair cell irregularities, such as extra or missing OHCs and microstructural changes in these cells, are prominent in the low- and mid-frequency region (Lonsbury-Martin *et al.*, 1988). From a developmental perspective, there could be more immature OHCs that are innervated with afferent or reciprocal fibers in the apical cochlear partition (Pujol, 2001; Thiers *et al.*, 2002) where the efferent synapses are sparse (Guinan *et al.*, 1984). The lack of efferent control over these OHCs in the low-frequency region could be the basis for the generation of SOAEs and the low-pass filter shape in their frequency-amplitude distribution. This could account for the developmental reduction in SOAE amplitudes and downward shift in SOAE frequencies (Burns *et al.*, 1994; Braun, 2006). Therefore, the appearances of the SOAEs provide clues for their underlying generating mechanisms.

B. Suppression and modulation: A nonlinear mechanism

Until the present study, there has been no report on the low-frequency modulation of SOAEs. The major findings of the present study are the suppression of SOAE amplitudes, the generation of multiple sidebands around SOAE components, and the phase-dependent AM of SOAE. These results are parallel to studies on low-frequency biasing of DPOAEs in both experimental animals (Frank and Kössl, 1996; Bian *et al.*, 2002, 2004; Bian, 2004, 2006) and humans (Scholz *et al.*, 1999; Bian and Scherrer, 2007). The details of the suppression and modulation of the SOAEs, in particular, are quite similar to the results obtained from the cubic difference tone ($2f_1$-f_2, CDT). For example, suppression with L_{bias}, presence of single or multiple sidebands, the bell-shaped quasi-static modulation pattern, and the two peaked period modulation pattern are common in both SOAE and CDT. Similar phenomena have been found with other types of OAEs, such as stimulus-frequency OAEs (Neely *et al.*, 2005) and tone-burst evoked OAEs (Zwicker, 1981). In the latter study, "suppression period patterns" with a progressive change from two amplitude maxima to a single peak when decreasing the level of a low-frequency suppressor were observed. This observation is similar to the period modulation patterns shown in Fig. 10. This suggests that these different OAEs may share the same generating mechanisms: the nonlinear transduction processes of the cochlear OHCs.

In addition, this generating mechanism can be evident by the nonlinear interference between externally presented acoustic signals and the SOAEs. For example, investigations on the interaction of an external tone with SOAEs have shown that they can be suppressed (Zurek, 1981; Rabinowitz and Widin, 1984; Bargones and Burns, 1988) and DPOAEs can be generated (Frick and Matthies, 1988; Norrix and Glattke, 1996), esp., when the external tone frequency is below the SOAE. One finding of the present study is that the suppression of SOAEs occurs more effectively if the bias tone is closer in frequency to the SOAE (Figs. 4 and 7). This reflects that the SOAEs may be generated locally on the cochlear partition due to the tonotopic tuning property and the hair cell transducer channels are saturated when the external stimulus reaches the SOAE generating site. The results of Rabinowitz and Widin (1984) that when the suppressor tone is set at 500 Hz, the level required to reduce the SOAE at 1350 Hz is above 70 dB SPL is inline with the 100 dB SPL for bias tones at or below 100 Hz (Fig. 4 top). The slopes of about 6–8 dB/octave on the iso-suppression and maximal sideband curves (Figs. 5 and 6) are consistent with the iso-modulation function in a recent study on low-frequency biasing of DPOAEs (Marquart *et al.*, 2007) and correlated with the tail of suppression tuning curves measured from SOAEs (Zurek, 1981; Bargones and Burns, 1988). These nonlinear interactions could generate more spectral and temporal information that are necessary for sensitive detection of sound.

Another finding indicates that the AM occurs prior to suppression as marked by the appearance of sidebands (Figs. 3 and 4). Indeed, this is similar to the results of Norrix and Glattke (1996) that when the suppressor frequency (f_s) is carefully placed closely below the frequency of an SOAE (f_{OAE}), a DPOAE can be generated at $2f_s$-f_{OAE}. In addition, when the SOAE is suppressed the DPOAE magnitude reaches a peak similar to Fig. 4. This DPOAE can be considered as an LSB II, i.e., f_{OAE}-$2(f_{\text{OAE}}$-$f_s)$, if the cochlear transducer is thought to be modulated by the beats (f_{OAE}-f_s) created by the two tones (Brown 1994; van der Heijden, 2005). When the suppressor tone is shifted more than an octave below the SOAE in the case of the low-frequency biasing, the hair cells at the SOAE generating site should be modulated at the f_s. The sidebands should present at $f_{\text{OAE}} - n \cdot f_s$, where n is an integer (Bian, 2006). This has been verified by direct observations of two-tone suppression on basilar membrane (BM) vibration (Rhode and Cooper, 1993) and auditory nerve (Temchin *et al.*, 1997) that the response of the characteristic frequency (CF) was reduced with the presence and growth of the spectral components at $f_{\text{CF}} \pm n \cdot f_s$. Such opposite behaviors of CF response and its sidebands occur with an AM of the CF ($f_s \ll f_{\text{CF}}$). These observations suggest that the sensory hair cells could phase lock to a lower frequency component in the incoming acoustic stimuli comparing to their own CF. In the time domain, this is probably crucial for a slower operation of the hair cells and the rest of the auditory system to preserve energy. Beats and roughness observed by Long (1998) also were occasionally reported by the subjects in the present study. Compared with the spectral and temporal features of the AM of SOAEs (Figs. 3 and 9), these perceptual consequences are not difficult to understand. Whether the DPs or simply the sidebands from AM are used for sensitive sound detection, they are all generated by the nonlinearity in the hair cell transducers. Therefore, suppression and distortion are spectral representations of the temporal actions of the sensory hair cells.

Because SOAEs are generated from the intrinsic processes of the inner ear, they tend to be influenced by any

subtle changes in the cellular environment of the hair cells. This sensitivity of SOAEs is demonstrated by the amplitude fluctuation when the bias tone level is low (Fig. 4 top and Fig. 7). The amplitude variability of SOAE is much higher than the DPOAEs with similar magnitudes and under similar biasing conditions (Bian and Scherrer, 2007). Without biasing, variability in SOAE amplitude is common (Lind and Randa, 1990; Burns, 1994; Smurzynski and Probst, 1998). This instability of SOAEs implicates that they are products of some dynamic processes of the hair cell transduction, possibly a limit-cycle oscillation of the stereocillial bundles (Talmadge *et al.*, 1991; Nadrowski *et al.*, 2004). The instability and spontaneous oscillation of the hair cell transduction channels seem necessary so that any sound induced change in the cochlear fluid pressure could modulate the amplitude and frequency of the movement of the hair bundles to provide a high sensitivity. In normal subjects, hearing thresholds at the SOAEs frequencies are indeed lower than other frequencies (Long and Tubis, 1988) that could cause the well-known "microstructures" in high-resolution audiograms.

C. Period modulation patterns

The period modulation pattern shows two peaks each corresponding to a zero crossing of the bias tone. Variations from the typical pattern are often found in three aspects: (1) merging of the two SOAE amplitude peaks with reducing biasing levels, (2) suppression of the first peak but not the second while increasing f_{bias}, and (3) delays of the SOAE peaks (Figs. 10 and 11). These changes in period modulation pattern determine the size and number of the spectral sidebands. Merging of the SOAE peaks indicates that BM displacement or stereocilial movement in the positive sound pressure direction is less effective in suppressing the SOAE than the opposite direction (Fig. 7 top). This asymmetry was observed in "low-side" suppression of the CF response on the BM (Rhode and Cooper, 1993) where a displacement towards scala tympani (ST) reduced the CF vibration more than the other direction. Moving the BM towards ST could push more OHC transduction channels into closed state, thus reducing the mechanical responses of the cells that produce SOAEs. In comparison, displacing the BM towards scala vestibuli (SV) can enhance the electrically evoked OAEs (Kirk and Yates, 1998).

For faster biasing, esp., at 100 Hz, suppression of SOAE during the displacements towards ST seems to prolong and prevent the SOAE amplitude from recovering when the cochlear partition returns to its resting position (Fig. 11). Such a period modulation pattern is similar to the period histogram of auditory nerve response during low-frequency suppression (Cai and Geisler, 1996). This prolonged effect probably correlates with the 5–10 ms time constant for the recovery of SOAE from a suppression produced by a high-frequency tone (Schloth and Zwicker, 1983). The mechanism of the multi-millisecond recovery time is unclear, perhaps due to the slow adaptation of the motor proteins in the OHC stereocilia. There is also a short delay of the SOAE peaks relative to the zero crossings of the bias tone, esp., at high L_{bias} and low f_{bias} (Fig. 10). This delay ($\leqslant$1 ms) indicates that the mechanism responsible for the suppression takes time to activate, possibly via a feedback to adjust the hair cell transducer gain (Bian *et al.*, 2004) or a fast adaptation of hair cell transduction channels (Ricci, *et al.*, 2000). Schloth and Zwicker (1983) also noticed that there was a less than 2 ms delay in the suppression and recovery of SOAEs. Moreover, it is shown that a suppressor should be placed 1–2 ms prior to an emission evoking click to take effect (Harte *et al.*, 2005). However, this time dependency of SOAE modulation is complicated by the bias tone level and frequency, e.g., the second peak can occur earlier than the zero-crossing in the unloading phase (Fig. 10).

V. SUMMARY AND CONCLUSION

Subjects with relatively large SOAEs were selected to receive low-frequency modulation under various signal conditions. The results showed a combined suppression and modulation of the SOAE amplitudes at high biasing levels. In the spectral domain, reduction in SOAE amplitude was observed with the generation and growth of sidebands when the L_{bias} was increased. For a fixed bias tone, the extent of suppression and behavior of modulation for an SOAE varied depending on the frequency and amplitude of the particular SOAE. The ability of the bias tone to suppress and modulate SOAE decreased with the frequency of SOAEs at a 6–8 dB/octave rate. The quasi-static modulation patterns of SOAEs demonstrated a bell shape which was associated with the first derivative of a nonlinear F_{Tr} of the cochlea. In the time domain, SOAEs showed an AM depending on the phase of the bias tone. Within a biasing cycle, the SOAE envelope peaked twice near the zero crossings of the bias tone. This typical period modulation pattern varied systematically with the L_{bias} and f_{bias}. The temporal behaviors of SOAE amplitudes reflect a time-dependent shifting of the operating point on the cochlear F_{Tr}. These comparable results to a recent observation on low-frequency biasing of DPOAEs in humans (Bian and Scherrer, 2007) suggest that the cochlear nonlinearity inherited with the hair cell transduction could be involved in the generation of SOAEs. More research is needed to elucidate the relation between the cochlear nonlinearity and other mechanisms (e.g., Shera, 2003) in the formation of SOAEs. The influence of SOAEs on the accuracy of estimating the cochlear F_{Tr} should be determined for possible clinical applications of the low-frequency biasing technique.

ACKNOWLEDGMENTS

Assistance from Nicole Scherrer in recruiting subjects is appreciated. The authors thank the clinical staff in the Department of Speech and Hearing Science at ASU for sharing their equipment and the subject for participating in the study. This work was supported by Grant No. R03 DC006165 from the National Institute on Deafness and Other Communication Disorders of the NIH.

Aibara, R., Welsh, J. T., and Goode, R. L. (**2001**). "Human middle-ear sound transfer function and cochlear input impedance," Hear. Res. **152**, 100–109.
Bargones, J. Y., and Burns, E. M. (**1988**). "Suppression tuning curves for spontaneous otoacoustic emissions in infants and adults," J. Acoust. Soc. Am. **83**, 1809–1816.

Bian, L. (**2006**). "Spectral fine-structures of low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 3872–3885.
Bian, L. (**2004**). "Cochlear compression: Effects of low-frequency biasing on quadratic distortion product otoacoustic emission," J. Acoust. Soc. Am. **116**, 3559–3571.
Bian, L., and Chertoff, M. E. (**2006**). "Modulation patterns and hysteresis: Probing cochlear dynamics with a bias tone," in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 93–100.
Bian, L., and Scherrer, N. M. (**2007**). "Low-frequency modulation of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **122**, 1681–1692.
Bian, L., Chertoff, M. E., and Miller, E. (**2002**). "Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **112**, 198–210.
Bian, L., Linhardt, E. E., and Chertoff, M. E. (**2004**). "Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **115**, 2159–2172.
Braun, M. (**2006**). "A retrospective study of the spectral probability of spontaneous otoacoustic emissions: Rise of octave shifted second mode after infancy," Hear. Res. **215**, 39–46.
Braun, M. (**2000**). "Inferior colliculus as candidate for pich extraction: Multiple support from statistics of bilateral spontaneous otoacoustic emissions," Hear. Res. **145**, 130–140.
Brown, A. M. (**1994**). "Modulation of the hair cell motor: A possible source of odd-order distortion," J. Acoust. Soc. Am. **96**, 2210–2215.
Burns, E. M., Campbell, S. L., and Arehart, K. H. (**1994**). "Longitudinal measurements of spontaneous otoacoustic emissions in infants," J. Acoust. Soc. Am. **95**, 385–394.
Cai, Y., and Geisler, C. D. (**1996**). "Suppression in auditory-nerve fibers of cats using low-side suppressors. II. Effect of spontaneous rates," Hear. Res. **96**, 113–125.
Frank, G., and Kössl, M. (**1996**). "The acoustic two-tone distortions 2fl-f2 and f2-fl and their possible relation to changes in the operating point of the cochlear amplifier," Hear. Res. **98**, 104–115.
Frick, L. R., and Matthies, M. L. (**1988**). "Effects of external stimuli on spontaneous otoacoustic emissions," Ear Hear. **9**, 190–197.
Gold, T. (**1948**). "Hearing. II. The physical of the basis of the action of the cochlea," Proc. R. Soc. London, Ser. B **135**, 492–498.
Guinan, J. J., Jr., Warr, W. B., and Norris, B. E. (**1984**). "Topographic organization of the olivocochlear projections from the lateral and medial zones of the superior olivary complex," J. Comp. Neurol. **226**, 21–27.
Harte, J. M., Elliott, S. J., Kapadia, S., and Lutman, M. E. (**2005**). "Dynamic nonlinear cochlear model predictions of click-evoked otoacoustic emission suppression," Hear. Res. **207**, 99–109.
Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.
Kirk, D. L., and Yates, G. K. (**1998**). "Enhancement of electrically evoked oto-acoustic emissions associated with low-frequency stimulus bias of the basilar membrane towards scala vestibuli," J. Acoust. Soc. Am. **104**, 1544–1554.
Lind, O., and Randa, J. S. (**1990**). "Spontaneous otoacoustic emissions: Incidence and short-time variability in normal ears," J. Otolaryngol. **19**, 252–259.
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., Martin, G. K., Probst, R., and Coats, A. (**1988**). "Spontaneous otoacoustic emissions in a nonhuman primate: II. Cochlear anatomy," Hear. Res. **33**, 69–94.
Long, G. R. (**1998**) "Perceptual consequences of the interactions between spontaneous otoacoustic emissions and external tones. I. Monaural diplacusis and aftertones," Hear. Res. **119**, 49–60.
Long, G. R., and Tubis, A. (**1988**). "Investigations into the nature of the association between threshold microstructure and otoacoustic emissions," Hear. Res. **36**, 125–138.
Manley, G. A., and Taschenberger, G. (**1993**). "Spontaneous otoacoustic emissions from a bird: A preliminary report," 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. 33–39.
Marquardt, T., Hensel, J., Mrowinski, D., and Scholz, G. (**2007**). "Low-frequency characteristics of human and guinea pig cochleae," J. Acoust. Soc. Am. **121**, 3628–3638.
Martin, G. K., Lonsbury-Martin, B. L., Probst, R., and Coats, A. (**1985**). "Spontaneous otoacoustic emissions in the nonhuman primate: A survey," Hear. Res. **20**, 91–95.
Martin, P., Bozovic, D., Choe, Y., and Hudspeth, A. J. (**2003**). "Spontaneous oscillation by hair bundles of the bullfrog's sacculus," J. Neurosci. **23**, 4533–4548.
McFadden, D., and Plattsmier, H. S. (**1984**). "Aspirin abolishes spontaneous otoacoustic emissions," Hear. Res. **16**, 251–260.
Moulin, A., Collet, L., and Morgan, A. (**1993**). "Interrelations between transiently evoked otoacoustic emissions, spontaneous otoacoustic emissions and acoustic distortion products in normally hearing subjects," Hear. Res. **65**, 216–233.
Nadrowski, B., Martin, P., and Jülicher, F. (**2004**). "Active hair bundle motility harnesses noise to operate near an optimum of mechanosensitivity," Proc. Natl. Acad. Sci. U.S.A. **101**, 12195–12200.
Neely, S. T., Johnson, T. A., Garner, C. A., and Gorga, M. P. (**2005**). "Stimulus-frequency otoacoustic emissions measured with amplitude-modulated suppressor tones," J. Acoust. Soc. Am. **118**, 2124–2127.
Norrix, L. W., and Glattke, T. J. (**1996**). "Distortion product otoacoustic emissions created through the interaction of spontaneous otoacoustic emissions and externally generated tones," J. Acoust. Soc. Am. **100**, 945–955.
Ohyama, K., Wada, H., Kobayashi, T., and Takasaka, T. (**1991**). "Spontaneous otoacoustic emissions in the guinea pig," Hear. Res. **56**, 111–121.
Palmer, A. R., and Wilson, J. P. (**1982**). "Spontaneous and evoked acoustic emissions in the frog *Rena esculenta*," J. Physiol. (London), **324**(Suppl.), 66P.
Penner, M. J., Glotzbach, L., and Huang, T. (**1993**). "Spontaneous otoacoustic emissions: Measurement and data," Hear. Res. **68**, 229–237.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2017–2067.
Pujol, R. (**2001**). "Neural anatomy of the cochlea: Development and plasticity," in *Physiology of the Ear*, 2nd ed., edited by A. F. Jahn and J. Santos-Sacchi (Singular, NY), pp. 515–528.
Rabinowitz, W. M., and Windin, G. P. (**1984**). "Interaction of spontaneous otoacoustic emissions and external sounds," J. Acoust. Soc. Am. **76**, 1713–1720.
Rhode, W. S., and Cooper, N. P. (**1993**). "Two-tone suppression and distortion product on the basilar membrane in the hook region of cat and guinea pig cochleae," Hear. Res. **66**, 31–45.
Ricci, A. J., Crawford, A. C., and Fettiplace, R. (**2000**). "Active hair bundle motion linked to fast transducer adaptation in auditory hair cells," J. Neurosci. **20**, 7131–7142.
Schloth, E., and Zwicker, E. (**1983**). "Mechanical and acoustical influences on spontaneous oto-acoustic emissions," Hear. Res. **11**, 285–193.
Scholz, G., Hirschfelder, A., Marquardt, T., Hensel, J., and Mrowinski, D. (**1999**). "Low-frequency modulation of the $2f_1-f_2$ distortion product otoacoustic emissions in the human ears," Hear. Res. **130**, 189–196.
Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized occhlear standing waves," J. Acoust. Soc. Am. **114**, 224–262.
Smurzynski, J., and Probst, R. (**1998**). "The influence of disappearing and reappearing spontaneous otoacoustic emissions on one subject's threshold microstructure," Hear. Res. **115**, 197–205.
Stewart, C. E., and Hudspeth, A. J. (**2000**). "Effects of salicylates and aminoglycosides on spontaneous otoacoustic emissions in the Tokay gecko," Proc. Natl. Acad. Sci. U.S.A. **97**, 454–459.
Talmadge, C. L., Tubis, A., Wit, H. P., and Long, G. R. (**1991**). "Are spontaneous otoacoustic emissions generated by self-sustained cochlear oscillators?" J. Acoust. Soc. Am. **89**, 2391–2399.
Temchin, A. N., Rich, N. C., and Ruggero, M. A. (**1997**). "Low-frequency suppression of auditory nerve responses to characteristic tones," Hear. Res. **113**, 29–56.
Thiers, F. A., Burgess, B. J., and Nadol, J. B., Jr. (**2002**). "Reciprocal innervation of outer hair cells in a human infant," J. Assoc. Res. Otolaryngol. **3**, 269–278.
van der Heijden, M. (**2005**). "Cochlear gain control," J. Acoust. Soc. Am. **117**, 1223–1233.
Whittemore, K. R., Merchant, S. N., Poon, B. B., and Rosowski, J. J. (**2004**). "A normative study of tympanic membrane motion in humans using a laser Doppler vibrometer (LDV)," Hear. Res. **187**, 85–104.
Zurek, P. M. (**1981**). "Spontaneous narrowband acoustic signals emitted by human ears," J. Acoust. Soc. Am. **69**, 514–523.
Zwicker, E. (**1981**). "Masking-period patterns and cochlear acoustical responses," Hear. Res. **4**, 195–202.

Phoneme representation and classification in primary auditory cortex

Nima Mesgarani, Stephen V. David, Jonathan B. Fritz, and Shihab A. Shamma[a)]
Electrical and Computer Engineering & Institute for Systems Research, University of Maryland, College Park, Maryland 20742

(Received 5 June 2007; revised 15 October 2007; accepted 30 October 2007; corrected 20 March 2008)

A controversial issue in neurolinguistics is whether basic neural auditory representations found in many animals can account for human perception of speech. This question was addressed by examining how a population of neurons in the primary auditory cortex (A1) of the naïve awake ferret encodes phonemes and whether this representation could account for the human ability to discriminate them. When neural responses were characterized and ordered by spectral tuning and dynamics, perceptually significant features including formant patterns in vowels and place and manner of articulation in consonants, were readily visualized by activity in distinct neural subpopulations. Furthermore, these responses faithfully encoded the similarity between the acoustic features of these phonemes. A simple classifier trained on the neural representation was able to simulate human phoneme confusion when tested with novel exemplars. These results suggest that A1 responses are sufficiently rich to encode and discriminate phoneme classes and that humans and animals may build upon the same general acoustic representations to learn boundaries for categorical and robust sound classification. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2816572]

PACS number(s): 43.64.Sj, 43.71.Qr [WPS] Pages: 899–909

I. INTRODUCTION

Humans reliably identify many phonemes and discriminate them categorically, despite considerable natural variability across speakers and distortions in noisy and reverberant environments that limit the performance of even the best speech recognition algorithms.[1,2] Trained animals have also been shown to discriminate phoneme pairs categorically and to generalize to novel situations.[3–10] The neurophysiological basis of these perceptual abilities in humans and animals remains uncertain. However, there is experimental evidence for cortical encoding of phonetic acoustic features regarded as critical for distinguishing classes of consonant-vowel (CV) syllables, such as voice-onset time.[11–14] Key questions include the nature and location of the neural representations of different phonemes and, more specifically, whether the neural responses of the primary auditory cortex (A1) are sufficiently rich to support the phonetic discriminations observed in humans and animals.

The general issue of the neural representation of complex patterns is common to all neuroscience and has been investigated in many sensory modalities. In the visual system, recent studies have shown that responses of approximately 100 cells in the inferior temporal cortex are sufficient to account for the robust identification and categorization of several object categories.[15] In the auditory system, a recent study has shown that neurometric functions derived from single unit recordings in the ferret primary auditory cortex closely parallel human psychometric functions for complex sound discrimination.[16] An important aspect of our approach in the present study is the inclusion of temporal features of the response in the analysis. This is crucial because phonemes are *spectro-temporal* patterns, and hence analyzing their neural representation at a single cell or ensemble level requires consideration of the interactions between the stimuli and the intrinsic dynamics of individual neurons.

In the present study, we recorded responses of A1 neurons to a large number of American English phonemes in a variety of phonemic contexts and derived from many speakers. Our results demonstrate that (I) time-varying responses from a relatively small population of primary auditory cortical neurons (<100) can account for distinctive aspects of phoneme identification observed in humans,[17] and that (II) well known acoustic features of phonemes are indeed explicitly encoded in the population responses in A1.

The analysis of the categorical representation of phonemes across a neuronal population presented in this paper remains largely model-independent in that only relatively raw response measures (e.g., peri-stimulus time histograms, PSTHs) are used in the computations and illustrations. The one key departure from this rule is necessitated by the desire to organize the display of the population responses according to their best frequency, spectral scale, and temporal dynamics. These response properties are quantified using the measured spectro-temporal receptive field (STRF) model of the neurons.[18,19]

[a)]Author to whom correspondence should be addressed. Electronic mail: sas@isr.umd.edu

II. EXPERIMENTAL PROCEDURES

The protocol for all surgical and experimental procedures was approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Maryland and consistent with NIH Guidelines.

A. Surgery

Four young adult, female ferrets were used in the neurophysiological recordings reported here. To secure stability of the recordings, a stainless steel head post was surgically implanted on the skull. During implant surgery, the ferrets were anesthetized with Nembutal (40 mg/kg) and Halothane (1–2%). Using sterile procedures, the skull was exposed and a headpost was mounted using dental cement, leaving clear access to primary auditory cortex in both hemispheres. Antibiotics and analgesics were administered as needed.

B. Neurophysiological recording

Experiments were conducted with awake head-restrained ferrets. The animals were habituated to this setup over a period of several weeks, and usually remained relaxed and relatively motionless throughout recording sessions that may last 2–4 h. Recordings were conducted in a double-walled acoustic chamber. Small craniotomies (1–2 mm in diameter) were made over the primary auditory cortex before recording sessions. Physiological recordings were made using tungsten microelectrodes (4–8 MΩ). Electrical signals were amplified and stored using an integrated data acquisition system (Alpha Omega). Spike sorting of the raw neural traces was done offline using a custom principal component analysis (PCA) clustering algorithm. Our requirements for single unit isolation of stable waveforms included (1) that the waveform and spike rate remained stable throughout the recording, and (2) that the inter-spike interval for each neuron was distributed exponentially with a minimum latency of 2 ms.

C. Speech stimuli and data analysis

Stimuli were phonetically transcribed continuous speech from the TIMIT database.[20] Thirty different sentences (3 s, 16 kHz sampling) spoken by different speakers (15 male and 15 female) were used to sample a variety of speakers and contexts. A large stimulus set was used, that extended the original set from 30 to 90 sentences, and also increased speaker diversity to 45 male and 45 female speakers. In all recordings, each sentence was presented five times.

D. Mean phoneme representation

The TIMIT phonetic transcriptions were used to align the responses of each neuron to all the instances of a given phoneme and then averaged to compute the peri-stimulus time histogram (PSTH) response to that phoneme, as illustrated in Fig. 1(A) (10 ms time bins). We did not attempt to compensate for the relatively short latency of neural responses in the ferret, since this was roughly constant and consistent for all A1 neurons (15–20 ms). We also computed the auditory spectrogram of each phoneme using the following procedure: Let $S(t,f)$ be the auditory spectrogram of the speech stimulus computed using a model of cochlear frequency analysis,[21] and let $r(t)$ be the corresponding neural response. For phoneme k, which occurs at times t_{k_1}, $t_{k_2},\ldots,t_{k_n}$, the average spectrogram is

$$\hat{S}_k(t,f)=\frac{1}{n}\sum_{i=1}^{n} S(t_{k_i}+t,f) \tag{1}$$

and the average neural response is

$$\hat{r}_k(t)=\frac{1}{n}\sum_{i=1}^{n} r(t_{k_i}+t). \tag{2}$$

The total number of occurrences of each phoneme, n, ranged from 7 (e.g. /g/) to 72 (e.g., /i/) in the chosen sentences.

E. Measurement of neuronal tuning properties

We characterized each neuron by its spectro-temporal receptive field (STRF), estimated by normalized reverse correlation of the neuron's response to the auditory spectrogram of the speech stimulus.[18] Although methods such as normalized reverse correlation can produce unbiased STRF estimates in theory, practical implementation requires some form of regularization to prevent overfitting to noise along the low-variance dimensions. This in effect imposes a smoothness constraint on the STRF. The regression parameters were adjusted using a jackknife validation set to maximize the correlation between actual and predicted responses.[22] Figure 1(B) illustrates the STRF of one such neuron. We measured several tuning properties from each STRF: Best frequency (BF) was defined as the largest positive peak value of the STRF along its frequency dimension. The STRF scale and rate were estimated from the two-dimensional (2D) modulation transfer function (MTF) (Fig. 1(B)). The MTF is the 2D Fourier transform of the STRF that is then collapsed along its temporal or spectral dimensions (known also as the *rate* and *scale*) to obtain the purely *spectral* (*sMTF*) or *temporal* (*tMTF*) modulation transfer functions (Fig. 1(B)). The *best scale* (related to the inverse bandwidth) of an STRF is defined as the centroid of the sMTF (in "cycles/octave"), whereas "speed" or *best rate* of the STRF is defined as the centroid of the tMTF (in Hz), as illustrated in Fig. 1(B). To display the neural *population responses* for each phoneme, we generated two-dimensional "topographic" plots in which each row contained the average PSTH response of one neuron, sorted according to neural BF, scale or rate. The distribution of these three tuning properties in our sample was fairly broad, covering most BFs, best scales, and best rates (see Appendix). However, because the parameters were not distributed exactly uniformly, we interpolated the vertical axis of the smoothed PSTH (2D disk filter: 60 ms * 6 neurons) to have uniform spacing and then smoothed the PSTH display with the same 2D filter. We characterized each phoneme according to the *locus* of maximal response within the neural population along the BF, scale and rate dimensions. For example, to find the locus along the BF dimension, we determined the position of the maximum PSTH responses over time for neurons ordered along the BF axis. The same

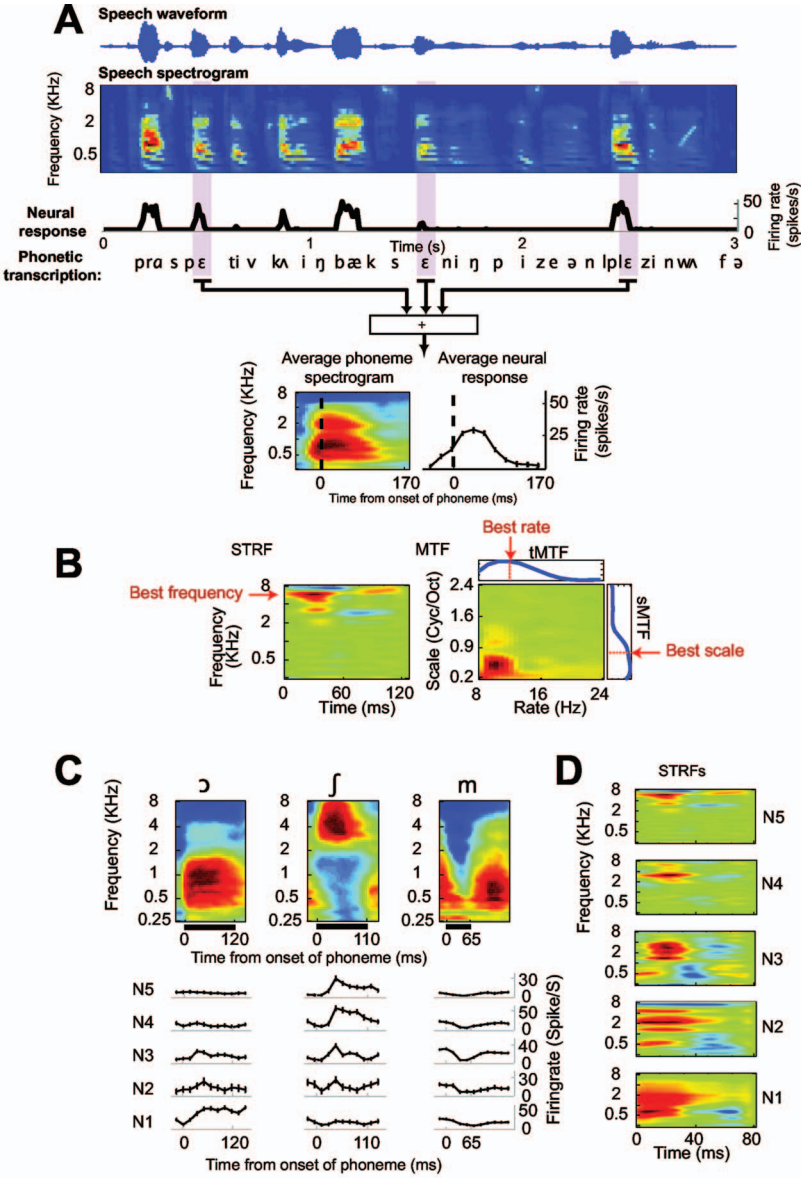


FIG. 1. Neuronal responses to phonemes in continuous speech. (A) The spectrograms of all /ɛ/ vowel exemplars were extracted and averaged to obtain one grand average auditory spectrogram (bottom left). In this and following average spectrogram plots, red areas indicate regions of higher than average energy and blue regions indicate weaker than average energy. The corresponding PSTH response to /ɛ/ was computed by averaging neural spike rates over the same time windows (bottom right). (B) The spectro-temporal receptive field (STRF) of a neuron as measured by normalized reverse correlation. Red areas indicate stimulus frequencies and time lags correlated with an increased response, and blue areas indicate stimulus features correlated with a decreased response. The neuron's BF was defined to be the excitatory peak of the STRF (red arrow). The modulation transfer function (MTF) is computed by taking the absolute value of the 2D Fourier transform of the STRF. We then collapse along the temporal or spectral dimensions (known also as the *rate* and *scale*) to obtain the purely *spectral (sMTF)* or *temporal (tMTF)* modulation transfer functions. The *best scale* (proportional to the inverse of bandwidth) of a STRF was defined as the centroid of the sMTF (in "cycles/octave"), whereas "speed" or *best rate* of the STRF is defined as the centroid of the tMTF (in Hz). The choice of *centroid* for best-scale parameter results in a compressed range but it does not affect the ordering of neurons along this dimension. (C) Average auditory spectra of three phonemes (/ɔ/, /ʃ/, /m/). Below each spectrogram is the PSTH response of five example neurons (labeled N1–N5). (D) The STRFs of these neurons indicate a diversity of spectro-temporal tuning properties.

procedure was repeated for PSTHs ordered along the scale and rate axes to obtain the three coordinates of the locus.

F. Phoneme classification and confusions

To examine the separation or overlap among the representations of different phonemes, we trained linear binary classifiers to discriminate each phoneme from all the others based on the neuronal population response. Formally, the neurons project the phoneme acoustic signals into a high dimensional space (i.e., the total number of neurons $\times$ the number of samples in each PSTH$=90\times22$). Because of the different selectivity of each neuron, different phonemes fall in specific subregions of this space.

A linear Support Vector Machine (SVM[35]) was trained to find the optimal hyperplanes for each phoneme, such that the hyperplane has the maximum distance (or "margin") to the closest data points (or "support vectors") in the two classes it separates. Using linear hyperplanes is intuitively appealing because the classifier's output is a weighted sum of the neural responses that can be interpreted easily. The output of each classifier is a scalar value indicating the distance of the data point to the hyperplane. Novel sounds are identified by choosing the classifier that produces the maximum distance to the boundary. We should emphasize that the order of the neural responses is not important in any way for classification.

G. Statistical analysis

The significance of correlations between the pattern of phoneme confusion predicted by the neural classifier and confusion observed for human perception[17] was ascertained by a randomized t test. Random correlations were computed between neural and perceptual confusion matrices after randomly shuffling phoneme identity (20 000 shuffles). The significance of the correlation between the actual confusion matrices was taken as the probability that such a correlation could be produced by the randomly shuffled matrices.

H. Measuring the acoustic distance among phonemes

The average auditory spectrogram of each phoneme was computed as described above.[21] The acoustic similarity between any pair of phonemes was then defined as the Euclidean distance between their average spectrograms.

III. RESULTS

A. Diversity of single-unit responses to phonemes

Physiological responses were recorded from 90 single units in A1 of four ferrets (*Mustela putorius*) during the monaural presentation of continuous speech stimuli (see Fig. 1(A)). The recorded neurons were broadly distributed in their spectral tuning and dynamic response properties as shown by population range of their best frequency (BF), best scale, and best rate (documented in the scatter plots in Fig. 6 in the Appendix). These neural tuning properties are based on measurements of the spectro-temporal receptive fields of the neurons (STRFs) as depicted in Fig. 1(B) and described in detail earlier in Sec. II. Figure 1(C) illustrates the PSTH responses of five single units (N1–N5) to three different phonemes (vowel /ɔ/, fricative /ʃ/ and nasal /m/) whose average auditory spectra are depicted in Fig. 1(C). The spectro-temporal receptive fields (STRFs) of the five selected neurons are shown in Fig. 1(D).

Each phoneme activates these five neurons differentially, depending on the match between the neuron's STRF and the spectro-temporal structure of the stimulus. For instance, the vowel /ɔ/ drives N1 very effectively because of the low BF of the neuron (~700 Hz). By contrast, the fricative /ʃ/ maximally activates N4 and N5, which have the highest BFs (~3 and ~7 kHz, respectively). Finally, the response pattern of the nasal /m/ is unique in that it causes a depression of responses in N2 and N3, reflecting the energy dip midway through the phoneme over all frequencies, but especially in the middle frequencies (~0.5–4 kHz).[23,24] In this manner, each phoneme evokes a unique response pattern across the population of A1 cells that differs from the evoked responses elicited by other phonemes.

B. Population responses to phoneme classes

To appreciate the unique response patterns evoked by different phonemes and, in particular, in order to highlight the acoustic features enhanced in the neural representation, it is best to view the ordered activity of the entire population simultaneously. This ordering depends entirely on the neuronal tuning properties to be emphasized. For instance, inspired by the tonotopic organization of the auditory pathway, the most common way to organize neural PSTHs has been by frequency according to the BF of the units.[25,26] However, unlike the receptive fields of fibers in the auditory nerve, A1 neurons exhibit systematic variations of tuning along multiple feature axes, including bandwidth, asymmetry, and temporal dynamics.[14,27,28]

Here we consider the ordered representation of phoneme responses along BF and two other dimensions derived from the STRF: best scale and best rate (see Sec. II and Fig. 1(B)). Best scale is *inversely proportional* to bandwidth and indicates how wide a range of sound frequencies are integrated into the neural response. Best rate indicates the dynamic agility of a neuron's responses and hence reflects the temporal modulation of the stimulus spectrum that best drives the neuron. The coordinates of each cell along these three dimensions can be estimated using a variety of techniques and stimuli. The most common techniques include tuning curves or iso-response functions measured from tones[28] and STRFs measured from ripples.[29] We use the speech-based STRFs to estimate these parameters for each cell.[18]

1. Encoding of vowels

Population responses to 12 American-English vowels are summarized in Fig. 2. Panels in the top row (Fig. 2(A)-I) display the average auditory spectrogram of each vowel computed from all of its samples encountered in the speech database (see Sec. II for details). The vowels are organized according to their articulatory configurations along the Open/Closed and Front/Back axes,[23] as illustrated at the top of Fig. 2: /o/, /ɔ/, /ɑ/, /ʌ/, /æ/, /ɛ/, /e/, /ə/, /i/, /ɪ/, /ɨ/, /ʉ/. The three middle vowels (/ɛ/, /e/, /ə/) are tightly clustered near the midpoint of the Front/Back and Open/Closed axes, and are difficult to order accurately along this one-dimensional representation of the vowels.

The averaged spectra (top row) reveal that Mid/Back vowels (/o/, /ɔ/, /ɑ/, and /ʌ/) have relatively concentrated activity at low to medium frequencies (~0.4–2 KHz), whereas Front vowels sometimes have two peaks spaced over a larger frequency range (~0.3 and ~4 KHz). This is consistent with the known distribution of the three formants (F1, F2, and F3) in these vowels,[23] namely, that they have F1 and F2 that are closely spaced, creating compact single broad peak spectra at intermediate frequencies (reminiscent of the center-of-gravity hypothesis of Chistovich and Lublinskaya[30]). As the vowels become more "Front"ed, the single peak broadens and splits (/æ/ to /ə/). Continuing this trend, Front/Closed vowels (/i/, /ɪ/, /ɨ/, /ʉ/) exhibit relatively narrow and well separated formant peaks with F1 at low and F2 at high frequencies.

These averaged phoneme spectra are broadly reflected in the response distributions ordered along the BF axis; neurons with BFs matching regions of high energy in a phoneme spectrum tend to give strong responses to that phoneme (Fig. 2(A)-II). However, notable differences of unknown significance exist such as the relative weakness of the low BF peaks in /e/ and /ə/, and of the high BF peak in /i/). More striking, however, are the response distributions along the best scale axis, which roughly indicates the *inverse* of the vowels' spectral bandwidths (Fig. 2(A)-III). Here, consistent with the bandwidths of the spectral peaks discussed earlier, Central/Open vowels tend to evoke maximal responses in broadly tuned cells commensurate with their broad spectra (low scales <1 Cyc/Oct) while Closed vowels evoke maximal responses in narrowly tuned cells (scales >1 Cyc/Oct), as indicated by the blue and red boxes in Fig. 2(A)-III, respectively.[31] Response distributions in the best rate panels (Fig. 2(A)-IV) reveal a trend in the dynamics of the vowels as one moves along the Front/Back axis. Specifically, Front

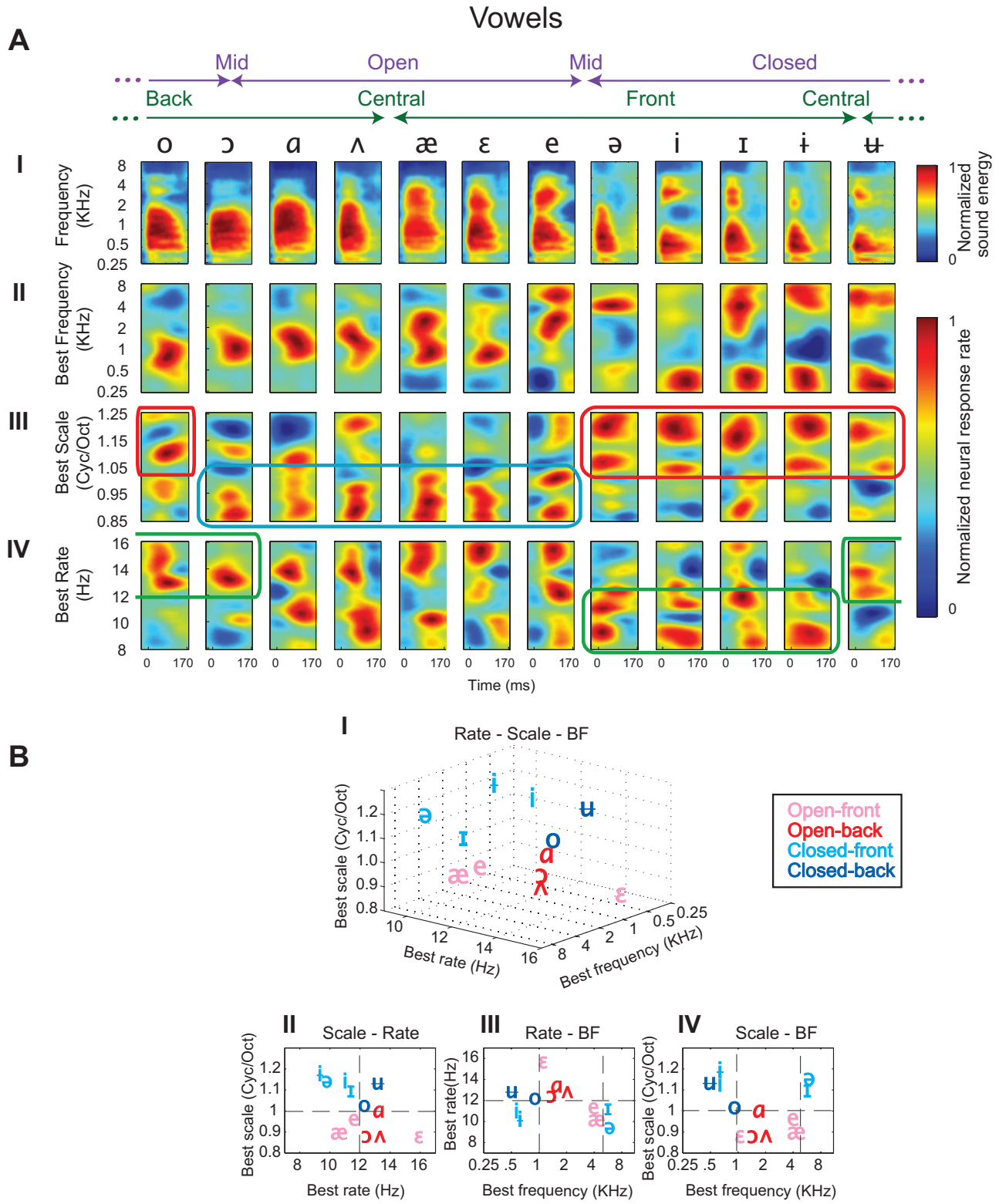


FIG. 2. Population response to vowels. (A) I. Average auditory spectrogram of 12 vowels organized approximately according to their open-closed and front-back articulatory features. The arrows at top indicate the *degree* of these features, with arrow "tips" representing minima (mid or central) and midpoints representing maxima. For example /ʌ/ is maximally open, but is neutral (central) on the front/back axis. Note also that the axes are presumed to loop around the page from right to left (dashed ends joining) creating a circular representation (II, III, IV): Average PSTH responses of 90 neurons to each vowel. Within each heat map, each row indicates the average response of a single neuron to the corresponding phoneme. Red regions indicate strong responses, and blue regions indicate weak responses. The average PSTH responses are sorted by neurons' best frequency (II), best scale (III) and best rate (IV) to emphasize the role of that parameter in the encoding of each vowel. (Details of the analysis and generation of these plots are given in Sec. II). (B) I. Each vowel is plotted at the centroid frequency, rate and scale of its average neuronal population response. The centroid values are calculated from the average PSTH responses sorted by the corresponding parameter (2A). "Open" vowels are shown in red, "Closed" vowels in blue, "Front" vowels with *light* font and "Back" vowels with *dark*. To visualize the contribution of each tuning property to vowel discrimination, the location of each vowel is also shown collapsed in 2–D plots of (II) scale-rate, (III) rate-BF and (IV) scale-BF. All other details of the analysis and generation of these plots are given in Sec. II (Experimental Procedures).

vowels (/ə/, /i/, /ɪ/, /ɨ/) evoke relatively stronger responses in the slower cells (with best rates $<\sim 12$ Hz), as compared to the more Back vowels (/ʉ/, /o/, /ɔ/) as highlighted by the green boxes in Fig. 2(A)-IV. The remaining more Central vowels (/ɑ/, /ʌ/, /æ/, /ɛ/, /e/) exhibit all dynamics. This response pattern may reflect the longer durations required to complete the articulatory excursions toward or away from Closed vowels towards the front of the vocal tract.

Figure 2(B) provides a compact summary of the population response to vowels. Each vowel is placed at the *locus* of maximum response in the neural population along the BF, best scale, and best rate axes. To highlight more clearly which of the three features best segregates them, the 3D display is projected onto each of the three marginal planes (Figs. 2(B)-II and 2(B)-IV). It is readily evident in these displays that the Open and Closed vowels separate along the scale axis above and below 1 Cyc/Oct (horizontal dashed lines in Figs. 2(B)-II and 2(B)-IV). They are also distinguished by BF, with the Open vowels clustering in the range 1.0–4.5 KHz (vertical dashed lines in Fig. 2(B)-III). Finally, the best rate axis segregates the Front/Back vowels (as discussed earlier), with Central and Back vowels located at high rates (>12 Hz), and Front vowels below it. It remains to be confirmed, however, whether these locations, which reflect the vowels' overall spectro-temporal similarity, can explain the perceptual confusion among them[32].

2. Encoding of consonants

Population responses to 15 consonants are shown in Fig. 3 in the same format already described for vowels. Three properties are commonly used to organize and classify consonants: place of articulation, manner of articulation, and voicing.[23,24,33] Here we examined how these three properties are encoded in the responses of the neuron population.

The distributions of the responses to the consonants sorted along the BF axis (Fig. 3(A)-II) approximate the features of their averaged spectra (Fig. 3(A)-I), which in turn are known to be closely related to place of articulation cues. For instance, the difference between the more forward places of constriction for /s/ compared to /ʃ/ is mirrored by the downward shift of the highpass spectral edge. Similarly the high-frequency noise burst at the onset of the forwardly constricted /t/ contrasts with the lower-frequency distribution of the other plosives (/p/ and /k/). However, there are also some notable differences in detail between the two sets of plots. There is generally a slight delay of about 20 ms in the neural responses relative to the spectrograms (presumably due to the latency of cortical responses). In addition, however, there are substantial differences between the responses and spectrograms in certain phonemes. For example, high BF responses to /f/ in Fig. 3(A)-II are strong despite their relative weakness in the spectrograms. Similarly, the low BF responses to /v/ are not consistent with the spectrogram. In other consonants, there are differences in the "timing" of certain frequency regions such as the rapid onset of high frequencies in the spectrogram of /t/ relative to its more delayed response, or in the continuity of the spectral regions in /ʃ/, /d/ and /ŋ/. The origin of all these differences is unclear and may reflect the nonlinearity of neural responses and/or our limited sampling of the neural population (90 neurons).

Response distributions along the best scale and best rate axes (Figs. 3(A)-III and 3(A)-IV) capture well the essential *manner of articulation* cues that supply the information necessary to discriminate plosives, fricatives, and nasals in continuous speech. For example, the broad distinction between "plosives" and "continuants" (e.g. /p/, /t/, /k/, /b/, /d/, /g/ versus /s/, /ʃ/, /z/, /n/, /m/, /ŋ/) is evident in the distribution of responses along the scale and rate axes (Figs. 3(A)-III and 3(A)-IV). Thus, plosives with their sudden and spectrally broad onsets display relatively strong activation in broadly tuned (low scales <1.1 cyc/oct) and fast (rates >12 Hz) cells (regions outlined in red in Figs. 3(A)-III and 3(A)-IV) compared to the more suppressed responses to longer duration unvoiced fricatives and nasals (outlined in blue in Fig. 3(A)-IV). Note also the brief suppressed response preceding the onset of all plosives due to the (silent) voice-onset-time (VOT) in all panels within the red box (Figs. 3(A)-III and 3(A)-IV).

Finally, the third cue of voicing is associated with the harmonic structure of voiced spectra near the low to mid-frequency range (0.2–1 kHz), and to a lesser extent the weak energy at low BFs near the fundamental of the voicing. Only this latter cue seems to distinguish consistently the voiced (/b/, /d/, /g/, /v/, /ð/, /z/, /m/, /n/, /ŋ/) from unvoiced (/p/, /t/, /k/, /f/, /s/, /ʃ/) consonants in our data as indicated by the green outlined region of Fig. 3(A)-II. However, such a strong low BF response as an indicator of "voicing" is missing in many of the vowel responses discussed earlier (e.g., the Open/Back vowels in Fig. 3(A)-II). Instead, its presence seems to correlate with the low F1 of the Closed vowels there. Therefore, our data suggest that the low-frequency voicing is reliably represented only in consonant responses, and perhaps in vowels where the F1 is low enough to amplify it.[34] However, there may well be a different and separate representation of voicing in the auditory cortex, for example, in terms of the pitch it evokes, or the harmonicity of its spectral components.[35]

Figure 3(B) illustrates the locus of the population response to each consonant in a plot of best frequency, best rate and best scale similar to that used with vowels earlier. The lower panels of Fig. 3(B) are projections of the three-dimensional (3D) plot onto its three marginal planes. Members of the three groups of consonants—plosives (red), fricatives (blue), and nasals (green)—are loosely grouped together in this parameter space. For instance, plosives tend to drive broadly tuned (scale <0.9 Cyc/Oct) and fast (rates >12 Hz) cells (Figs. 3(B)-II). Rate is also a distinguishing feature between plosives on the one hand, and nasals and (most) fricatives on the other (above and below 12 Hz, respectively). Similarly, phoneme groups roughly segregate along the BF axis, with unvoiced fricatives occupying the highest frequencies (>4 kHz), unvoiced plosives falling between 2 and 4 kHz, and other voiced phonemes falling below 2 kHz (Figs. 3(A)-III and 3(A)-IV). As with vowels, this plot of the neural loci of consonants reveals the

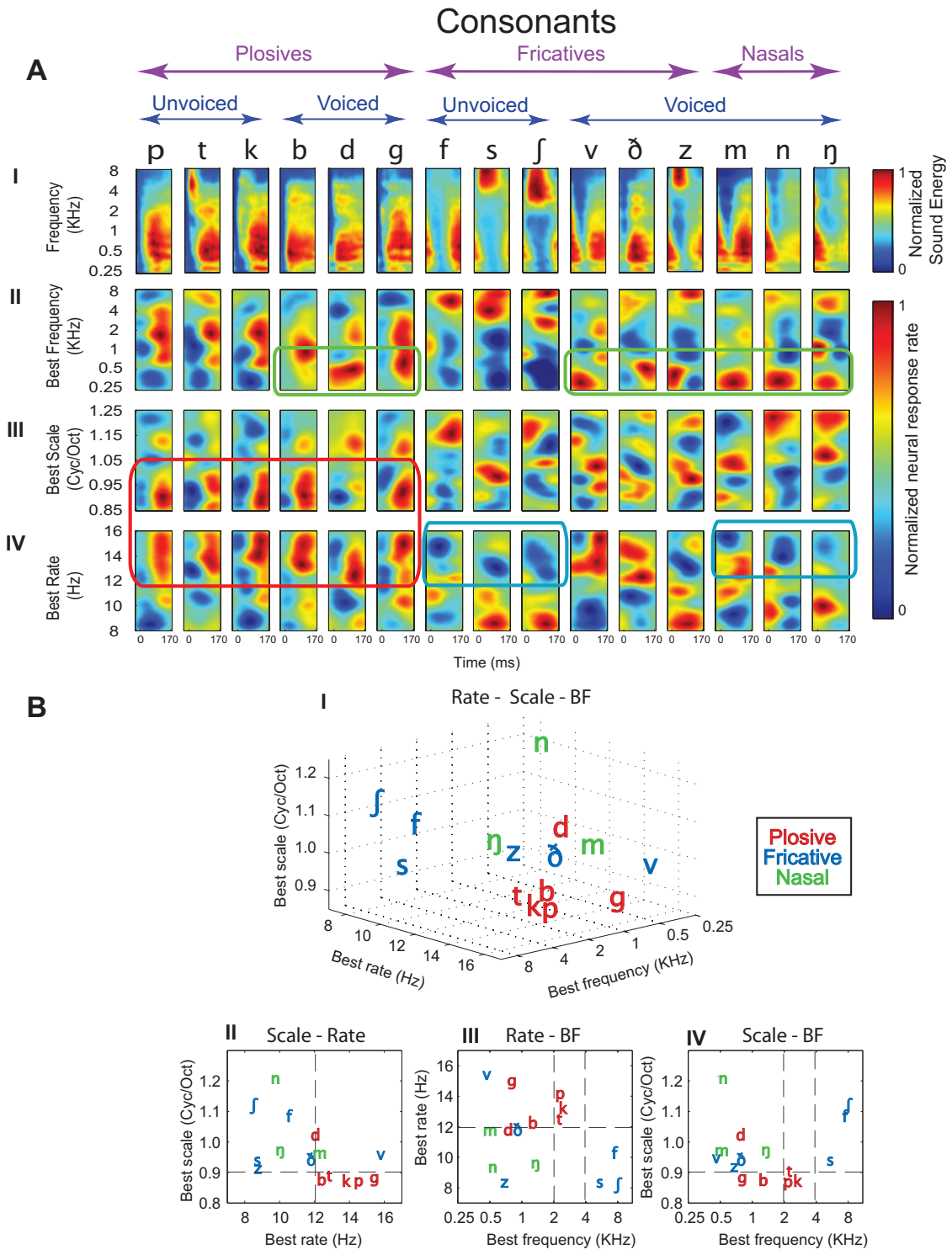


FIG. 3. Population response to consonants. (A) I. Average spectrogram of 15 consonants phonemes grouped as six plosives, six fricatives and three nasals. Each of the plosive and fricative-groups contains three voiced and three unvoiced phonemes (see arrows at top). (II, III, IV) Average PSTH responses of the neural population to each consonant, plotted as in Fig. 2(A). The average PSTH responses are sorted by neurons' best frequency (III), best scale (II) and best rate (IV) to emphasize the role of that parameter in the encoding of consonants. (All other details of the analysis and generation of these plots are given in Sec. II). (B) Each consonant is placed at the centroid frequency, rate and scale of its neuronal population response, measured from the corresponding PSTH responses (A). Plosive phonemes are plotted in red, fricatives in blue and nasals in green. The locus of each consonant is also shown collapsed in 2D plots of (II) scale-rate, (III) rate-BF and (IV) scale-BF. All other details of the analysis and generation of these plots are given in Sec. II (Experimental Procedures).

relative distances among them and perhaps explains the pattern of perceptual confusion observed between them, as we shall elaborate next.

C. Phoneme confusions

Average phoneme responses give useful insights into the mean representation of each phoneme, but they fail to indicate how well the neural population can discriminate pho-

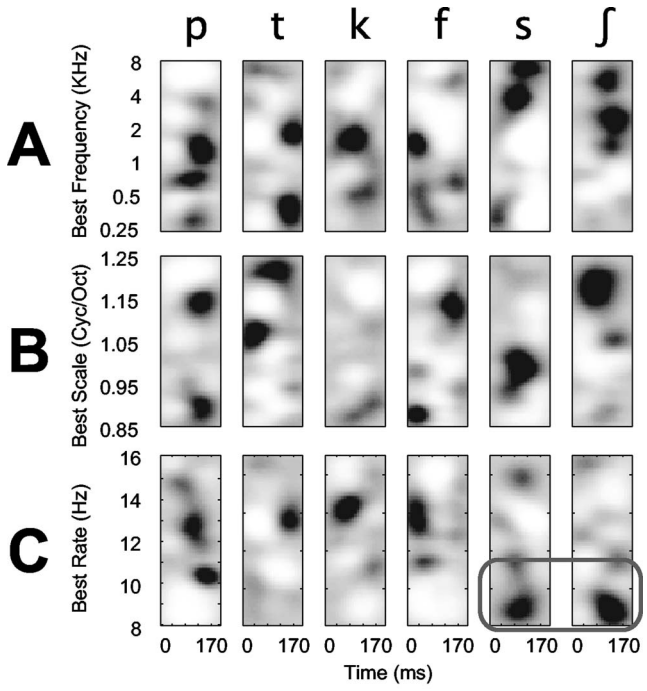


FIG. 4. Phoneme classification based on the population response. Classification masks for three unvoiced plosives (/p/, /t/, /k/) and three unvoiced fricatives (/f/, /s/, /ʃ/) sorted by neurons' best frequency (A), best scale (B) and best rate (C). Gray scale indicates the importance of the presence (*black* regions) or absence (*white* regions) of neural response for the classification of that phoneme. The output of each phoneme classifier is a scalar, computed as the sum of the population PSTH multiplied by the mask. Thus the order of the mask/PSTH is irrelevant to the output of the classifier.

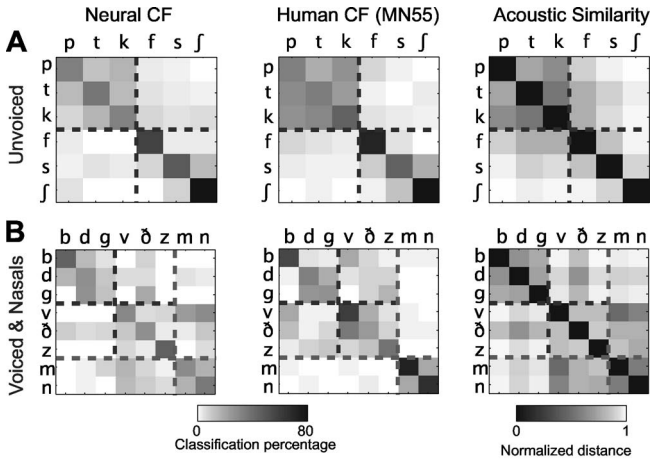


FIG. 5. Neural and human phoneme confusions, and phonemes acoustic similarity. Consonant confusion matrices from neural phoneme classifiers (left panels) and human psychoacoustic studies (Ref. 17) (middle panels). Gray scale indicates the probability of reporting a particular phoneme (column) for an input phoneme (row). (Right panels) The acoustic similarity between phoneme pairs defined as the Euclidian distance between their average auditory spectrograms. (A) Confusion matrices and phonemic distances for unvoiced consonants. Dashed lines separate the plosives /p/, /t/, /k/ from fricatives /f/, /s/, /ʃ/. (B) Confusion matrices and phonemic distances for voiced consonants. Dashed lines separate the plosives /b/, /d/, /g/ from fricatives /v/, /ð/, /z/ and the nasal consonants /m/ and /n/ from the rest.

nemes, given the natural acoustic variability among samples of the same phoneme during continuous speech. To delineate perceptual boundaries implied by the responses to the phonemes, we trained a linear classifier for each phoneme to separate it from all others, based on the PSTHs of the neural population.[36] To determine the identity of a novel phoneme, the population response was applied to all the classifiers, each computing the likelihood of its designated phoneme. The classifier indicating the maximum likelihood was taken as the identity of the input phoneme. To train and test the classifiers, we divided the speech data into 100 train and test subsets. In each subset, 90% of the data was randomly chosen for training and the remaining 10% was used for testing. The classification accuracy and the confusion matrices reported here are the average results from these 100 subsets.

Once trained, each linear classifier can be viewed as a mask that selects, by multiplication with the population response, the neurons and response latencies that most effectively distinguish the associated phoneme from all others. Figure 4 displays the masks computed for the unvoiced consonants /p/, /t/, /k/, /f/, /s/, /ʃ/. The masks are ordered in the same way as the PSTHs in Fig. 3(A) (i.e., by BF, best scale, and best rate). In the masks, black regions signify neurons and response latencies for which a strong response provides evidence for presence of the phoneme, and white regions signify strong responses that provide evidence against that phoneme. The masks in Fig. 4 differ from the mean neural responses in Fig. 3(A) in that they emphasize the *unique* features of each phoneme. For example, the mean responses to /ʃ/ (Fig. 3(A)-II) indicate strong responses in high and medium BF neurons, but in the masks the mid-BF neurons (2 kHz) are given higher weights. This differential weighting reflects the fact that both /ʃ/ and /s/ evoke strong responses from high BF neurons, but only /ʃ/ evokes responses from the mid-BF neurons. Similarly, the /p/, /t/, /k/ masks reflect only the features that distinguish these phonemes from each other. The BF masks (Fig. 4(A)) emphasize the low (750 Hz), high (>2 kHz), and medium (0.3–1.5 kHz) spectral regions for the /p/, /t/, /k/ bursts, respectively. Note also how the rate masks (Fig. 4(C)) distinguish plosives /p/, /t/, /k/ from the long fricatives /s/, /ʃ/ by enhancing the regions outlined in the rectangle, namely the slow rates of the fricatives (<11 Hz) relative to the faster rates of the plosives. It should be noted that the classifier performance does not depend in any way on the *order* of the neural responses, which is solely used for analysis and display purposes.

The extent to which the neural phoneme representations can account for the perception of *individual* phoneme exemplars can be assessed by studying the pattern of pair-wise confusions by the classifier. Figure 5(A) shows the confusion matrix measured from classifications of the neural data. Labels along each row indicate the phoneme presented, and columns report the probability of the phoneme output by the classifier.[17,37] The classifier was trained on two sets of data. In a small set of 20 neurons, we succeeded in measuring responses to 330 s of speech (90 sentences) to be used in the training; these are shown in Fig. 5. In a larger set, training was based on responses from all sampled neurons in which at least 90 s of speech stimuli (30 sentences) were presented; these results are shown in Fig. 7 of the Appendix. In an ideal case in which all phonemes were accurately identifiable, we would expect to see a diagonal confusion matrix. Off-diagonal values represent misidentification. The phonemes

are arranged based on voiced-unvoiced and plosive, fricative, nasal consonant categories to facilitate comparison with a previous study of human perception[17,37] (replicated in Fig. 5(B)). The dashed boxes delineate the three major phoneme categories: plosives, fricatives, and nasals. In both neural and perceptual data, phonemes within each category—plosives (/p/, /t/, /k/), fricatives (/f/, /s/, /ʃ/), and nasals (/m/, /n/)—tend to be more confusable within the group than across categories. The correlation coefficient between the complete neural and perceptual matrices is 0.78 ($p=0.0002$, randomized t test). Ignoring the confusions between voiced and unvoiced consonants improves the similarity to 0.86, with a correlation of 0.95 for only the unvoiced consonants and 0.71 for their voiced counterparts. At least some of the difference between confusion matrices reflects noise due to limited sampling of neural responses, and/or limited data for training the phoneme classifiers. For example, when we computed the same confusion matrix for the entire population of 90 neurons (trained only on 90 s of speech), the correlation between neural and human confusion matrices fell to 0.70 ($p=0.001$), a change that may reflect the added dimensions and free parameters as new neurons are included in the analysis, while the amount of training data decreases at the same time. (Appendix; Fig. 7).

Alternatively, we explored the sensitivity of the classification in Fig. 5 to the number of neurons included (using the same training material). As expected, the results indicate that percentage of correct classification (averaged across all consonant phonemes) improves as the number of randomly selected neurons is increased (Appendix; Fig. 8). More detailed exploration of this issue should take into account the differential contribution of specific neurons to different phonemes, e.g., high BF neurons to the classification of /s/ and /ʃ/.

Finally, we also explored the extent to which both the neural and human confusion matrices are a reflection of the acoustic similarity (or "distances") among the phonemes at the level of the auditory spectrograms (see Sec. II). Figure 5 illustrates that such a phoneme "similarity matrix" fundamentally resembles the human and neural confusion matrices (with correlation coefficients of 0.66 and 0.93, respectively). In fact, the neural matrix encodes remarkably well the details of the phoneme acoustic similarity, such as the confusions between /v/ and the nasals /m/, /n/, and also between /ð/ and the voiced consonants /b/, /d/, /g/.

IV. DISCUSSION

Neuronal responses to continuous speech in the primary auditory cortex of the naive ferret reveal an explicit multidimensional representation that is sufficiently rich to support the discrimination of many American English phonemes. This representation is made possible by the wide range of spectro-temporal tuning in A1 to stimulus frequency, scale and rate. The great advantage of such diversity is that there is always a unique subpopulation of neurons that responds well to the distinctive acoustic features of a given phoneme and hence encodes that phoneme in a high-dimensional space.

As an example, consider the perception of the plosive consonant /k/ in a CV syllable, which is identified by a conjunction of several acoustic features: an initial silent voice-onset time (VOT), an onset burst of spectrally broad noise, and the direction of the following formant transitions.[23] Each of these features can be encoded in the cortical responses along different dimensions. Thus, neurons selective for broad spectra respond selectively to the noise burst. Rapid neurons respond well following the VOT, whereas directional neurons selectively encode the vowel formant transitions. In this manner, /k/ is encoded *robustly* by a rich pattern of activation that varies in time across the neural population. This neuronal activation pattern constitutes the phoneme representation in A1 and presumably forms the input to a set of neural "phoneme classifiers" in higher auditory areas. If one acoustic feature is distorted or absent, the pattern along the other dimensions (and hence the percept) remains stable.

We have focused here on describing a few prominent features of the response distributions that correspond to well-known distinctive acoustic features of the consonants considered.[24] There are clearly many other aspects and more details of the responses that reflect intricate articulatory gestures, contextual effects, or speaker-dependent variability that can only be reliably considered with a much larger sample of responses. One example is the distribution of the *directionality index* of the responses in the neighborhood of a consonant,[38] an attribute that would indicate whether the formants are upward or downward sweeping, or if they are converging towards or diverging away from a locus frequency.

Humans confuse the phonemes of their native tongue when placed in unusual or noisy contexts. Typically, phonemes that share some acoustic features tend to be more confusable than those that do not. This was confirmed by the similarity we found between the acoustic distance and the human confusion matrices. Similarly, since A1 responses in our naive ferrets also preserve the relative acoustic distances between the phonemes (as they would presumably for other complex sounds), we are led to the conjecture that human phoneme perception can (in principle) be explained in large measure by basic auditory representations such as the auditory spectrogram and the cortical spectro-temporal analysis common to many mammalian (and also avian) species.[6,9,10,39,40]

The representation of phonemic features across a population of filters tuned to BF, scale and rate suggests a strategy for improved speech recognition systems, and further study may reveal additional strategies for speech processing. However, many questions about the neural representation of phonemes still remain unanswered; for example, how can one extrapolate from such neurophysiological findings to the human perceptual ability to perceive phonemes categorically (also found in monkeys,[11] cats,[8] chinchillas,[3] birds[9] and rats,[41]), and to shift categorical boundaries arbitrarily between phoneme pairs?

While the human ability to discriminate native phonemes is the result of many years of training, naïve ferrets lack such a history. Hence ferret perception of clean phonemes may be comparable to humans perception of noisy phonemes. In both cases, confusion patterns would reflect the acoustic distances between the phonemes. However, if ferrets were trained to actively discriminate phonemes, it is

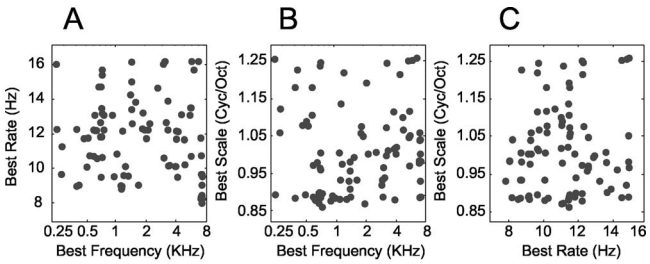


FIG. 6. Joint distribution of neural parameters. Joint distributions of best frequency, best rate (A), best frequency, best scale (B) and best rate, best scale (C) of 90 neurons.

likely that dimensions useful for this specific discrimination would be emphasized, creating the heightened sensitivity necessary to perform the task. This is presumably what happens in humans as they learn the phonemes of a given language, and what the classifier essentially simulates in our analysis when it learns the masks and boundaries that enable robust phoneme discriminations. Therefore, from a neural perspective, one may view the masks as either a subsequent layer of synaptic weights *or* as pattern of behaviorally driven plasticity of A1 receptive fields—the end result of perceptual learning in which neurons adapt their tuning along the dimensions appropriate for the phoneme discrimination task. This same general principle would apply to the discrimination between members of any set of complex sound, using frequency, rate and scale as well as additional cortical response dimensions, such as pitch, spatial location, and loudness.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant Nos. R01DC005779 and F32DC008453), the Air Force Office of Scientific Research and the Southwest Research Institute.

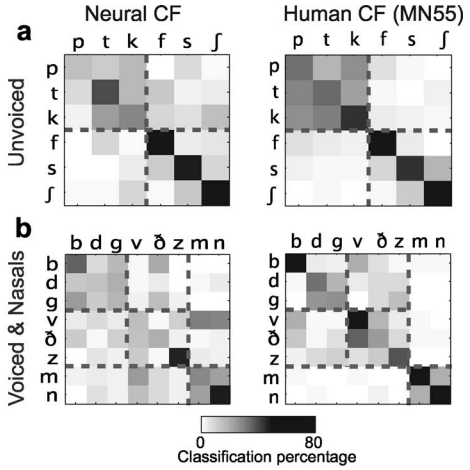


FIG. 7. Phoneme confusions from 90 neurons. (Left column) Consonant confusion matrices from neural phoneme classifiers using entire population of 90 neurons and 90 s of speech. (Right column) human psychoacoustic studies. Gray scale indicates the probability of reporting a particular phoneme (column) for an input phoneme (row). (a) Confusion matrices for unvoiced consonants. Dashed lines separate the plosives /p/, /t/, /k/ from fricatives /f/, /s/, /ʃ/. (b) Confusion matrices for voiced consonants. Dashed lines separate the plosives /b/, /d/, /g/ from fricatives /v/, /ð/, /z/ and the nasal consonants /m/ and /n/ from the rest.

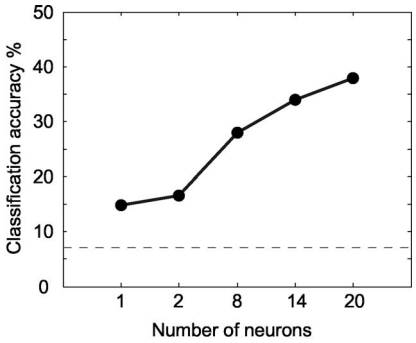


FIG. 8. Dependence of phoneme classification accuracy to the number of neurons. Classification accuracy as a function of the number of neurons used by the classifier. The dashed line indicates chance performance (7% for 14 phonemes) (see Sec. II for details).

APPENDIX

Here we provide additional information regarding: (A) uniformity of the sampling of the neural parameters; (B) the phoneme confusions from an SVM recognizer using a larger number of neurons, but with significantly fewer speech responses on which to train the classifier; (C) an exploration of the recognition accuracy with fewer numbers of neurons.

1. Joint distribution of neural parameters

To ensure that the response patterns in Figs. 2(A) and 3(A) are representative of the neural population in the cortex, we examined the uniformity of the coverage of the parameters of neural STRFs in our sample of 90 neurons. Specifically, the joint distributions of the different neural receptive field parameters (best frequency, best scale, and best rate) are shown in the three panels of Fig. 6, revealing fairly uniform coverage over all frequencies, bandwidths, and different dynamics (see Sec. II for further details).

2. Phoneme confusions from 90 neurons

Phoneme confusions derived from responses of the entire population of 90 neurons, but using only 90 s of speech, are displayed in Fig. 7. The correlation coefficient between the neural and human phoneme confusion (0.70; $p=0.001$) is still reasonable but is significantly less than that of the patterns in Fig. 8 (see method for more details).

3. Dependence of phoneme classification accuracy to the number of neurons

The number of neurons is a crucial variable in determining the accuracy of the phoneme classification as illustrated in the results of Fig. 8. Here the classification accuracy was computed as a function of the number of neurons used in training the classifier. For each condition, 100 random subsets of neurons were taken and the classification accuracy was averaged over all subsets. Note that the accuracy based

on the 20 neurons in this plot is still only at 37% (7% is chance performance). Presumably, adding more neurons increases the performance.

[1]R. P. Lippmann, "Speech recognition by machines and humans," Speech Commun. **22**, 1–15 (1997).
[2]S. Greenberg, W. Ainsworth, A. N. Popper, and R. R. Fay, *Speech Processing in the Auditory System* (Springer-Verlag, New York, 2004), Vol. 18.
[3]P. K. Kuhl and J. D. Miller, "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," Science **190**, 69–72 (1975).
[4]P. K. Kuhl and D. M. Padden, "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," J. Acoust. Soc. Am. **73**(3), 1003–1010 (1983).
[5]P. K. Kuhl and D. M. Padden, "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," J. Acoust. Soc. Am. **73**(3), 1003–1010 (1983).
[6]K. R. Kluender, A. J. Lotto, L. L. Holt, and S. L. Bloedel, "Role of experience for language-specific functional mappings of vowel sounds," J. Acoust. Soc. Am. **104**(6), 3568–3582 (1998).
[7]F. Pons, "The effects of distributional learning on rats' sensitivity to phonetic information," J. Exp. Psychol. Anim. Behav. Process **32**(1), 97–101 (2006).
[8]R. D. Hienz, C. M. Aleszczyk, and B. J. May, "Vowel discrimination in cats: Acquisition, effects of stimulus level, and performance in noise," J. Acoust. Soc. Am. **99**(6), 3656–3668 (1996).
[9]M. L. Dent, E. F. Brittan-Powell, R. J. Dooling, and A. Pierce, "Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (Melopsittacus undulatus)," J. Acoust. Soc. Am. **102**(3), 1891–1897 (1997).
[10]A. J. Lotto, K. R. Kluender, and L. L. Holt, "Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica)," J. Acoust. Soc. Am. **102**(2 Pt 1), 1134–1140 (1997).
[11]M. Steinschneider, Y. I. Fishman, and J. C. Arezzo, "Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey," J. Acoust. Soc. Am. **114**(1), 307–321 (2003).
[12]M. Steinschneider, D. Reser, C. E. Schroeder, and J. C. Arezzo, "Tonotopic organization of responses reflecting stop consonant place of articulation in primary cortex (A1) of the monkey," Brain Res. **674**, 147–152 (1995).
[13]M. Steinschneider, I. O. Volkov, Y. I. Fishman, H. Oya, J. C. Arezzo, and M. A. Howard, "Intracortical responses in human and monkey auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter," Cereb. Cortex **15**, 170–186 (2005).
[14]J. J. Eggermont and C. W. Ponton, "The neurophysiology of auditory perception: From single units to evoked potentials," Audiol. Neuro-Otol. **7**(2), 71–99 (2002).
[15]C. P. Hung, G. K. Kreiman, T. Poggio, and J. J. DiCarlo, "Fast readout of object identity from macaque inferior temporal cortex," Science **310**, 863–866 (2005).
[16]K. Walker, B. Ahmed, and J. W. Schnupp, "Linking cortical spike pattern codes to auditory perception," J. Cogn Neurosci., Oct 5 (Epub) (2007).
[17]G. Miller and P. Nicely, "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352 (1955).
[18]F. E. Theunissen, S. V. David, N. C. Singh, A. Hsu, W. E. Vinje, and J. L. Gallant, "Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli," Network **12**(3), 289–316 (2001).
[19]D. J. Klein, J. Z. Simon, D. A. Depireux, and S. A. Shamma, "Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex," J. Comput. Neurosci. **20**(2), 111–136 (2006).
[20]S. Seneff and V. Zue, "Transcription and alignment of the timit database," J. S. Garofolo, editor, National Institute of Standards and Technology (NIST), Gaithersburg, MD (1988).
[21]X. Yang, K. Wang, and S. A. Shamma, "Auditory representation of acoustic signals," IEEE Trans. Inf. Theory **38**(2), 824–839 (Special issue on wavelet transforms and multi-resolution signal analysis) (1992).
[22]S. V. David and J. L. Gallant, "Predicting neuronal responses during natural vision," Network **16**(2–3), 239–260 (2005).
[23]P. Ladefoged, *A Course in Phonetics*, 5th ed. (Harcourt Brace, Orlando, 2006).
[24]K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, MA, 1980).
[25]S. Shamma, "Speech processing in the auditory system. Part I: The representation of speech sounds in the responses of the auditory-nerve," J. Acoust. Soc. Am. **78**(5), 1612–1621 (1985).
[26]E. D. Young and M. B. Sachs, "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403 (1979).
[27]C. E. Schreiner, H. L. Read, and M. L. Sutter, "Modular organization of frequency integration in primary auditory cortex," Annu. Rev. Neurosci. **23**, 501–529 (2000).
[28]H. L. Read, J. A. Winer, and C. E. Schreiner, "Functional architecture of auditory cortex," Curr. Opin. Neurobiol. **12**(4), 433–440 (2002).
[29]D. A. Depireux, J. Z. Simon, D. J. Klein, and S. A. Shamma, "Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex," J. Neurophysiol. **85**, 1220–1234 (2001).
[30]L. A. Chistovich and V. V. Lublinskaya, "The center of gravity effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli," Hear. Res. **1**, 185–195 (1979).
[31]We emphasize that this response pattern is unlikely to be due to a nonuniform sampling of the scale and frequency variables, since no such bias in the joint distribution of the scale frequency is evident in Fig. 6(A). Furthermore, note that high scale neurons can be driven well by spectra with low frequencies as in phoneme /o/. The opposite is true for vowel /e/ where low scale units are driven well by high frequency energy.
[32]W. Klein, R. Plomp, and L. C. Pols, "Vowel spectra, vowel spaces and vowel identification," J. Acoust. Soc. Am. **48**(4), 999–1009 (1970).
[33]T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice* (Prentice–Hall, Englewood Cliffs, NJ, 2002).
[34]O. Deshmukh, C. Espy-Wilson, A. Salomon, and J. Singh, "Use of temporal information: Detection of the periodicity and aperiodicity profile of speech," IEEE Trans. Speech Audio Process. **13**(5), 776–786 (2005).
[35]D. Bendor and X. Wang, "The neuronal representation of pitch in primate auditory cortex," Nature (London) **436**, 1161–1165 (2005).
[36]V. N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995).
[37]J. B. Allen, *Articulation and Intelligibility* (Morgan and Claypool, 2005).
[38]D. Depireux, J. Z. Simon, and S. Shamma, "Measuring the dynamics of neural responses in primary auditory cortex," Comments in Theoretical Biology **5**(2), 89–118 (1998).
[39]N. Kowalski, D. Depireux, and S. Shamma, "Analysis of dynamic spectra in ferret primary auditory cortex: Prediction of single-unit responses to arbitrary dynamic spectra," J. Neurophysiol. **76**(5), 3524–3534 (1996).
[40]L. M. Miller, M. A. Escabi, H. L. Read, and C. E. Schreiner, "Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex," J. Neurophysiol. **87**, 516–527 (2002).
[41]C. T. Novitski *et al.*, Program 800.18/Poster E45, Neural coding of speech sounds in naïve and trained rat primary auditory cortex, Society for Neuroscience, Atlanta (2006).

Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise

Wookeun Song[a)]
Sound Quality Research Unit, Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7B, DK-9220 Aalborg East, Denmark and Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark

Wolfgang Ellermeier
Institut für Psychologie, Technische Universität Darmstadt, Alexanderstraße 10, D-64283 Darmstadt, Germany

Jørgen Hald
Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark

(Received 15 March 2007; accepted 18 November 2007)

The potential of spherical-harmonics beamforming (SHB) techniques for the auralization of target sound sources in a background noise was investigated and contrasted with traditional head-related transfer function (HRTF)-based binaural synthesis. A scaling of SHB was theoretically derived to estimate the free-field pressure at the center of a spherical microphone array and verified by comparing simulated frequency response functions with directly measured ones. The results show that there is good agreement in the frequency range of interest. A listening experiment was conducted to evaluate the auralization method subjectively. A set of ten environmental and product sounds were processed for headphone presentation in three different ways: (1) binaural synthesis using dummy head measurements, (2) the same with background noise, and (3) SHB of the noisy condition in combination with binaural synthesis. Two levels of background noise (62, 72 dB SPL) were used and two independent groups of subjects ($N=14$) evaluated either the loudness or annoyance of the processed sounds. The results indicate that SHB almost entirely restored the loudness (or annoyance) of the target sounds to unmasked levels, even when presented with background noise, and thus may be a useful tool to psychoacoustically analyze composite sources. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2822669]

PACS number(s): 43.66.Cb, 43.60.Fg, 43.66.Pn [RAL] Pages: 910–924

I. INTRODUCTION

The localization of problematic sound sources in a sound field is becoming increasingly important in areas such as automotive engineering and the aerospace, and consumer electronics industry. Typically, array techniques, such as near-field acoustic holography (NAH) (Maynard *et al.*, 1985; Veronesi and Maynard, 1987) and beamforming (Johnson and Dudgeon, 1993) have been employed to identify the noise sources of interest. In beamforming, a microphone array can be placed at a certain distance from the source plane and therefore it is easier to use in comparison with NAH, when there are obstacles close to the test object. Furthermore, the output of a beamformer is typically the sound pressure contribution at the center of the array in the absence of the array and this can be easily transformed to the sound pressure contribution at both ears by incorporating binaural technology (Møller, 1992). Hald (2005) proposed a scaling factor, which can be applied to the output of the delay-sum beamformer in order to obtain sound power estimates.

Since conventional physical measures, such as sound pressure or intensity, do not take into account how human listeners perceive sounds, there is growing interest in predicting specific psychoacoustic attributes from objective acoustical parameters (Ellermeier *et al.*, 2004b; Zwicker and Fastl, 2006). That also holds for microphone-array measurements in that it is desirable to identify problematic noise sources by mapping the sound fields of interest in terms of psychoacoustic attributes (Song, 2004; Yi, 2004) and by determining the directional contribution from individual sources (Song *et al.*, 2006).

Recently, spherical microphone arrays have been investigated for the recording and analysis of a sound field (Meyer, 2001; Meyer and Agnello, 2003; Petersen, 2004; Rafaely, 2004, 2005a). The major advantage of spherical microphone arrays where the microphones are distributed along the surface of a rigid sphere is that they permit steering a beam toward three-dimensional space with an almost identical beam-pattern, independent of the focused angle. Park and Rafaely (2005) validated the spherical microphone measurements in an anechoic chamber and measured the directional characteristics of reverberant sound fields. Rafaely (2005b) showed that spherical-harmonics and delay-sum beamforming provide similar performance when the highest spherical-harmonics order employed equals the product of the wave number and sphere radius. At lower frequencies, however, spherical harmonics beamforming allows the use of higher

a)Electronic mail: wksong@bksv.com

orders of spherical harmonics and thus better resolution. Note, though, that this improved resolution comes at the expense of robustness, i.e. the improvement of signal-to-noise ratio in the beamformer output.

Some studies examined the possibility of recording the higher-order spherical harmonics in a sound field and reproducing them by wavefield synthesis or ambisonics (Daniel *et al.*, 2003; Moreau *et al.*, 2006). But these methods require a large number of loudspeakers and a well-controlled environment such as an anechoic chamber. In order to render the recorded sound field binaurally, by contrast, the binaural signals obtained via either synthesis or recording can be played through a pair of headphones by feeding the left and right ear signal exclusively to each channel. Duraiswami and coworkers (Duraiswami *et al.*, 2005; Li and Duraiswami, 2005) studied theoretically how the free-field pressure obtained from spherical-harmonics beamforming (SHB) can be synthesized binaurally. The advantages of SHB, however, have not been demonstrated by means of psychoacoustic experiments in which subjective responses are collected to (a) validate the procedure, and (b) show that individual sources may successfully be isolated.

Therefore, the current study reports on a series of experiments to investigate the validity of using beamforming when auralizing a desired sound source in the presence of background noise or competing sources. The goals of this study are twofold:

1. To develop and verify the auralization of a desired source using beamforming. Procedures for estimating the pressure contribution of individual sources have already been suggested, but a scaling procedure will have to be developed to obtain the correct sound pressure level at the center of the array. To verify the procedure, the sound signals synthesized by beamforming will have to be compared with dummy head measurements.
2. To measure the effect of background noise suppression using beamforming on perceptual sound attributes, such as loudness and annoyance, derived from a listening experiment. To investigate the effects of noise suppression, the subjects' attention shall be controlled in such a way that they either judge the target sound (sound separated from background noise), or the entire sound mixture (including background noise).

To achieve these goals, the study employed ten stimuli from a study by Ellermeier *et al.* (2004a) which had been shown to cover a wide range with respect to loudness and annoyance. By playing them back in the presence of competing noise sources impinging from other directions, it may be investigated whether measuring the sound field with a spherical microphone array and processing it by SHB will recover the target source. Such a measurement protocol will be useful in situations in which only a desired source should be auralized, but in which background noise cannot be reduced or controlled during the measurement.

II. THEORETICAL BACKGROUND

A. Binaural synthesis

Reproduction of binaural signals via headphones is a convenient way of recreating the original auditory scene for the listener. The recording can be performed by placing a dummy head in a sound field, but it can also be synthesized on a computer. The binaural impulse response (BIR) from a "dry" source signal to each of the two ears in anechoic conditions can be described as (Møller, 1992):

$$h_{\text{left}}(t) = b(t) * c_{\text{left}},$$
$$h_{\text{right}}(t) = b(t) * c_{\text{right}}, \tag{1}$$

where the asterisk ($*$) represents convolution, b denotes the impulse response of the transmission path from a dry source signal to free-field pressure at the center of head position and c represents the impulse response of the transmission path from the free-field pressure to each of the two ears, i.e., head-related impulse response (HIR). The binaural signals can then be obtained by convolving a dry source signal with the binaural impulse response functions h. When using a spherical microphone array, SHB is able to approximate b for a given sound source by measuring the impulse response functions (IRF) from a dry source signal to each microphone of the array, and calculating the directional impulse response function (see Sec. II B 3) toward the dry source. The advantage of using SHB in comparison with a single-microphone measurement is the ability of focusing on a target source, i.e., obtaining the approximation of b, while suppressing background noise from other sources.

B. Spherical-harmonics beamforming

A theoretical description of SHB is presented in the following and a method to arrive at binaural auralization using SHB is proposed.

1. Fundamental formulation

For any function $f(\Omega)$ that is square integrable on the unit sphere, the following relationship holds (Rafaely, 2004):

$$F_{nm} = \oint f(\Omega) Y_n^{m*}(\Omega) d\Omega, \tag{2}$$

$$f(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F_{nm} Y_n^m(\Omega), \tag{3}$$

where the asterisk (*) represents complex conjugate, Y_n^m are the spherical harmonics, Ω is a direction, and $d\Omega = \sin\theta d\theta d\phi$ for a sphere. The spherical harmonics are defined as (Williams, 1999)

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta)\exp^{im\phi} \tag{4}$$

where n is the order, P_n^m are the associated Legendre polynomials, and $i=\sqrt{-1}$. Equation (3) shows that any square integrable function can be decomposed into spherical-harmonics coefficients. Rafaely (2004) defined the relation-

ship in Eqs. (2) and (3) as the spherical Fourier transform pair. The sound pressure on a hard sphere with radius $r=a$, $p(\Omega,a)$, and the directional distribution of incident plane waves, $w(\Omega)$, are square integrable and therefore we can introduce the two spherical transform pairs $\{p(\Omega,a),P_{nm}\}$ and $\{w(\Omega),W_{nm}\}$ according to Eqs. (2) and (3).

The goal of spherical-harmonics beamforming is to estimate the directional distribution $w(\Omega)$ of incident plane waves from the measured pressure on the hard sphere. To obtain a relation between the pressure on the sphere and the angular distribution of plane waves, we consider first the pressure on the hard sphere produced by a single incident plane wave. The pressure $p_\ell(\Omega_\ell,\Omega)$ on the hard sphere induced by a single plane wave with a unit amplitude and incident from the direction Ω_ℓ can be described as (Williams, 1999)

$$p_\ell(\Omega_\ell,\Omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}R_n(ka)Y_n^{m*}(\Omega_\ell)Y_n^m(\Omega), \tag{5}$$

where k is the wave number, and R_n is the radial function:

$$R_n=4\pi i^n\left[j_n(ka)-\frac{j_n'(ka)}{h_n^{(1)\prime}(ka)}h_n^{(1)}(ka)\right]. \tag{6}$$

Here, j_n is the spherical Bessel function, $h_n^{(1)}$ the spherical Hankel function of the first kind, and j_n' and $h_n^{(1)\prime}$ are their derivatives with respect to the argument. The total pressure $p(\Omega,a)$ on the hard sphere created by all plane waves can be found then by taking the integral over all directions of plane wave incidence. Using Eq. (5) and the spherical Fourier transform pair of $w(\Omega)$ we get

$$\begin{aligned}p(\Omega,r=a)&=\oint p_\ell(\Omega_\ell,\Omega)w(\Omega_\ell)d\Omega_\ell\\&=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}R_n(ka)Y_n^m(\Omega)\oint w(\Omega_\ell)Y_n^{m*}(\Omega_\ell)d\Omega_\ell\\&=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}W_{nm}R_n(ka)Y_n^m(\Omega).\end{aligned} \tag{7}$$

By comparing Eq. (7) with the spherical Fourier transform pair of $p(\Omega,a)$, the spherical Fourier transform coefficients of $w(\Omega)$ can be obtained as

$$W_{nm}=\frac{P_{nm}}{R_n(ka)}. \tag{8}$$

Substituting these coefficients in the spherical Fourier transform pair of $w(\Omega)$ results in

$$w(\Omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{P_{nm}}{R_n(ka)}Y_n^m(\Omega). \tag{9}$$

This shows that the directional distribution of plane waves can be obtained by dividing the pressure coefficients P_{nm} by the radial function R_n in the spherical Fourier domain.

We now introduce a set of M microphones mounted at directions Ω_i, $i=1,\ldots,M$, on the hard sphere with radius a. The Fourier transform expression for P_{nm} has the form of a continuous integral over the sphere, but the sound pressure is known only at the microphone positions. Therefore, we must use an approximation of the form:

$$P_{nm}\approx\tilde{P}_{nm}\equiv\sum_{i=1}^{M}c_i p(\Omega_i)Y_n^{m*}(\Omega_i). \tag{10}$$

The weights c_i applied to the individual microphone signals and the microphone positions Ω_i are chosen in such a way that

$$H_{mn\mu\nu}\equiv\sum_{i=1}^{M}c_i Y_\nu^{\mu*}(\Omega_i)Y_n^m(\Omega_i)=\delta_{\nu n}\delta_{\mu m} \tag{11}$$

$$\text{for } n\leqslant N,\nu\leqslant N,$$

where N is the maximum order of spherical harmonics that can be integrated accurately with Eq. (10). The value of N will depend on the number M of microphones. Therefore, the beamformer response for the direction Ω is calculated by substituting Eq. (10) in Eq. (9) and by limiting the spherical harmonics order to N:

$$b(\Omega)\equiv\sum_{i=1}^{M}\left[\sum_{\nu=0}^{N}\frac{1}{R_\nu(ka)}\sum_{\mu=-\nu}^{\nu}c_i Y_\nu^{\mu*}(\Omega_i)Y_\nu^\mu(\Omega)\right]p(\Omega_i). \tag{12}$$

2. Pressure scaling

Equation (12) is the typical beamformer output, but does not provide the correct pressure amplitude of an incident plane wave. Therefore, the goal here is to derive a scaling factor that gives rise to the correct estimate of the pressure amplitude. Ideally one may derive the scaling factor for each focus direction by calculating the beamformer response to a plane wave incident from that direction. Such a procedure would, however, significantly increase the computational effort. In particular at the lower frequencies, where the spatial aliasing is very limited, the "in-focus plane wave response" is fairly independent of the focus angle of the beamformer. One could therefore calculate the in-focus plane wave response for a single focus direction and apply that quantity for scaling of the beamformer output for all focus directions. But as shown in the following, it is possible to derive an analytical expression for the angle-averaged in-focus plane wave response. Use of that simple analytical expression requires less computation and provides a scaling that is better as an average over all directions.

We assume now a plane wave incident with a unit amplitude from the direction Ω_ℓ. By inserting Eq. (5) in Eq. (12) followed by use of Eq. (11) we get the beamformer response for an arbitrary focus direction Ω:

$$b(\Omega,\Omega_\ell)=\sum_{\nu=0}^{N}\sum_{\mu=-\nu}^{\nu}Y_\nu^\mu(\Omega)\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{R_n}{R_\nu}Y_n^{m*}(\Omega_\ell)\times\sum_{i=1}^{M}c_iY_\nu^{\mu*}(\Omega_i)Y_n^m(\Omega_i)=\sum_{\nu=0}^{N}\sum_{\mu=-\nu}^{\nu}Y_\nu^\mu(\Omega)\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{R_n}{R_\nu}Y_n^{m*}(\Omega_\ell)H_{mn\mu\nu}. \tag{13}$$

Only the in-focus response is needed, i.e., in the direction of plane wave incidence, $\Omega=\Omega_\ell$. This response will have a fairly constant amplitude and phase independent of the angle of the plane wave incidence, so it can be well represented by the angle averaged response $\bar{b}$. When we perform such an averaging, we can make use of the following orthogonality of the spherical harmonics:

$$\oint Y_\nu^\mu(\Omega)Y_n^{m*}(\Omega)d\Omega=\delta_{\nu n}\delta_{\mu m}. \tag{14}$$

Use of Eq. (14) in connection with Eq. (13) leads to the following expression for the angle averaged in-focus response,

$$\bar{b}\equiv\frac{1}{4\pi}\oint b(\Omega_\ell,\Omega_\ell)d\Omega_\ell=\frac{1}{4\pi}\sum_{\nu=0}^{N}\sum_{\mu=-\nu}^{\nu}H_{\mu\nu\mu\nu}. \tag{15}$$

And if in Eq. (15) we use Eq. (11), we get

$$\bar{b}\equiv\frac{(N+1)^2}{4\pi} \tag{16}$$

provided N is not larger than the spherical-harmonics order the beamformer was designed for, see Eq. (11). Equation (16) provides the average beamformer output, when focusing at infinite distance toward an incident plane wave of unit amplitude. If we wish the response to equal the amplitude of the incident plane wave, we therefore have to divide the output by $\bar{b}$ of Eq. (16). Notice that Eq. (15) shows a general approach, which may be applied to frequencies higher than those the microphone array is designed for. However, assuming no spatial aliasing (i.e., $R_n(ka)=0$ for $n>N$) the array beam pattern is independent of the focused direction. This means that Eq. (16) may be derived directly by substituting Eq. (11) in Eq. (13) and subsequently by using the spherical harmonics addition theorem [Rafaely, 2004, Eq. (20)].

So far we have considered plane wave incidence and focusing at an infinite distance. Consider instead the case of a monopole point source and focusing of the beamformer at the distance r_0 of the point source. The free-field sound pressure produced at the origin by this monopole is

$$p_{\text{center}}=\frac{e^{ikr_0}}{kr_0}. \tag{17}$$

The sound pressure at the microphone positions on the hard sphere can be expressed in spherical harmonics as in Eq. (5), but now with the following radial function (Bowman *et al.*, 1987):

$$R_n(ka)=4\pi i h_n^{(1)}(kr_0)\left[j_n(ka)-\frac{j_n'(ka)}{h_n^{(1)\prime}(ka)}h_n^{(1)}(ka)\right]. \tag{18}$$

Using the radial function of Eq. (18) in the beamforming processing, and averaging over all directions for the point source, leads to the same average in-focus beamformer output as in Eq. (16). If we wish the output to be the free-field pressure at the center of the array [Eq. (17)], then we have to scale the beamformer output by the following factor:

$$\frac{4\pi e^{ikr_0}}{(N+1)^2kr_0}. \tag{19}$$

3. Binaural auralization using SHB

Scaling the beamformer output by Eq. (19) provides the directional free-field pressure contributions at the center position in the absence of the array. Beamforming measurement and processing should then be taken for each sound event to be reproduced by the loudspeaker setup (described in Sec. III C): The type of sound cannot be changed after the measurement is done. But performing the measurement and processing for each sound is very time consuming. For this reason, directional impulse response functions will be calculated and used for simulating the total transmission from each loudspeaker input to each of the two ears.

Provided we measured the frequency response function (FRF) $t(\Omega_i)$ from a loudspeaker input to each microphone position on the sphere, the coefficients of the loudspeaker FRF's spherical Fourier transform T_{mn} can then be obtained by replacing $p(\Omega_i)$ by $t(\Omega_i)$ in Eq. (10),

$$T_{nm}\equiv\sum_{i=1}^{M}c_it(\Omega_i)Y_n^{m*}(\Omega_i). \tag{20}$$

Substituting Eq. (20) in Eq. (9) yields the directional response of the beamformer,

$$s(\Omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{T_{nm}}{R_n(ka)}Y_n^m(\Omega). \tag{21}$$

The directional impulse response can then be obtained by taking the inverse temporal fast Fourier transform (FFT) of $s(\Omega)$. If there is more than one loudspeaker, then the contribution from sound sources in other directions than the one focused on has to be taken into account and the total output of the beamformer at a particular direction Ω can be expressed as

$$y(\Omega)=\sum_{\ell=1}^{N_d}s_\ell(\Omega)x_\ell, \tag{22}$$

where N_d denotes the number of loudspeakers, $s_\ell(\Omega)$ represents the directional response of the ℓth loudspeaker in the focused direction Ω, and x_ℓ is the input signal of the ℓth loudspeaker. This will be a fairly good approximation since the contribution from other directions than those of the sources is negligible. Finally, the binaural signal can be obtained by multiplying $y(\Omega)$ with the HRTFs in the focused direction Ω.

Park and Rafaely (2005) suggested that the maximum spherical harmonics order in SHB should be limited to $N \leqslant ka$ in order to avoid noise originating from the high-order spherical harmonics. With the spherical microphone array used in this study, this would cause the beamformer output to become omnidirectional below 390 Hz. However, it was found that the order-limiting criterion can be relaxed in the following way without generating a high noise contribution:

$$N = \begin{cases} \lfloor ka \rfloor + 1, & \lfloor ka \rfloor + 1 \leqslant N_{\max} \\ N_{\max}, & \lfloor ka \rfloor + 1 > N_{\max} \end{cases} \tag{23}$$

where $\lfloor ka \rfloor$ represents the largest integer smaller than or equal to ka, and $N_{\max}$ is the maximum order of spherical harmonics for which the array can provide accurate integration [see Eq. (10)]. The number of spherical harmonics, $(N+1)^2$, should not exceed the number of microphones, and therefore $N_{\max}$ should be 7 in the current study where 64 microphones were used. Relaxing this condition by introducing higher spherical harmonic orders will reduce robustness and introduce greater uncertainties but our measurement and simulation experience shows that the use of spherical harmonic orders equal to $ka+1$ [as defined in Eq. (23)] produces only minor numerical instabilities.

C. Psychoacoustical considerations

The goal of the empirical part of the present study is to validate the beamforming method proposed, and—more specifically—to show how its use will help to psychoacoustically characterize target signals in a background of noise.

While, from a methodological perspective, it may be interesting to investigate the *detectability* of a target source in the presence of noise, in practice, the sources of interest are almost always well above threshold, or at best partially masked. Often, the focus of industrial applications is restricted to identifying the most problematic source in a mixture (Hald *et al.*, 2007; Nathak *et al.*, 2007), and to modify it to reduce its negative impact. Therefore, from a psychoacoustical perspective, some kind of suprathreshold subjective quantification of the salience of the target source in the background noise is called for. For the present investigation, the suprathreshold attributes of loudness and annoyance were chosen, since the former has been extensively studied (for reviews, see Moore, 2003; Zwicker and Fastl, 2006), and the latter is of particular relevance for noise control engineering (e.g., Marquis-Favre *et al.*, 2005; Versfeld and Vos, 1997).

As will be detailed in Sec. III, a between-subjects design was employed, investigating the two attributes in two independent groups of listeners. This was done in order to avoid potential carry-over effects that might produce artifactual correlations between loudness and annoyance.

Measuring the loudness or annoyance of the target stimuli under various conditions of partial masking required a scaling method that is relatively robust with respect to changes in context. A two-step category scaling procedure that uses both initial verbal labels to "anchor" the judgments and subsequent numerical fine-tuning possesses this property. It has been shown (Ellermeier *et al.*, 1991; Gescheider, 1997) to largely preserve the "absolute" sensation magnitudes even if the experimental context is changed. It was felt that the most widespread suprathreshold scaling procedure, namely Stevens' magnitude estimation, by virtue of the instructions to judge ratios of successive stimuli would encourage "relative" judgment behavior which might make it hard to compare the results across the different auralization methods used. Finally, the chance that in some conditions the target sounds might be entirely masked (yielding judgments of zero or undefined ratios), appeared to make ratio instructions unfeasible.

III. METHOD

A. Subjects

Twenty-eight normal-hearing listeners between the age of 21 and 34 (12 male, 16 female) participated in the experiment. All listeners were students at Aalborg University except for one female participant. The subjects' hearing thresholds were checked using standard pure-tone audiometry in the frequency range between 0.25 and 8 kHz and it was required that their pure-tone thresholds should not fall more than 20 dB below the normal curve (ISO 1998) at more than one frequency. The subjects were also screened for known hearing problems and paid an hourly wage for their participation. The subjects were not exposed to the sounds employed prior to the experiment.

B. Apparatus

The experiment was carried out in a small listening room with sound-isolating walls, floors, and ceiling. The room conforms with the ISO (1992) standard. The listeners were seated in a height-adjustable chair with a headrest. They were instructed to look straight ahead and were not allowed to move their head during the experiments. Their head movement was monitored by a camera installed in the listening room. Two monitors, one in the control room and the other in the listening room, were displayed at the same time with the help of a VGA splitter. A small loudspeaker placed in the control room played the same sound as the subject listened to so the experimenter could monitor the sound playback and the listener's behavior.

A personal computer with a 16-bit sound card (RME DIGI96) was used for D/A conversion of the signals. The sound was played with a sampling rate of 48 kHz and delivered via an electrostatic headphone (Sennheiser HE60) connected through an amplifier (Sennheiser HEV70) with a fixed volume control to assure constant gain. An external amplifier (t.c. Electronic Finalizer) between the headphone amplifier and the sound card controlled the playback level.

Playback and data collection were controlled by a customized software developed in C#. The software read the session files to assign a subject to the defined session, played the stimuli using the ASIO driver, collected subjects' responses, and wrote the responses into text files.

C. Measurements

The three different types of measurements, i.e., microphone, dummy head, and spherical microphone array, were

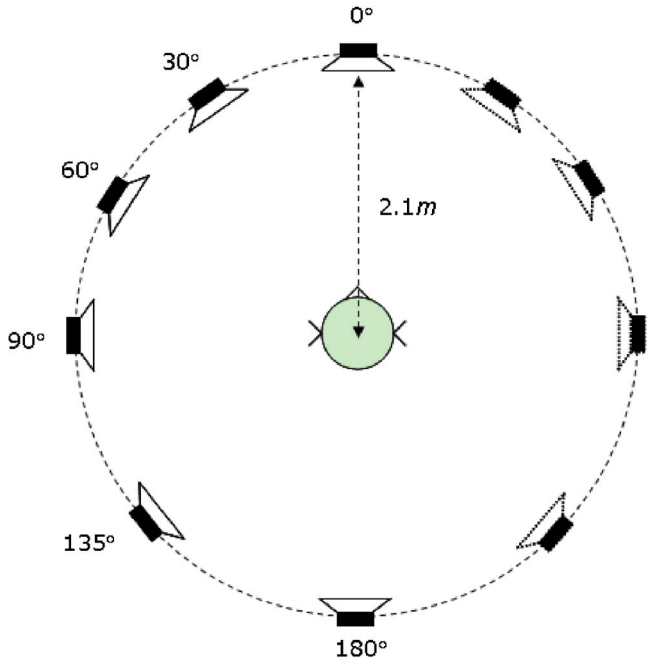


FIG. 1. (Color online) The loudspeaker setup in the anechoic chamber.

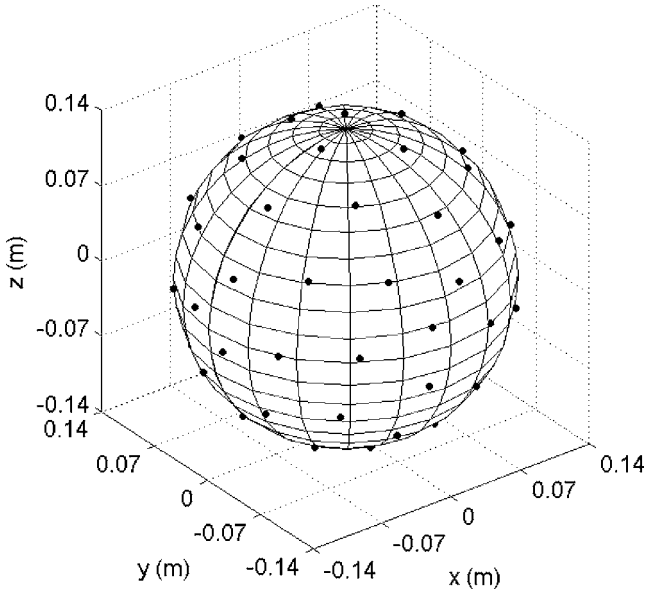


FIG. 2. The array consisting of 64 microphones placed on the hard surface of a sphere having a 14-cm radius. The dots on the sphere indicate the microphone positions.

performed in an anechoic chamber. Six loudspeakers were positioned at 2.1 m away from the center of the setup and their positions are shown in Fig. 1 (placed on the left-hand side). A setup of ten loudspeakers was simulated by flipping the four loudspeakers to the right-hand side. The loudspeaker in the frontal direction was used as the desired source through which the recorded sounds were synthesized and the rest of the loudspeakers served to create background noise sources. Since the microphone array and the required hardware was available for a very limited time, it was decided to record time data to permit changing some of the parameters without repeating the measurements. The input and output time data were recorded by means of the Data Recorder in the Brüel & Kjær software (type 3560) with a frequency range of 6.4 kHz. The loudspeakers were excited by random pink noise. The IRFs between speaker excitations and microphone responses were calculated using the autospectrum and cross spectrum of input and output and taking the inverse FFT of the calculated frequency response function in MATLAB. In order to remove the influence of reflections caused by the supporting structure and by other loudspeakers than the measured one, an 8-ms time window was applied to the calculated IRFs.

The loudspeaker responses were measured at the center position of the setup using a 1/2-in. pressure field microphone (Brüel & Kjær type 4134). The microphone was placed at 90° incidence to the loudspeakers during the measurement with the help of three laser beams mounted in the room. The measured IRFs were compared with the simulated ones to validate the recorded sound field using SHB. Responses of each loudspeaker at each ear of a dummy head were measured by placing the artificial head VALDEMAR (Christensen and Møller, 2000) at the center of the loudspeaker setup. Care was taken that the IRFs in both ears have the same delay when measuring the loudspeaker response in the frontal direction. The dummy head measurements were compared with the ones synthesized from SHB. The HRTFs employed in this study to perform binaural synthesis using SHB were taken from a database containing artificial-head HRTFs measured at 2° resolution (Bovbjerg *et al.*, 2000; Minnaar, 2001).

IRFs of each loudspeaker at the microphones of the array were obtained by positioning a spherical microphone array at the center of the setup. The position of the microphone array was adjusted carefully so that the beamformed sound pressure mapping could localize the correct angular position of each loudspeaker. The microphone array with a radius of 14 cm consisted of 64 microphones (1/4-in. microphone, Brüel & Kjær type 4951) that were evenly distributed on the surface of the hard sphere in order to achieve the constant directivity pattern in all directions. Figure 2 displays the position of microphones marked by dots on a sphere. In an earlier study, the array was applied to the issue of noise source localization, and the detailed specifications and characteristics of the array are described in Petersen (2004). In total, six loudspeaker positions and 64 microphones produced 384 IRFs.

The headphone transfer functions (PTFs) were measured in the listening room with the same dummy head and equipment used for the IRF measurement. The PTF measurement was repeated five times and after each measurement the headphone was repositioned. The upper panel of Fig. 3 shows that the repetitions have similar spectral shape in the frequency range of the investigation. An average of these five measurements was taken and smoothed in the frequency domain by applying a moving average filter corresponding to the 1/3 octave bands. The inverse PTF was calculated from the average PTF using fast deconvolution with regularization (Kirkeby *et al.*, 1998) (see the lower panel of Fig. 3).

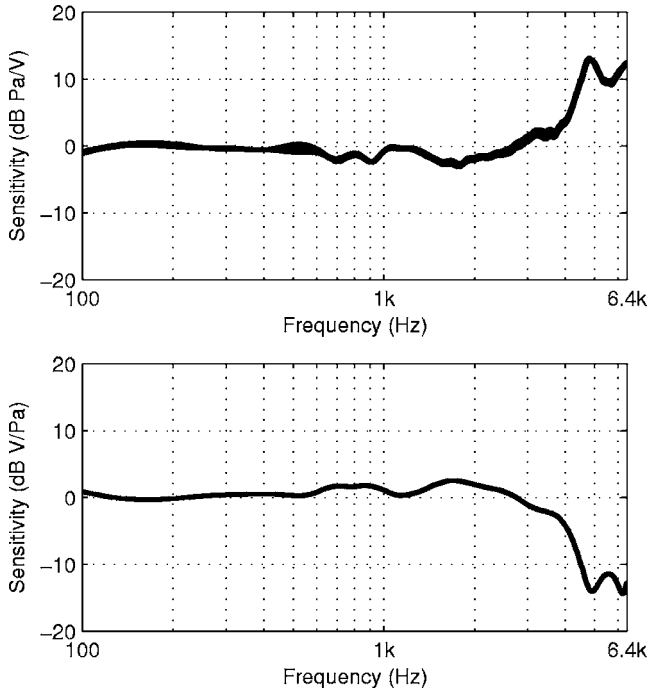


FIG. 3. Five headphone transfer functions (upper panel) measured at the left ear of the dummy head and the inverse filter derived (lower panel).

D. Stimuli

A set of 10 environmental and product sounds was selected from the 40 stimuli used by Ellermeier *et al.* (2004a). The ten sounds chosen were recorded in a sound-insulated listening room, except for two outdoor recordings of automotive sounds. About half of them were everyday sounds (e.g., door knocking, water pouring) and the rest were product sounds (e.g., kitchen mixer, razor, car). Both the perceived loudness and the annoyance of the selected sounds were almost equally spaced according to the attribute scales obtained in the reference study (Ellermeier *et al.*, 2004a). The length of stimuli varied from 0.8 to 5 s, and their overall sound pressure level at the recorded position ranged from 45 to 75 dB SPL. The sounds had a sampling rate of 44.1 kHz originally, but were resampled to 48 kHz in order to meet the requirements of the listening test program.

The desired source was synthesized to be located in the frontal direction and the remaining nine loudspeakers generated background noise. The selected sounds were convolved with the dummy head IRFs in the frontal direction to obtain the desired stimuli, and white noise having the same duration as the target sounds was convolved with the dummy head IRFs corresponding to the other nine directions. For each loudspeaker position, a new random sequence of white noise was created, and the signals convolved with the BIR at each ear were simply added to obtain the background noise. By doing so, the generated background noise was perceived to be diffuse. Two different levels of background noise were employed. The low level of background noise was adjusted to have the same sound pressure level as the bell sound (62 dB SPL), which was located in the middle of the attribute scale and the high level was defined to be 10 dB higher than the low one. In this way, the effect of the background noise level could be investigated. It was expected that some of the sounds would be partially masked by the background noise thereby affecting the attribute-scale responses.

The directional pressure contribution was obtained by recording the sound field using the spherical microphone array and applying SHB to the recorded data. Thus directional impulse response functions were calculated by using the IRFs at each of the microphone positions on the sphere as input to SHB processing. The resulting directional impulse response functions were convolved with HRTFs in the frontal direction to obtain the binaural IRFs, which still contain the contributions from background noise sources, though greatly reduced by the beamforming. In this case, the perception of the background noise is different from that with traditional binaural synthesis in that the noise is perceived to originate from the frontal direction. Thus in this study the influence of the level and perceptual quality of the background noise are confounded.

Subjects were asked to judge either the annoyance or the loudness of 50 stimuli, which were produced by combining three different processing modes (original, original+noise, SHB+noise), with two different noise levels, for the ten sounds selected. The same calibration tone as in the reference study (Ellermeier *et al.*, 2004a) was used and the level at the center position of the loudspeaker setup was adjusted to be 88 dB SPL when playing the calibration tone. A 100-ms ramp was applied to the beginning and end of each stimulus in order not to generate impulsive sounds. The inverse PTF was applied to the stimuli as a final step of the processing.

E. Procedure

The subjects were randomly assigned to one of two groups, one judging the loudness, the other the annoyance of the sounds. During the experiment, the participants were instructed to judge the entire sound event in one session, and the target sound only in the other. When judging the target sound only, they were asked to ignore the background noise and not to give ratings based on the direct comparison between the target sound and the background noise. The listeners were instructed to combine any of the components they heard for rating the entire sound mixture. These two ways of judging the sound attributes were chosen to check whether the effect of suppressing the background noise by SHB processing is different dependent on which part of a stimulus is being judged.

In each group, half of the subjects started judging the target sound only and proceeded to judge the entire sound (target plus background). The other half completed those two tasks in the opposite order. Note that each subject made but a single rating of each of the 50 experimental stimuli, i.e., there were no repetitions. The subjects spent approximately 1.5 h to complete the experiment. The participants were asked to judge either the loudness or the annoyance of the sounds by using a combined verbal/numerical rating scale, i.e., category subdivision (see Ellermeier *et al.*, 1991), shown in Fig. 4.

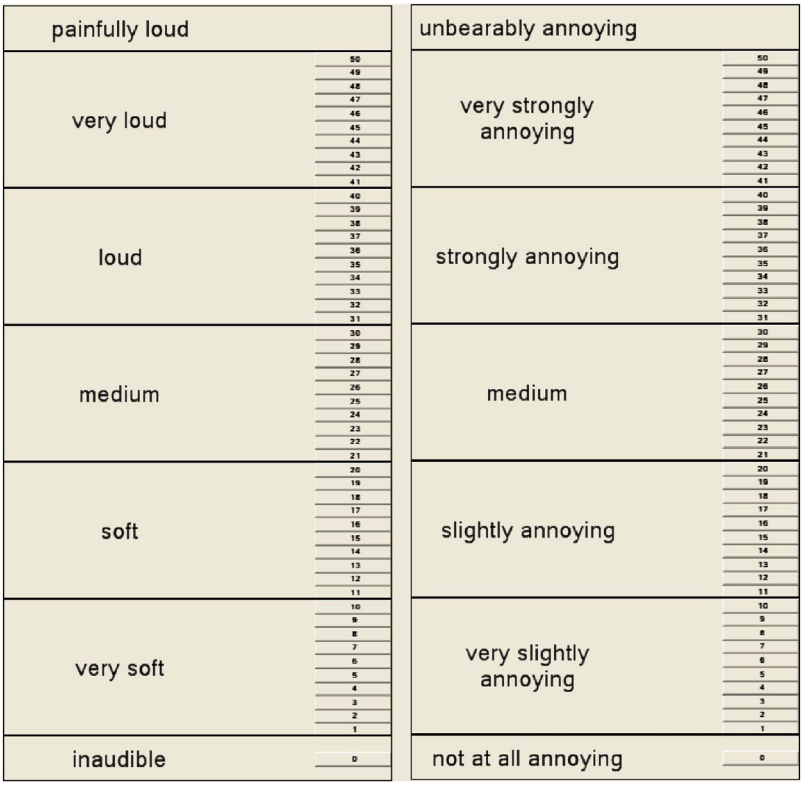


FIG. 4. (Color online) Category subdivision scales for loudness (left) and annoyance (right).

1. Training

There were two types of training prior to the main experiment. The goal of the first training unit was to give the subjects an opportunity of listening to the target sounds and to get an idea on what they had to focus, if the target was presented in background noise. To that effect, 20 buttons were displayed on a PC screen in two columns. The first column was labeled "target sound" and the second one "target sound+noise." The noise level was randomly selected from either the high- or the low-level condition. The participants were asked to first listen to the target sound only and then to the target sound with noise. During the training, the experimenter was present in the listening room and the subjects could ask any questions related to the understanding of the task. During the second training unit, the subjects received practice with rating the attribute, e.g., loudness or annoyance, of either target sound only or the entire sound dependent on which session they started with. The aim was to familiarize the participants with the procedure. This training unit consisted of only ten stimuli sampled to cover the entire range of sound pressure levels.

If the subjects started with judging the entire sound, they completed the training on the rating procedure first and were practiced in distinguishing target and background before starting with the second part of the experiment. Subjects, who judged the target sound in the first block, finished the two training units in a sequence prior to the main experiment.

2. Loudness scaling

For loudness scaling, the scale shown in Fig. 4 was displayed on a computer screen together with a reminder indicating whether they have to judge the target sound or the entire sound. The scale consisted of five verbal categories which were subdivided into ten steps and labeled "very soft" (1–10), "soft" (11–20), "medium" (21–30), "loud" (31–40), and "very loud" (41–50). The end points of the scale were used and labeled as "inaudible" (0) and "painfully loud" (beyond 50). On each trial, one sound was presented at a time, and the subjects were asked to decide which category the sound belonged to and then to fine-tune their judgment by clicking a numerical value within that category. That input started the next trial with a 1-s delay. The subjects were not allowed to make their rating while a sound was played. In order to avoid the situation where subjects rated the target sounds based on identifying them and recalling previous ratings, they were told that the level of the target sound might vary between trials.

3. Annoyance scaling

The format of the annoyance scale used was the same as that of the loudness scale (see Fig. 4). The five verbal categories were "very slightly annoying" (1–10), "slightly annoying" (11–20), "medium" (21–30), "strongly annoying" (31–40), and "very strongly annoying" (41–50). The lower end point was labeled as "not at all annoying" (0) and the higher one "unbearably annoying" (beyond 50). In the target sound only session, an "inaudible" button was placed below the category scale and subjects were asked to press it when they could not detect the target sound due to strong background noise.

The annoyance instructions were based on proposals by Berglund *et al.* (1975) and Hellman (1982). That is, a scenario was suggested, leading the participants to imagine a

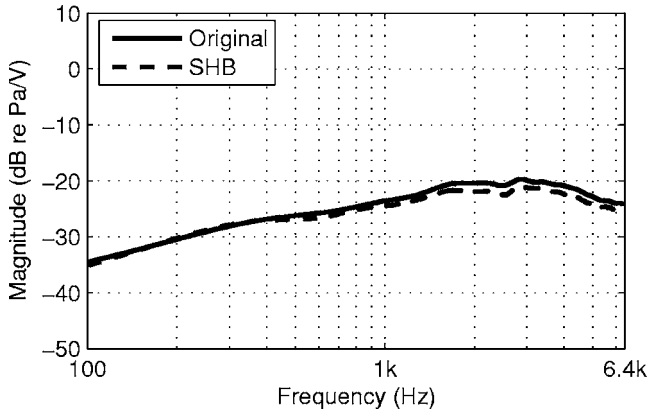


FIG. 5. Free-field loudspeaker response (30°): Measured (solid line) and synthesized (dashed line) using SHB.

situation in which the sounds could interfere with their activity: "After a hard day's work, you have just been comfortably seated in your chair and intend to read your newspaper."

IV. RESULTS

Here, the simulated sound field using SHB is compared with both the microphone and the dummy head measurements to illustrate the expected level difference induced by the beamforming in monaural and binaural responses. Moreover, the discrepancies in perceptual quality among the processing modes are demonstrated in both loudness and annoyance ratings obtained in the listening experiments.

A. Recording the sound field using SHB

In order to evaluate the success of the SHB simulation, the simulated and measured loudspeaker responses were compared. The loudspeaker responses at the 64 microphones placed on the sphere were measured and used as the input to the SHB calculation. The directional impulse response function toward each loudspeaker was calculated and compared with the direct measurement using a microphone positioned at the center position of the setup. The simulated and measured responses were compared in the frequency range of interest from 0.1 to 6.4 kHz, and an example for the loudspeaker placed at 30° is displayed in Fig. 5.

Generally, the agreement between the simulated and measured responses was good and the maximum discrepancy was approximately 2 dB in all loudspeaker directions. There was a tendency for the error to increase at high frequencies. In the current investigation where N_{max} is 7 and the radius of the array is 14 cm, spatial aliasing is expected above 2.7 kHz and thereby corrupts the spatial response. This could be the main reason for the inaccuracies at high frequencies.

The binaural response to the six loudspeakers was simulated by convolving the directional impulse response with the HRTF for the same direction as the loudspeaker (see Sec. II B 3). Subsequently, the simulated responses were compared with those measured with a dummy head and an example of the results is displayed in Fig. 6. The graphs represent the combination of the free-field loudspeaker response and the HRTF. In general, the two curves have similar shape and amplitude and the same tendency as for the free-field

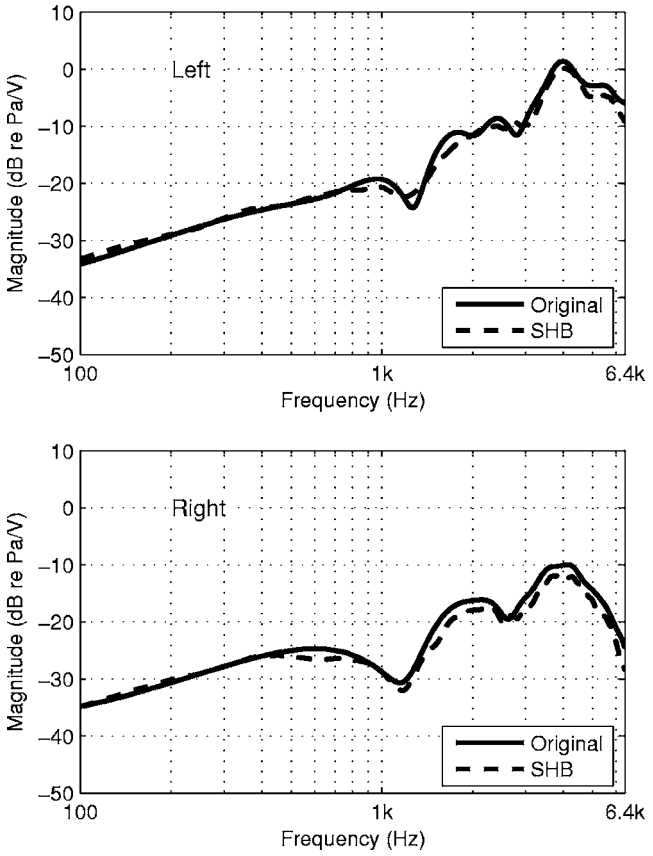


FIG. 6. Loudspeaker response (30°) at both ears: Measured (solid line) and synthesized (dashed line) using SHB.

response was observed, i.e., that the error grows slightly at high frequencies. These investigations confirm that the proposed method of combining SHB and binaural synthesis can generate binaural signals physically close to the measured ones.

B. Signal-to-noise ratio

The two measurement techniques, i.e., based on a dummy head and SHB, respectively, may be compared physically in terms of their monophonic signal-to-noise (S/N) ratios for each sound sample in the noisy conditions. Since the monophonic response for each loudspeaker was estimated both with a single microphone and with a microphone array, it was possible to separate the pressure contribution of the sound samples presented in the frontal direction from that of the noises in other directions. The monophonic S/N ratio for each sound sample was calculated simply by dividing the rms pressure of the signal by that of the noise.

Figure 7 shows the resulting S/N ratios of dummy head (original+noise) and SHB synthesis in both background noise conditions. The lower panel indicates the results of the low level noise condition and the upper panel the high level one. Notice that the S/N ratio of the bell sound is 0 and −10 dB in the original+noise condition for the low and high background noise levels, respectively (see Sec. III D). In general, the S/N ratio increases monotonically for the sounds ordered along the abscissa and there is a constant 10 dB difference between the low and high background noise con-

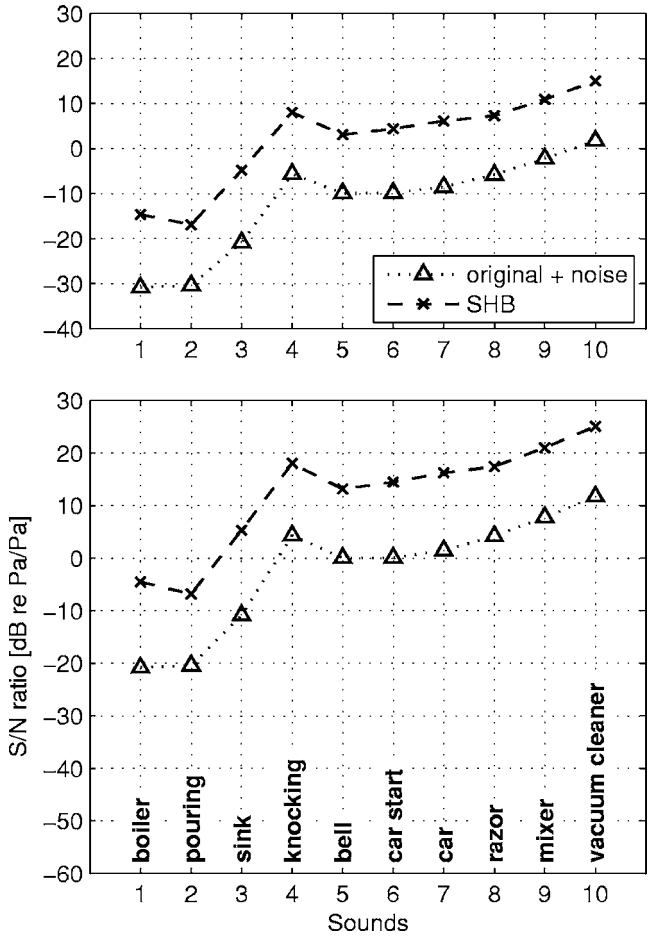


FIG. 7. Monophonic S/N ratio of dummy head (original+noise) and SHB measurements in the low (lower panel) and high (upper panel) background noise conditions.

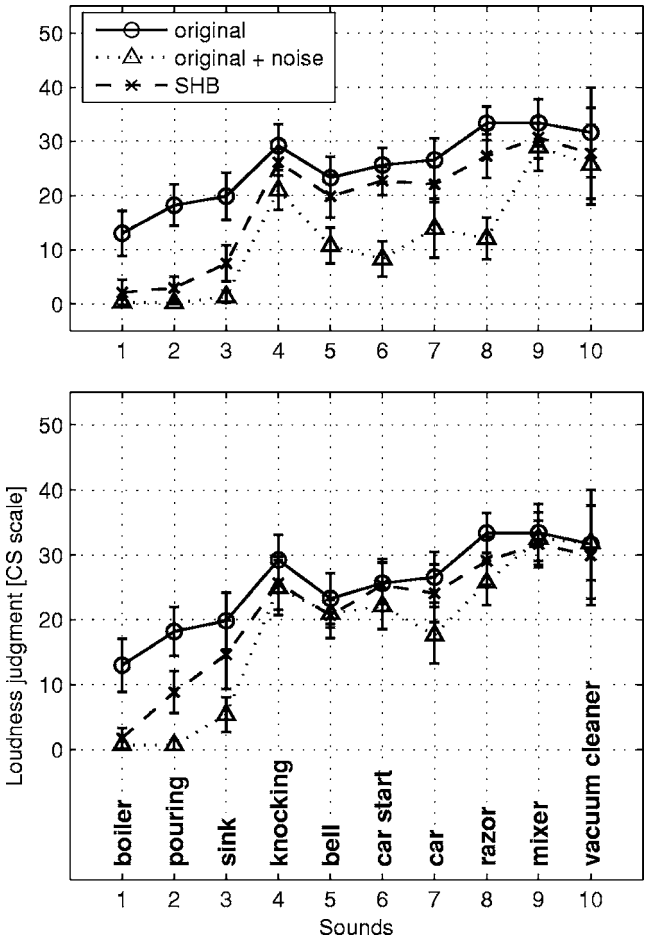


FIG. 8. Loudness judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners focused on the target sound only.

ditions. Thus, the effect of noise on the psychoacoustical scales is expected to be dominant in the low level sounds, e.g., for sound 1 to 3, for both measurement techniques. SHB increases the S/N ratio by approximately 15 dB for all sound samples, and thus the effect of the noise on loudness will be smaller for SHB in comparison with the dummy head technique.

C. Loudness scaling

The subjective loudness judgments were averaged across the 14 subjects for each sound in the three processing modes (original, original+noise, SHB) and 95%-confidence intervals were determined. The outcome is plotted in Fig. 8, for judgments of the target sound only, and in Fig. 9, for judgments of the entire sound event. The upper graph in Figs. 8 and 9 represents the high background noise condition and the lower graph the low background noise condition. Both graphs share the same ratings for the original condition plotted with solid lines. The sounds on the abscissa were arranged in the order of the mean ratings obtained in the reference study (Ellermeier *et al.*, 2004a). It appears that the present sample of subjects judged the knocking sound to be somewhat louder than in the reference study.

In the "target sound only" conditions (see Fig. 8), the target loudness was considerably reduced by adding noise to the target sound (compare the dotted and solid line) due to partial masking. It appears that SHB (dashed line in Fig. 8) partially restored the loudness of the target sounds. This was confirmed by performing a three-factor analysis of variance[1] (ANOVA) (Montgomery, 2004) with the two processing modes (SHB; original+noise), the two noise levels, and the ten sounds all constituting within-subjects factors. The analysis showed a highly significant main effect of processing mode [$F(1,13)=44.5$, $p<0.001$], as well as significant interactions ($p<0.001$) of processing mode with all other factors. That suggests that SHB did indeed suppress the background noise, thereby partially restoring loudness to the original levels. With the low-level masking noise (lower panel of Fig. 8) that was true for relatively "soft" target sounds (pouring and sink) while with the high-level masking noise (upper panel) the "loud" targets were the ones benefiting most from the release from masking produced by the SHB auralization. Notice on the other hand the difference between the two synthesis techniques in terms of S/N ratio is almost constant across different sounds and not dependent on the background noise level (see Fig. 7). Thus, a simple objective measure such as S/N ratio may not be suitable for predicting the effect of background noise suppression using beamforming on psychoacoustic attributes. Most subjects, however, could not de-

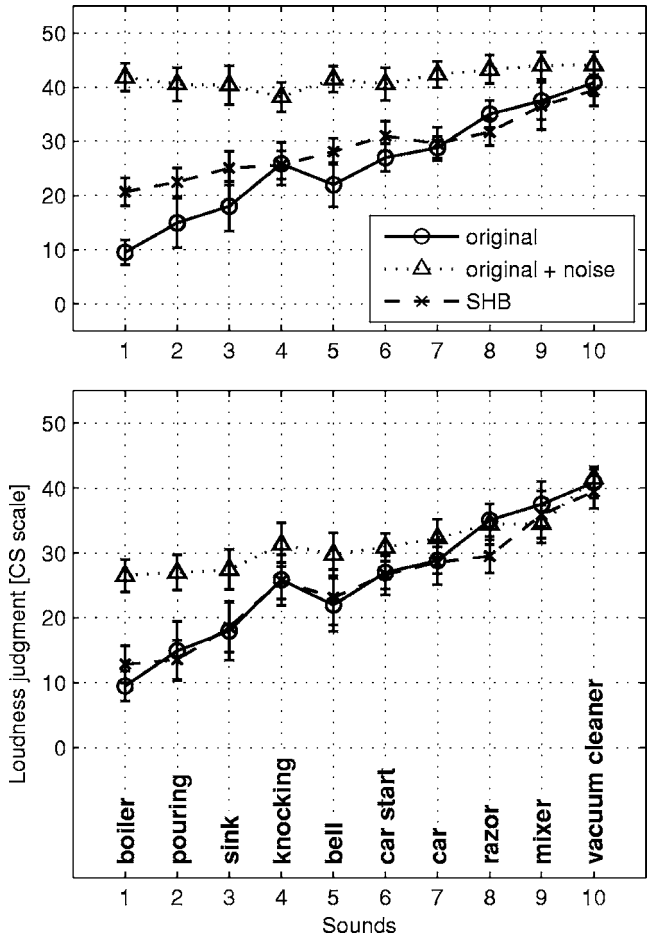


FIG. 9. Loudness judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners judged the entire sound event.

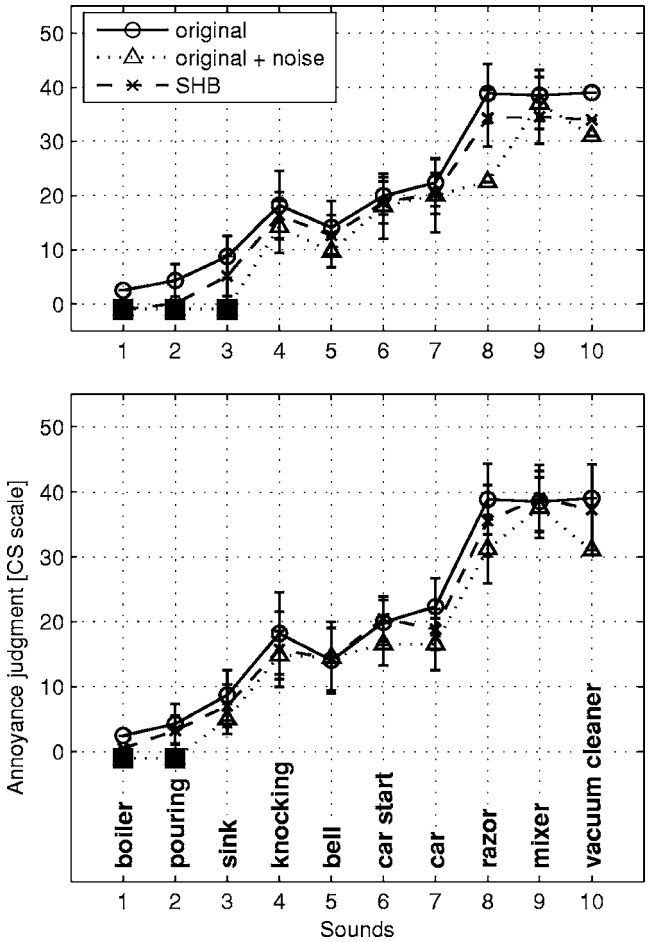


FIG. 10. Annoyance judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners focused on the target sound only. If the majority of the participants did not hear the target, the data points were marked with closed squares.

tect the "boiler" sound in both noise conditions since this sound was completely masked by background noise. This may be seen in Fig. 7 in that for the boiler sound a very low S/N ratio was obtained, even after the processing. Furthermore, the subjective ratings of the "knocking" sound almost coincided with those of the original sound, revealing that the subjects extracted this impulsive sound from the background much easier than other sounds. The high confidence intervals obtained for the vacuum-cleaner sound occurred because the target sound was so similar to background noise that it was difficult to distinguish one from the other.

Judging the entire sound event (see Fig. 9) made the suppression of the masker even more obvious in that the loudness functions for the original and SHB conditions almost coincide. That is, the SHB processing, though simulating a "noisy" listening situation, sufficiently suppresses the noise to approximate listening to the original targets in quiet. The significance of that effect was confirmed by a three-factor ANOVA showing a highly significant main effect of processing mode $[F(1,13)=229.7,\ p<0.001]$, and a processing mode × sound interaction $[F(9,117)=20.94,\ p<0.001]$. Only when the background noise level is high (upper panel in Fig. 9) and the target level is low, one can observe some noise "leaking" into the SHB condition, and the ratings to fall between those of the original sounds in quiet, and of the original sounds with noisy background.

These results imply that an evaluation of individual target sound sources in a background of noise or competing sources can be achieved by steering the beam toward the target sound source using SHB. The results are not dependent on whether listeners are asked to judge the loudness of the target sound or the entire sound event.

D. Annoyance scaling

The average annoyance data are depicted in Fig. 10 (target sounds rated) and Fig. 11 (entire sound rated) with the sound samples ordered in the same way as in Figs. 8 and 9. The lower plot shows the low noise condition and the upper the high noise condition. In the experimental condition in which the participants were asked to judge the annoyance of the target only (Fig. 10), and did not hear it (i.e. pressed the inaudible button, which occurred in 11.9% of all annoyance trials), a "−1" was recorded. To account for this qualitatively different response reflecting a lower, but indeterminate level of annoyance, the median of all responses was substituted for

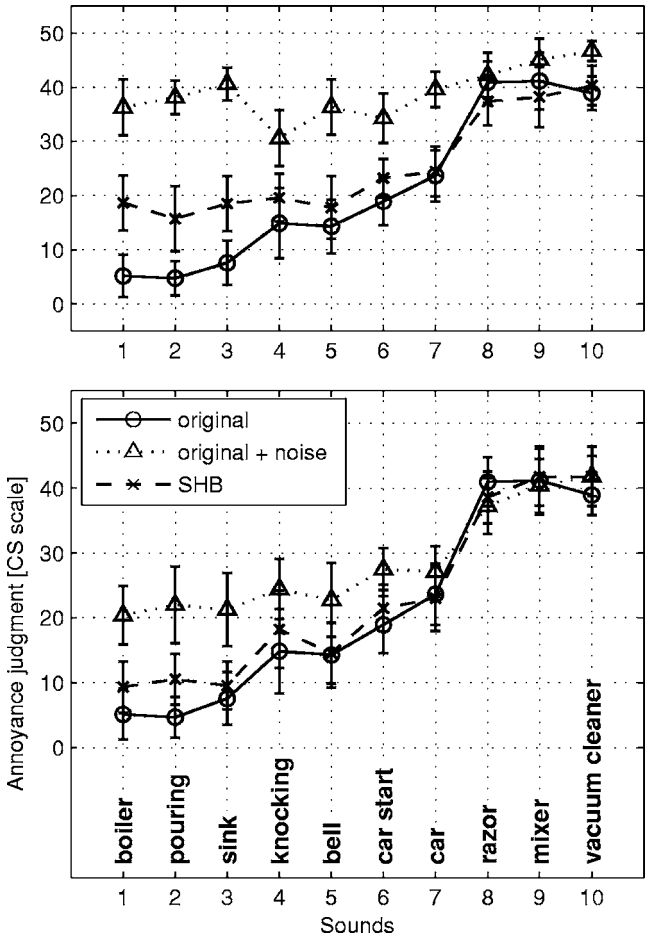


FIG. 11. Annoyance judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners judged the entire sound event.

the mean in all graphical depictions when a judgment of "not heard" had occurred. It is evident in Fig. 10 that in three (respectively, two) cases the majority of the participants did not hear the target when presented in background noise of high (respectively, low) level. In one instance, the target (boiler sound in high-level noise; top panel of Fig. 10) was not even detected after SHB processing.

When the subjects were asked to focus on the annoyance of the target sound only (see Fig. 10), it appears that the different processing conditions do not affect the ratings very much: The three curves in Fig. 10 (upper and lower panel) are hardly distinguishable. Furthermore, the level of background noise does not seem to affect the annoyance ratings significantly: $F(1,13)=2.2$, $p=0.166$. This indicates that even though the sounds were contaminated by noise, the subjects were able to judge the annoyance of the target sound consistently by identifying the target's annoying features. Therefore, the advantage of using SHB cannot be shown in this case, because in contrast to the results of the loudness scaling there is hardly a background noise effect in the first place. A four-factor analysis of variance with the two attributes (loudness and annoyance) constituting an additional between-subjects factor revealed that the annoyance ratings of the target sounds were significantly different from the corresponding loudness judgments, as was evident in the significant interactions of the attribute judged with the processing mode [$F(1,26)=5.22$, $p=0.03$], and the three-way interaction with processing mode and sound [$F(9,234)=2.36$, $p=0.014$].[2]

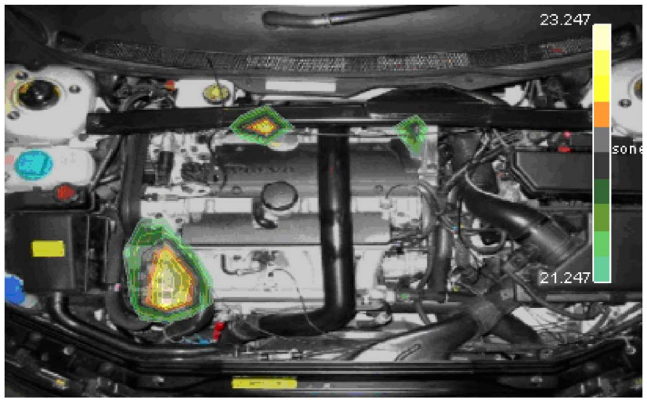


FIG. 12. (Color online) Loudness mapping of an engine compartment between 15 and 18 bark at 4000 rpm. See the text for details.

When the annoyance of the entire sound event is judged (see Fig. 11), the results are quite similar to those obtained for loudness. The effect of SHB processing is highly significant [$F(1,13)=158.43$, $p<0.001$], and the ratings obtained with SHB resemble those of the original sounds, with discrepancies emerging for the low-level sounds only. When loudness and annoyance are contrasted with respect to judgments of the entire sound, the interaction of the attributes with processing mode, sound level, and their combinations are no longer statistically significant (compared to judgments focusing on the targets, see the previous discussion), suggesting that the general pattern is quite similar for loudness and annoyance. This indicates that the annoyance percept is largely based on loudness if the subjects' attention is drawn to the entire sound mixture.

V. DISCUSSION

In an earlier investigation (Song, 2004), a comparison between traditional sound pressure maps and loudness maps derived from microphone array measurements was made and it was found that source identification in terms of psychoacoustic attributes improves the detectability of problematic sources. On the other hand, the mapping of some attributes cannot be derived due to the lack of metrics algorithms. Hence there is a need for auralizing the target sound identified as being devoid of background noise for further listening experiments.

Figure 12 shows the loudness map of an engine compartment of a passenger car with a five-cylinder, four-stroke engine. The engine was running at constant 4000 rpm without any external load applied. A 66-channel wheel array of 1 m diameter was mounted parallel to the car engine compartment at a distance of 0.75 m. In Fig. 12, it is obvious that the blank hole placed at the opposite side of the oil refill cap and the power steering pump at the lower left corner were the dominant sources in this operating condition. One might want to investigate attributes other than loudness, e.g., the

annoyance of those two sound sources, i.e., an attribute for which no agreed-upon objective metric exists. This could be done by having subjects judge the annoyance of the binaurally auralized sound of each target source at a time. This is a typical scenario for the use of source localization in practical applications in the automotive and consumer electronics industries.

Thus, the theoretical scaling of the SHB output derived in this paper and its experimental validation can be utilized for deriving a procedure to measure the auditory effects of individual sound sources. Since the method is based on steering the beam of a microphone array in three-dimensional space, no physical modifications of the sound field need to be made in contrast to typical dummy-head measurements. The details of the procedure proposed here will be discussed in the following.

A block diagram of the procedure for auralizing a target sound source binaurally is depicted in Fig. 13. This can easily be implemented together with classical beamforming applications in order to investigate problematic sources. Sound pressure signals are first measured at each microphone position on a rigid sphere, and converted to the frequency domain. Spherical harmonics beamforming is applied to steer a beam toward the target source (S_n) in each frequency band. A limited number of spherical-harmonics orders are used in SHB in order to avoid noise from the high-order spherical harmonics [see Eq. (23)].

The output of SHB, $P_{\mathrm{SHB}}(f)$, is scaled according to Eq. (19) to obtain the free-field pressure, $P_s(f)$, in the absence of the array with the assumption of a point source distribution on the source plane. The corresponding pressure time data, $P_s(t)$, are calculated by taking the inverse FFT of the scaled free-field pressure, $P_s(f)$. Finally, the binaural pressure signal can be acquired by convolving the free-field pressure with the HRTF in the source direction. Since HRTF databases are usually measured at discrete points on a full sphere, it is required to take either the nearest functions if the HRTFs are measured with a fine spatial resolution, or to interpolate between nearby points. The detailed procedure for interpolating HRTFs is described by Algazi *et al.* (2004) with respect to reproducing the measured sound field binaurally with the possibility of head tracking.

In the present paper, the analysis was restricted to the pressure contribution from a single direction. But, in many situations, such as in the professional audio industry, it is required to auralize distributed sources, i.e., the contribution from an area, and even the entire sound field as authentically as possible. An example of this kind of sound reproduction is the recording of sound fields in a car cabin while driving and reproducing it for head-tracked listening tests. In such situations, the measurements with a dummy head will have to be repeated many times in a well-controlled environment, which is very time-consuming, and may even be impossible due to lack of repeatability. Applying the procedure developed here to more than one direction enables the recording of full three-dimensional sound fields by one-shot array measurements and therefore allows listeners to turn their head while preserving the spatial auditory scene.

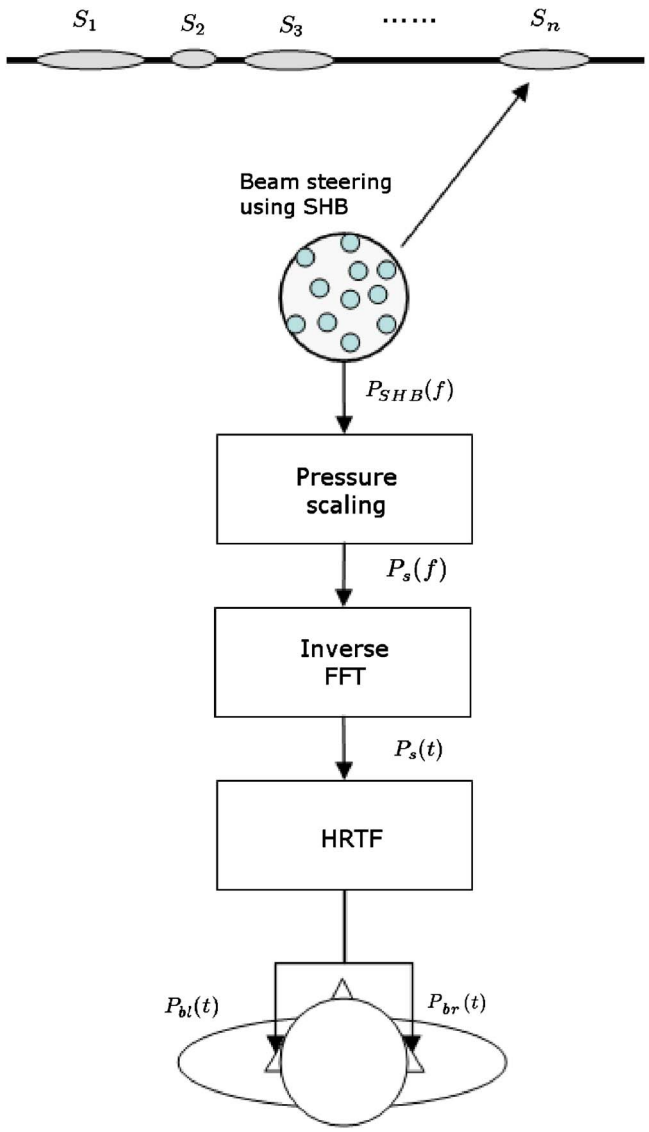


FIG. 13. (Color online) Binaural auralization of a desired sound source. Sound pressure signals are measured at each microphone position, and converted to the frequency domain. Spherical-harmonics beamforming (SHB) is applied to steer the beam toward a desired sound source and the output, $P_{\mathrm{SHB}}(f)$, is scaled to generate the free-field pressure, $P_s(f)$. The HRTF in the source direction is convolved with the pressure time signal, $P_s(t)$, obtained from the inverse FFT, and this results in binaural signals, $P_{bl}(t)$ and $P_{br}(t)$, at each ear.

VI. CONCLUSION

(1) A theoretical proposal was made for scaling the output of a spherical-harmonics beamformer, in order to estimate the free-field pressure at the listener's position in the absence of the microphone array. The comparison of measured and simulated responses (both monaural and binaural) to an array of loudspeakers showed that there is good agreement in the frequency range between 0.1 and 6.4 kHz. Notice that the simulated binaural responses were generated using an HRTF database, which was based on measurements using different instruments, physical structures, and a different anechoic chamber. Therefore, any differences between the two sets of responses contain the discrepancies between the earlier and current measurements.

(2) When the subjects judged target sounds partially masked by noise, their loudness was greatly reduced, but spherical harmonics beamforming managed to largely restore *loudness* to unmasked levels, except at low S/N ratios. By contrast, judgments of target *annoyance* were hardly affected by noise at all, suggesting that annoying sound features are extracted regardless of partial masking.
(3) When the subjects were asked to judge the entire sound events, SHB led to ratings close to those obtained in the original unmasked condition for both loudness and annoyance by suppressing background noise. The subjective judgments were largely explained by the percept of loudness: The loudness and annoyance data sets were highly correlated.
(4) The background noise level had significant effects by either producing partial masking (of targets) or contributing to the overall loudness (when the entire sound was judged). Judgments of target annoyance constituted an exception in that they were not affected by overall level.
(5) Implications of the study for sound-quality applications were sketched and a general procedure of deriving binaural signals using SHB was illustrated. The procedure can be used for evaluating the loudness and annoyance of individual sources in the presence of background noise.

ACKNOWLEDGMENTS

The experiments were carried out while the first two authors were at the "Sound Quality Research Unit" (SQRU) at Aalborg University. This unit was funded and partially staffed by Brüel & Kjær, Bang & Olufsen, and Delta Acoustics and Vibration. Additional financial support came from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

[1]Since the prerequisite normal-distribution assumption was met in the vast majority of experimental conditions—as verified by a Kolmogorov–Smirnov goodness-of-fit test—standard parametric analyses of variance were performed.

[2]In the ANOVAs, all not heard judgments were treated as values of −1. The pattern of statistical significances remained essentially the same when these problematic cases were excluded from the analysis.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2004**). "Motion-tracked binaural sound," J. Audio Eng. Soc. **52**, 1142–1156.

Berglund, B., Berglund, U., and Lindvall, T. (**1975**). "Scaling loudness, noisiness, and annoyance of aircraft noise," J. Acoust. Soc. Am. **57**, 930–934.

Bovbjerg, B. P., Christensen, F., Minnaar, P., and Chen, X. (**2000**). "Measuring the head-related transfer functions of an artificial head with a high directional resolution," in Audio Engineering Society, 109th Convention, Los Angeles, preprint 5264.

Bowman, J. J., Senior, T. B. A., and Uslenghi, P. L. E. (**1987**). *Electromagnetic and Acoustic Scattering by Simple Shapes* (Hemisphere, New York).

Christensen, F., and Møller, H. (**2000**). "The design of VALDEMAR—An artificial head for binaural recording purposes," in Audio Engineering Society, 109th Convention, Los Angeles, preprint 5253.

Daniel, J., Nicol, R., and Moreau, S. (**2003**). "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," in Audio Engineering Society, 114th Convention, Amsterdam, The Netherlands, preprint 5788.

Duraiswami, R., Zotkin, D. N., Li, Z., Grassi, E., Gumerov, N. A., and Davis, L. S. (**2005**). "High order spatial audio capture and its binaural head-tracked playback over head-phones with HRTF cues," in Audio Engineering Society, 119th Convention, New York, preprint 6540.

Ellermeier, W., Westphal, W., and Heidenfelder, M. (**1991**). "On the 'absoluteness' of category and magnitude scales of pain," Percept. Psychophys. **49**, 159–166.

Ellermeier, W., Zeitler, A., and Fastl, H. (**2004a**). "Impact of source identifiability on perceived loudness," in ICA2004, 18th International Congress on Acoustics, Kyoto, Japan, pp. 1491–1494.

Ellermeier, W., Zeitler, A., and Fastl, H. (**2004b**). "Predicting annoyance judgments from psychoacoustic metrics: Identifiable versus neutralized sounds," in Internoise, Prague, Czech Republic, preprint 267.

Gescheider, G. A. (**1997**). *Psychophysics: The Fundamentals* (Erlbaum, London, NJ).

Hald, J. (**2005**). "An integrated NAH/beamforming solution for efficient broad-band noise source location," in SAE Noise and Vibration Conference and Exhibition, Grand Traverse, MI, preprint 2537.

Hald, J., Mørkholt, J., and Gomes, J. (**2007**). "Efficient interior NSI based on various beamforming methods for overview and conformal mapping using SONAH holography for details on selected panels," in SAE Noise and Vibration Conference and Exhibition, St. Charles, IL, preprint 2276.

Hellman, R. P. (**1982**). "Loudness, annoyance, and noisiness produced by single-tone-noise complexes," J. Acoust. Soc. Am. **72**, 62–73.

ISO. (**1992**). "Audiometric test methods. 2. Sound field audiometry with pure tone and narrow-band test signals," ISO 8253-2, Geneva, Switzerland.

ISO. (**1998**). "Reference zero for the calibration of audiometric equipment. 1. Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," ISO 389-1, Geneva, Switzerland.

Johnson, D. H., and Dudgeon, D. E. (**1993**). *Array Signal Processing: Concepts and Techniques* (Prentice Hall, London).

Kirkeby, O., Nelson, P. A., Hamada, H., and Orduna-Bustmante, F. (**1998**). "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**, 189–194.

Li, Z., and Duraiswami, R. (**2005**). "Hemispherical microphone arrays for sound capture and beamforming," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, pp. 106–109.

Marquis-Favre, C., Premat, E., and Aubre, D. (**2005**). "Noise and its effects—A review on qualitative aspects of sound. II. Noise and Annoyance," Acust. Acta Acust. **91**, 626–642.

Maynard, J. D., Williams, E. G., and Lee, Y. (**1985**). "Nearfield acoustic holography. I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413.

Meyer, J. (**2001**). "Beamforming for a circular microphone array mounted on spherically shaped objects," J. Acoust. Soc. Am. **109**, 185–193.

Meyer, J., and Agnello, T. (**2003**). "Spherical microphone array for spatial sound recording," in Audio Engineering Society, 115th Convention, New York, preprint 5975.

Minnaar, P. (**2001**). "Simulating an acoustical environment with binaural technology—Investigations of binaural recording and synthesis," Ph.D. thesis, Aalborg University, Aalborg, Denmark.

Møller, H. (**1992**). "Fundamentals of binaural technology," Appl. Acoust. **36**, 171–218.

Montgomery, D. C. (**2004**). *Design and Analysis of Experiments* (Wiley, New York).

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing* (Academic, London).

Moreau, S., Daniel, J., and Bertet, S. (**2006**). "3D sound field recording with higher order ambisonics—Objective measurements and validation of a 4th order spherical microphone," in Audio Engineering Society, 120th Convention, Paris.

Nathak, S. S., Rao, M. D., and Derk, J. R. (**2007**). "Development and validation of an acoustic encapsulation to reduce diesel engine noise," in SAE Noise and Vibration Conference and Exhibition, St. Charles, IL, preprint 2375.

Park, M., and Rafaely, B. (**2005**). "Sound-field analysis by plane-wave decomposition using spherical microphone array," J. Acoust. Soc. Am. **118**, 3094–3103.

Petersen, S. O. (**2004**). "Localization of sound sources using 3D microphone array," Master's thesis, University of Southern Denmark, Odense, Denmark.

Rafaely, B. (**2004**). "Plane-wave decomposition of the sound field on a sphere by spherical convolution," J. Acoust. Soc. Am. **116**, 2149–2157.

Rafaely, B. (**2005a**). "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Process. **13**, 135–143.

Rafaely, B. (**2005b**). "Phase-mode versus delay-and-sum spherical microphone array processing," IEEE Signal Process. Lett. **12**, 713–716.
Song, W. (**2004**). "Sound quality metrics mapping using beamforming," in Internoise, Prague, Czech Republic, preprint 271.
Song, W., Ellermeier, W., and Minnaar, P. (**2006**). "Loudness estimation of simultaneous sources using beamforming," in *JSAE Annual Congress*, Yokohama, Japan, preprint 404.
Veronesi, W. A., and Maynard, J. D. (**1987**). "Nearfield acoustic holography (NAH). II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322.
Versfeld, N. J., and Vos, J. (**1997**). "Annoyance caused by sounds of wheeled and tracked vehicles," J. Acoust. Soc. Am. **101**, 2677–2685.
Williams, E. G. (**1999**). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London).
Yi, S. (**2004**). "A study on the noise source identification considering the sound quality," Master's thesis, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea.
Zwicker, E., and Fastl, H. (**2006**). *Psychoacoustics: Facts and Models*, (Springer, Berlin), 3rd Ed.

Spectral loudness summation for sequences of short noise bursts

Jesko L. Verhey[a] and Michael Uhlemann
AG Neurosensorik, Institut für Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 19 January 2007; accepted 16 November 2007)

Recent loudness data of single noise bursts indicate that spectral loudness summation depends on signal duration. To gain insight into the mechanisms underlying this duration effect, loudness was measured as a function of signal bandwidth centered around 2 kHz for sequences of 10-ms noise bursts at various repetition rates and, for comparison, for single noise bursts of either 10- or 1000-ms duration. The test-signal bandwidth was varied from 200 to 6400 Hz. For the repeated noise bursts, the reference signal had a bandwidth of 400 Hz. For the single noise bursts, data were obtained for two reference bandwidths: 400 and 3200 Hz. In agreement with previous results, the magnitude of spectral loudness summation was larger for the 10-ms than for the 1000-ms noise bursts. The reference bandwidth had no significant effect on the results for the single noise bursts. Up to repetition rates of 50 Hz, the magnitude of spectral loudness summation for the sequences of noise bursts was the same as for the single short noise burst. The data indicate that the mechanism underlying the duration effect in spectral loudness is considerably faster than the time constant of about 100 ms commonly associated with the temporal integration of loudness.
 [DOI: 10.1121/1.2822318]

PACS number(s): 43.66.Cb, 43.66.Mk, 43.66.Ba [AJO] Pages: 925–934

I. INTRODUCTION

Many natural sounds including speech show fluctuations in level over time (Nelken *et al.*, 1999) and can be interpreted as sequences of events of short duration (Florentine *et al.*, 1996). Natural sounds usually extend over a large frequency range. Thus, for the understanding of the perception of natural sounds, it is important to know how the auditory system combines information of these short events across frequency.

The purpose of this paper is to measure the influence of the spectrum on the loudness of sequences of noise bursts. Previous studies on loudness of pulse trains measured the level difference between a sequence of short signals and a continuous sound with the same spectrum (e.g., Garner, 1948; Port, 1963b; Reichardt, 1970; Grimm *et al.*, 2002; van Beurden and Dreschler, 2005). The present study, in contrast, quantifies the amount of spectral loudness summation for sequences of noise bursts by measuring the level difference between equally loud sequences of noise bursts with different bandwidths.

Several loudness studies have shown that the level of a narrow-band sound is higher than that of an equally loud broadband sound (e.g., Fletcher and Munson, 1933; Zwicker and Feldtkeller, 1955; Zwicker *et al.*, 1957; Scharf, 1959, 1962; Cacace and Margolis, 1985; Schneider, 1988; Hübner and Ellermeier, 1993; Verhey and Kollmeier, 2002). This effect, commonly referred to as spectral loudness summation, is believed to be the result of an analysis of the incoming sound by a bank of overlapping critical band filters followed by a compressive nonlinearity that transforms the intensity to specific loudness and a final loudness summation across channels (e.g., Fletcher and Steinberg, 1924; Zwicker and Scharf, 1965; Moore *et al.*, 1997; Zwicker and Fastl, 1999).

The loudness of repetitive pulses depends on temporal properties, such as the pulse repetition rate and the duration of each pulse. Studies on the loudness of single bursts have shown that the level of a short signal has to be higher than the level of an equally loud long signal with the same spectrum (Munson, 1947; Port, 1963a; Poulsen, 1981; Florentine *et al.*, 1996; Buus *et al.*, 1997; Epstein and Florentine, 2005). The effect, known as temporal integration of loudness, can be accounted for by assuming that the auditory system integrates the intensity (Poulsen, 1981) or the neural activity (Zwislocki, 1969) over time by means of a leaky integrator. There is considerable variability across studies in the exact value of the build-up time constant ranging between 25 ms (Niese, 1959; Reichardt, 1965, 1970) and 100 ms (Zwicker, 1966) or 200 ms (Munson, 1947; Stevens, 1961; Zwislocki, 1969). The variation between the different estimated time constants may be at least partly due to the difficulties experienced by subjects when comparing the loudness of two sounds if their durations are markedly different (Port, 1963a; Reichardt and Niese, 1970).

In order to facilitate the comparison of short pulses and stationary sounds, Port (1963a) used sequences of short pulses. Using pulse trains was, additionally, a step toward more natural sounds (Port, 1963b). In general, loudness tended to increase as the repetition rate was increased (Port, 1963b; Zwicker, 1966; Reichardt, 1970). Port (1963b) and Zwicker (1966) suggested that the decay time constant may be different from the build-up time constant. Both studies obtained 1.4–5 times larger decay time constants. Such an approach is realized in recent loudness models applicable to time-varying sounds (Ogura *et al.*, 1991; Glasberg and Moore, 2002).

[a]Electronic mail: jesko.verhey@uni-oldenburg.de

Temporal integration of loudness also changes with level (Poulsen, 1981; Florentine *et al.*, 1996; Buus *et al.*, 1997; Epstein and Florentine, 2005, 2006). Florentine *et al.* (1996) showed that the data could be predicted by assuming the same loudness ratio between equal-level long and short signals for all levels. This hypothesis was called equal-loudness ratio model (Florentine *et al.*, 1996; Buus *et al.*, 1997) or equal-loudness ratio hypothesis [(ELRH), Buus *et al.*, 1999; Epstein and Florentine, 2005]. The assumption of an equal-loudness ratio which is independent of the spectral properties of the sound (Buus *et al.*, 1997) agrees qualitatively with experimental findings and model structure by Zwicker (1969, 1977) that spectral loudness summation precedes the temporal integration stage.

Recent data on spectral loudness summation as a function of duration challenged the ELRH (Verhey and Kollmeier, 2002; Fruhmann *et al.*, 2003; Anweiler, 2005). The ELRH predicts a duration-independent level difference between equally loud narrow-band and broadband signals with the same duration. Using a loudness matching procedure, Verhey and Kollmeier (2002) found, in contrast to this prediction, a larger spectral summation for short (10 ms) than for long (1000 ms) signals with the same reference intensity (see also Fruhmann *et al.*, 2003). Anweiler and Verhey (2006) showed a similar duration effect in spectral loudness summation using a categorical loudness scaling procedure.

Recent data by Grimm *et al.* (2002) on the loudness of fluctuating sounds indicate that a duration-dependent spectral loudness summation may also affect the loudness perception of time-varying sounds. They found for narrow-band signals that, at equal loudness, the level of an amplitude modulated signal had to be slightly higher than that of a stationary sound with the same bandwidth. In contrast, for the modulated broadband signal, it was slightly lower than the level of the equally loud unmodulated signal with the same bandwidth. They argued that this bandwidth effect was due to the duration effect in spectral loudness summation. They based the hypothesis on the assumption that an amplitude modulated sound could be interpreted as a sequence of short signals. Motivated by this interpretation of the results of Grimm *et al.* (2002), van Beurden and Dreschler (2005) measured the level difference between sequences of noise bursts and equally loud stationary noise signals for two bandwidths. In agreement with Grimm *et al.* (2002), they found smaller level differences for broad signals than for narrow signals, further supporting the hypothesis that a duration-dependent spectral loudness summation also affects the loudness of temporally varying sounds.

The results of the previous studies provide indirect evidence for differences in the spectral loudness summation of time-varying and stationary signals. The present study measured the level differences between equally loud sequences of noise bursts with different bandwidths in order to investigate (i) if the magnitude of spectral loudness summation of pulse trains was similar to that of a single noise burst and (ii) if spectral loudness summation depends on the repetition rate. The reference bandwidth (400 Hz) of the loudness matching procedure used in the present study was the same as in Fruhmann *et al.* (2003). Fruhmann *et al.* (2003) and Anweiler and Verhey (2006) argued that the loudness matching procedure used in the present study might be slightly biased by the choice of the reference. In order to test this hypothesis, spectral loudness summation was measured in an additional experiment for short and long signals with reference bandwidths of 400 and 3200 Hz.

II. EXPERIMENT I: SPECTRAL LOUDNESS SUMMATION OF SINGLE AND REPEATED NOISE BURSTS

A. Methods

1. Stimuli and apparatus

All stimuli were bandpass-noise signals geometrically centered around 2 kHz. Signal-iterated low-noise noise was used, i.e., signals were divided by their Hilbert envelope and then restricted to their original bandwidth (Verhey and Kollmeier, 2002; Kohlrausch *et al.*, 1997). The standard deviation σ relative to the mean m of the intensity of signal-iterated low-noise noise is approximately half of the ratio σ/m for Gaussian noise (Verhey and Kollmeier, 2002). Thus, by using single-iterated low-noise noise instead of Gaussian noise, it was possible to present the signal at higher levels without overload distortions.[1]

In general, signals consisted of sequences of noise bursts of 10-ms duration each including 2.5-ms cosine-square ramps at on- and offset of the noise burst. The stimuli were derived from a 45 500-sample buffer of a bandlimited noise with the desired level and bandwidth by multiplication with the desired temporal envelopes. Running noise was used, i.e., a new noise buffer was generated for each interval. The repetition rates of the noise bursts were 3, 6, 9, 12, 25, 50, 70 and 100 Hz. The interburst interval was equal to the inverse of the repetition rate reduced by the duration of the noise burst, e.g., 30 ms for a repetition rate of 25 Hz. For a repetition rate of 100 Hz, the noise bursts abut each other, i.e., the onset ramp of the next burst starts immediately after the end of the offset ramp of the previous noise burst. The number of noise bursts per interval was equal to the repetition rate in Hertz, e.g., three for a repetition rate of 3 Hz. In addition to the sequences of noise bursts, single noise bursts of either 10- or 1000-ms duration including 2.5-ms cosine-square ramps at on- and offset were used. Test signal bandwidths were 200, 400, 800, 1600, 3200, and 6400 Hz. The bandwidth of the reference signal was 400 Hz. The reference level was 70 dB SPL.[2]

Stimuli were generated digitally at a sampling rate of 44.1 kHz. A standard personal computer controlled stimulus generation and presentation and recorded results using the AFC software package for MATLAB developed at the University of Oldenburg. Stimuli were D/A converted (RME ADI-8 PRO), amplified (Tucker-Davis HB7), and presented monaurally via a Sennheiser HDA 200 headphone with free-field equalization. Subjects were seated in a sound-insulated booth. Responses were given by pressing the corresponding button on a computer keyboard.

2. Subjects

Twelve normal-hearing subjects (6 male, 6 female; aged 24 to 45 years) participated in the experiments. One subject (S8) was a member of the research group, the others were paid volunteers. None of them had previous experience in loudness matching experiments. All subjects had normal audiograms with hearing thresholds lower than or equal to 15 dB HL at the standard audiometric frequencies between 125 Hz and 8 kHz.

3. Procedure

Stimuli with different bandwidths were matched in loudness to a reference signal with fixed bandwidth and level using an adaptive two-interval, two-alternative forced-choice procedure. In each trial the subjects heard two sounds, the reference and the test signal. Loudness was only matched between signals with the same temporal properties, i.e., the same repetition rate for the pulse trains and the same duration for the single noise bursts. The end of the last noise bursts of a sequence determined the duration of the interval. For example, the duration of an interval was 990 ms for a repetition rate of 50 Hz and 677 ms for a repetition rate of 3 Hz.[3] For single noise bursts, the duration of the interval was equal to the duration of the noise bursts.

In general, the duration of the silent interval between the signal intervals was 500 ms. For repetition rates of 3 and 6 Hz, it was increased to 750 ms in order to facilitate the perceptual separation of the second interval from the first interval of a trial. A silence interval of 750 ms was also used for the single 10-ms noise bursts.

Test and reference signals were presented in random order and with equal *a priori* probability. The subjects indicated which signal was louder by pressing the corresponding key on a keyboard. A simple one-up one-down procedure was used, which converges at the 50% point of the psychometric function (Levitt, 1971). If the subject indicated that the test signal was the louder one, its level was reduced in the next trial, otherwise it was increased. At the beginning, the step size was 8 dB. It was divided by two after each reversal in the adaptive level tracking procedure. At a step size of 2 dB, it was held constant for the next four reversals. The level difference between test and reference signal at equal loudness for one track was determined by calculating the median of the levels at these last four reversals. Three matches were obtained for each subject and pair of stimuli. For each of the three matches a different starting level of the test signal at the beginning of the track was used. The starting levels were 10 dB above, 10 dB below, and at the reference level (0-dB difference).

To reduce biases that occur when stimuli from only one stimulus pair are matched in loudness in a series of trials, several interleaved adaptive tracks were used (Florentine *et al.*, 1996; Verhey and Kollmeier, 2002). Hence, concurrent loudness matches were obtained for all stimulus pairs tested. On each trial, the track was chosen randomly from all possible tracks, i.e., from all tracks that had not yet been terminated. To ensure that the interleaved tracks converged at roughly the same time the random choice of tracks was further restricted by requiring that each track be selected once in random order before any track could be reselected. If one track was terminated, the rule was applied to the choice of trials from the remaining unterminated tracks. Only tracks with the same temporal property were interleaved in a series of trials. It was chosen randomly from all possible temporal properties, i.e., the eight repetition rates and the short and long signals.

4. Evaluation of the data

The average of the data for the three different starting levels was considered the point of subjective equality. To examine the statistical significance of the effects of stimulus variables and differences among subjects, an analysis of variance (ANOVA) for repeated measures was performed (MATLAB 7.0). The dependent variable was the level difference between equally loud test signal and reference signal. Post hoc Scheffé tests for contrast (MATLAB 7.0) were performed to explore sources of significant effects and interactions.

B. Results

1. Individual data

The results of the loudness matching experiment with single noise bursts are shown in Fig. 1. Figure 1 shows individual data and standard deviations for 12 subjects. As a function of the bandwidth, Fig. 1 shows the level difference ΔL between the test signal and the 400-Hz-wide reference signal needed to produce the sensation of equal loudness. A positive ΔL means that the test signal was higher in level than the reference. Squares represent the results for the 1000-ms signals and the triangles those for the 10-ms signals. For each duration, the data points are connected with lines to indicate that they all produce the same loudness.

As expected, the data show no level difference when test signal and reference signal had the same bandwidth (400 Hz). In general, all subjects obtained a negative ΔL for test-signal bandwidths larger than the reference bandwidth and a level difference ΔL close to zero for a test-signal bandwidth of 200 Hz. The decrease in level at the point of equal loudness as the bandwidth increases reflects the effect of bandwidth on loudness known as spectral loudness summation.

Large individual differences were observed in the amount of spectral loudness summation. For example, for the 1000-ms signals (Fig. 1, squares), the difference between the 200- and the 6400-Hz wide signals ranged from 6 dB (S10) to 36 dB (S12). For most of the subjects the magnitude of spectral loudness summation was larger for 10-ms signals than for 1000-ms signals. As expected, especially large differences were obtained for large test-signal bandwidths. Seven of the 12 subjects (S1, S2, S7, S8, S9, S10, and S11) showed a difference of more than 3 dB between the two durations for the largest two bandwidths.

Figure 2 shows the level difference ΔL between test signal and the equally loud reference for sequences of repeated noise bursts with the lowest (3 Hz, inverted triangles) and the highest (100 Hz, circles) repetition rates used in the present study. For comparison, the data for the single noise

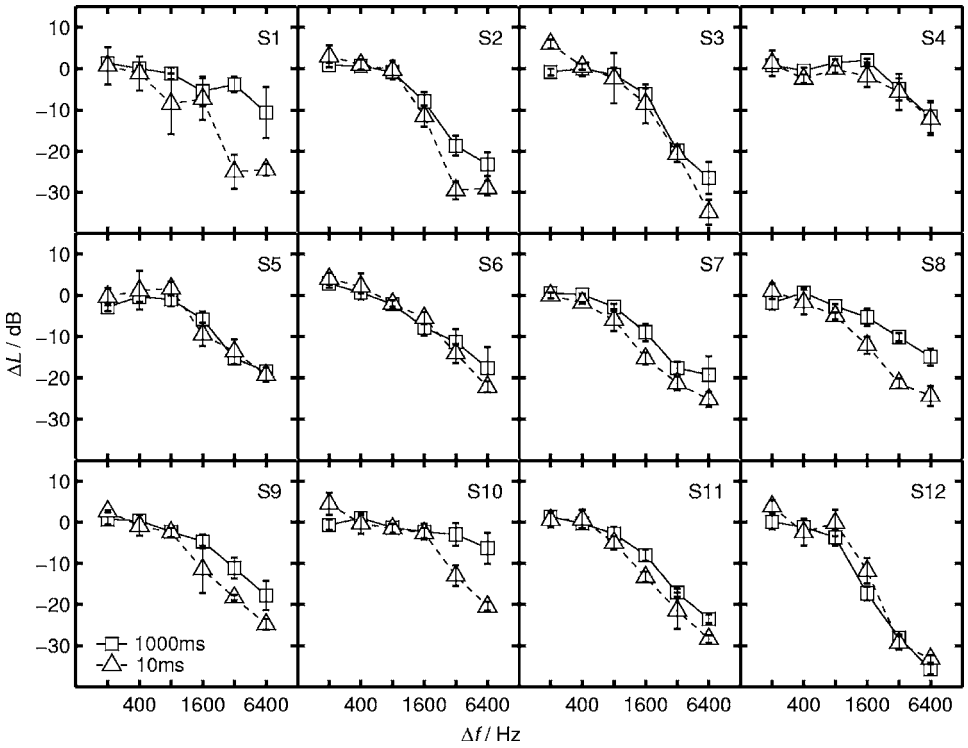


FIG. 1. Loudness summation for single noise pulses from twelve subjects. The reference bandwidth was 400 Hz. Signal duration was either 1000 ms (squares) or 10 ms (triangles). The level differences between the level of test and reference signal needed to obtain equal loudness are plotted as a function of the bandwidth of the test signal. The vertical bars show ±1 s.d. of the mean.

bursts are redrawn from Fig. 1 (solid line: 1000 ms; dashed line: 10 ms). The levels at the point of equal loudness for the repeated noise bursts are in the range of levels obtained for the single noise bursts. For subjects showing more than 3 dB lower levels for the short single noise bursts than for the long signals at the largest two bandwidths, there is a tendency of a larger spectral loudness summation for the 100-Hz pulse train than for the 1000-ms signal. For some subjects (e.g., S1, S8, and S10) the magnitude of spectral loudness summation for the 3-Hz repetition rate was larger than for the 100-Hz repetition rate.

2. Group data

Figure 3 shows average loudness matching data and standard errors for 12 subjects. Test signal levels that were perceived equally loud as the reference signal are plotted as a

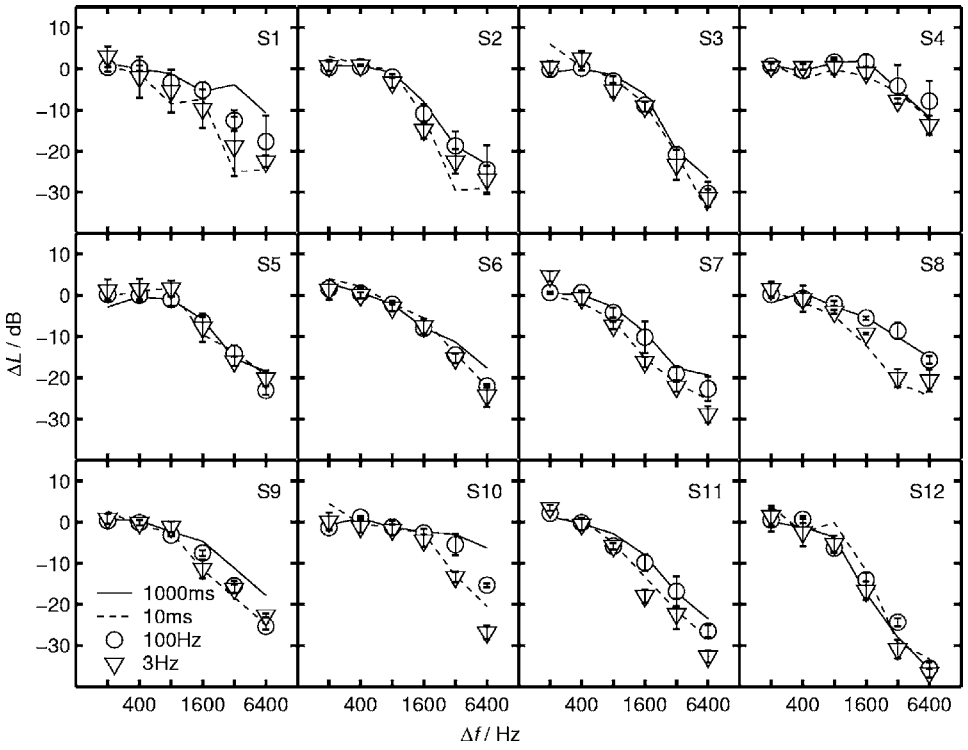


FIG. 2. Loudness summation for sequences of bandpass-noise pulses from 12 subjects. The reference bandwidth was 400 Hz. The repetition rate was either 3 Hz (inverted triangles) or 100 Hz (circles). The duration of each pulse was 10 ms. The level differences between the level of test and reference signal needed to obtain equal loudness are plotted as a function of the bandwidth of the test signal. The vertical bars show ±1 s.d. of the mean. In addition, the data from Fig. 1 for the single bursts are shown with thin dotted (10 ms) and thin solid (1000 ms) lines.

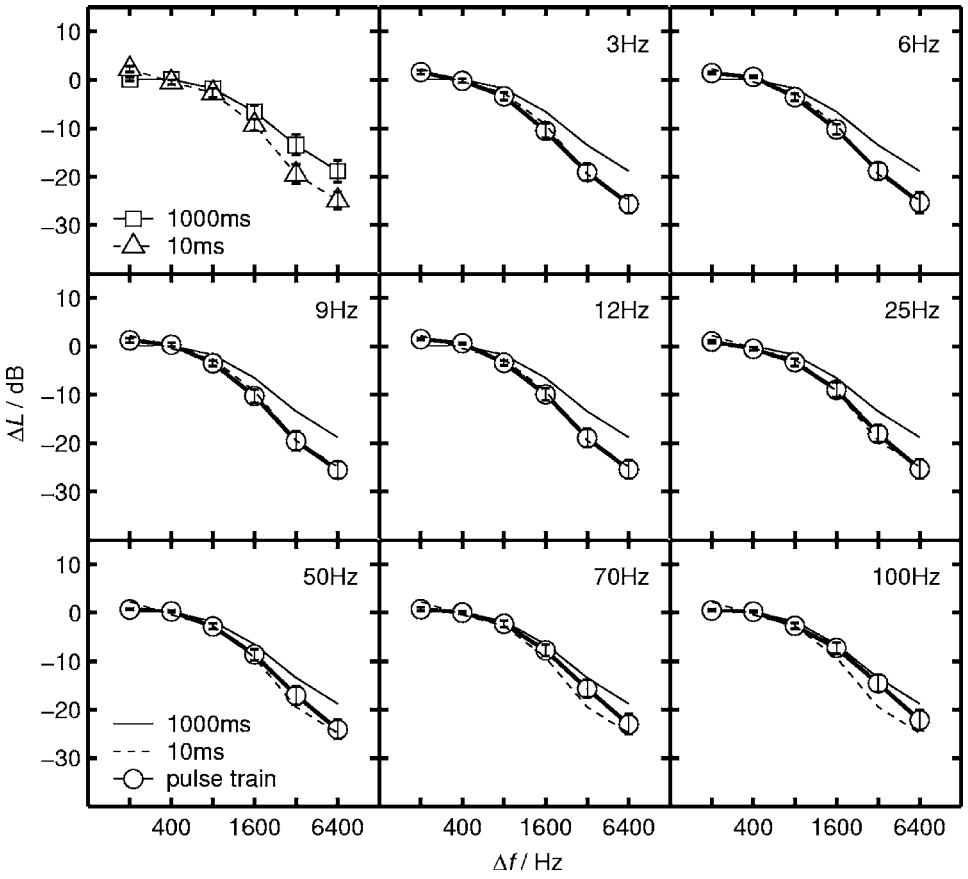


FIG. 3. Average spectral loudness summation. The top left panel shows data for single noise bursts of 10-ms (triangles, dashed line) and 1000-ms (squares, solid line) duration. The other panels show the data for sequences of noise bursts (circles, thick solid line). The repetition rate is indicated in the upper right corner of each panel. For comparison, the data for the single bursts are redrawn using the same line style as in the top left panel. Each panel shows the level difference between the test signal and the reference (Δf=400 Hz, L=70 dB SPL) at equal loudness as a function of the bandwidth. The reference bandwidth was 400 Hz. The vertical bars show ±1 s.e. of the mean.

function of the bandwidth of the test signal. Lines connect points of equal loudness. The top left panel shows the data for the single noise bursts (10 ms: triangles, 1000 ms: squares). Each of the other panels shows spectral loudness summation for sequences of noise bursts at different repetition rates (circles connected with a bold solid line) together with the data for the single bursts from the top left panel.

The average level differences between equally loud signals with the smallest (200 Hz) and the largest bandwidth (6400 Hz) is 19 dB for the 1000-ms noise burst and 27 dB for the single 10-ms noise burst. Approximately the same level differences (difference <3 dB) as for the single 10-ms noise burst were obtained for the pulse trains with repetition rates in the range from 3 to 50 Hz. The amount of spectral loudness summation decreased as the repetition rate was increased beyond 50 Hz. The level difference between the equally loud 200- and the 6400-Hz stimuli was only about 23 dB for a repetition rate of 100 Hz.

These effects were supported by a three-way ANOVA for repeated measures (Table I). Both stimulus variables (bandwidth and repetition rate) had significant effects on the level difference between the test and the reference signal ($p<0.01$). A Scheffé post hoc test showed no significant effect between the 1000-ms signal and the sequence of noise bursts repeated at a rate of 100 Hz ($p>0.05$). The results for all other repetition rates were significantly different from those for the long signal ($p<0.05$). The significant interactions of bandwidth and repetition rate, bandwidth and subject, repetition rate and subject, and bandwidth, repetition rate, and subject ($p<0.01$) showed that the amount of spectral loudness summation depended on the repetition rate of the signal and that this interaction in turn depended on the subject.

III. EXPERIMENT II: SPECTRAL LOUDNESS SUMMATION FOR TWO REFERENCE BANDWIDTHS

A. Methods

Stimulus generation, experimental setup, and data analysis were the same as in the previous experiment. The stimulus duration was either 10 or 1000 ms. In contrast to previ-

TABLE I. Three-way analysis of variance for repeated measures of loudness matching for normal-hearing subjects. The dependent variable is the level difference between test and reference signal at equal loudness. Bandwidth of the test signal (bw; six steps: 200, 400, 800, 1600, 3200, and 6400 Hz) and repetition rate [rep; ten steps: 1 (single 10-ms noise burst), 3, 6, 9, 12, 25, 50, 70, 100, and 133 Hz (1000-ms noise burst)] are fixed factors. Subject (sub; twelve steps: S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12) is a random factor.

Source	Sum Sq.	d.f.	Mean sq.	F	Prob>F
sub	16 437.9384	11	1 494.358	4.9132	≤0.0001
rep	1 706.8588	9	189.651	8.6863	≤0.0001
bw	186 551.0009	5	37 310.2002	128.4886	≤0.0001
sub×rep	2 161.4968	99	21.8333	2.7082	≤0.0001
sub×bw	15 970.7602	55	290.3775	36.0189	≤0.0001
rep×bw	2 255.6148	45	50.1248	6.2176	≤0.0001
sub×rep×bw	3 990.5963	495	8.0618	1.2284	≤0.0022

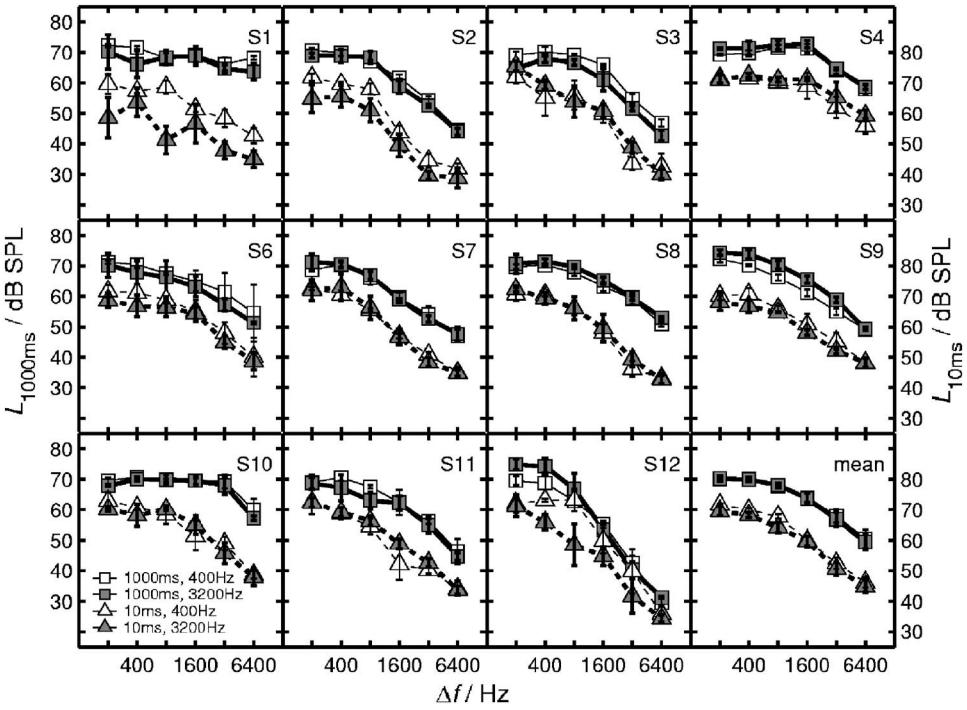


FIG. 4. Spectral loudness summation from 11 subjects for single noise burst with a duration of 1000 ms (squares) and 10 ms (triangles). The bottom right panel shows average data. All other panels show individual data. Open symbols indicate the level of the test signal which was perceived as equally loud as the 400-Hz wide reference at a level of 70 dB. The levels at equal loudness for the 3200-Hz reference bandwidth are shown with solid grey symbols. The level of the 3200-Hz wide reference was adjusted individually on the basis of the data of the first experiment (see Fig. 1). The data for the 10-ms signals are shifted downwards as indicated by the different scales on the right (for 10 ms) and the left margin (for 1000 ms). The vertical bars in the lower right panel (mean data) show ±1 s.e. The vertical bars in all other panels show ±1 s.d.

ous experiments, a loudness matching procedure with two different references was used. In addition to the 400-Hz wide reference at a level of 70 dB, a reference with a bandwidth of 3200 Hz was used. The level of the 3200-Hz wide reference was adjusted individually for each duration of the single noise bursts using the estimated levels of the 3200-Hz wide signal needed to produce the same loudness as 400-Hz wide reference from the first experiment. The tracks for all test-signal bandwidths and the two reference bandwidths were interleaved, i.e., a total of 12 concurrent loudness matches were obtained. Signal duration was either 10 or 1000 ms. All but one (S5) of the subjects of the first experiment participated in the second experiment.

B. Results

Figure 4 shows the level-bandwidth combinations which were perceived as equally loud. Solid grey symbols indicate the data for a reference bandwidth of 3200 Hz. Open symbols indicate the results for the 400-Hz wide reference. The data for the 10-ms (triangles) and 1000-ms signals (squares) are vertically shifted against each other by 10 dB. The scale on the left indicates the levels for the 1000-ms signal and the scale on the right those for the 10-ms noise bursts.

The reference bandwidth changed the shape of the curves for 10-ms signal for one subject (S12). Most of the subjects showed, however, similar results for the two reference bandwidths. Some of the subjects (e.g., S2 for the 10-ms signal and S3 for the 1000-ms signal) showed data for the two references that were vertically shifted against each other by on average about 3–5 dB indicating that the 3200-Hz wide reference had a slightly different loudness than the 400-Hz wide reference. This mismatch seems to correlate with the variability in the data of the first experiment for the 3200-Hz wide test signal. The standard deviations for subjects S2 and S3 were small in the first experiment and the differences between the data for the two references were also small in the second experiment. Subject S1 showed a large standard deviation for the 3200-Hz wide 10-ms test signal in the first experiment and had the largest average differences between the 10-ms data for the two references in the second experiment. This comparison between individual variability and the individual shift between the data for the two reference bandwidths suggests that the shift resulted from the inaccurate estimate of the level for the 3200-Hz reference from the first experiment, rather than reflecting a bias in the measurement procedure. The small average difference between the data (2 dB) for the two references (bottom right panel of Fig. 4) indicates that the choice of reference bandwidth had, on average, a negligible effect on the results.

These effects were supported by a four-way ANOVA for repeated measures (Table II). The bandwidth had a highly significant effect on the level difference between the test and the reference signal ($p<0.001$). Subject and duration were significant ($p<0.05$). The reference bandwidth showed no significant effect ($p>0.05$). This result supports the hypothesis that the data are not influenced by the fact that only the test stimulus is varied and the reference level is constant (see also Verhey, 1999). The significant interactions of bandwidth and subject ($p<0.001$) indicate individual difference in the amount of spectral loudness summation. The significant interactions of bandwidth and duration and bandwidth, duration and subject ($p<0.001$) indicate that spectral loudness summation depended on the duration and that this interaction in turn depended on the subject. The significant interaction of duration, subject, and reference bandwidth ($p<0.001$) indicates that for some subjects the reference bandwidth had an effect on the results for one of the two durations.

TABLE II. Four-way analysis of variance for repeated measures of loudness matching for normal-hearing subjects. The dependent variable is the level difference between test and reference signal at equal loudness. Bandwidth of the test signal (bw; six steps: 200, 400, 800, 1600, 3200, and 6400 Hz) and duration (dur; two steps: 10 and 1000 ms) and the reference bandwidth (ref; two steps: 400 and 3200 Hz) are fixed factors. Subject (sub; 11 steps: S1, S2, S3, S4, S6, S7, S8, S9, S10, S11, S12) is a random factor.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob$>F$
sub	7 113.7348	10	711.3735	2.4351	0.0469
dur	1 531.9473	1	1 531.9473	9.0028	0.0133
ref	232.917	1	232.917	2.9165	0.1185
bw	54 297.5445	5	10 859.5089	73.37	≤0.0001
sub × dur	1 701.6326	10	170.1633	1.7286	0.1642
sub × ref	798.6212	10	79.8621	1.011	0.4941
sub × bw	7 400.5076	50	148.0102	5.6266	≤0.0001
dur × ref	104.364	1	104.364	1.3111	0.2789
dur × bw	913.1152	5	182.623	6.7854	≤0.0001
ref × bw	32.6304	5	6.5261	0.87447	0.5050
sub × dur × ref	795.9937	10	79.5994	9.8618	≤0.0001
sub × dur × bw	1 345.7008	50	26.914	3.3345	≤0.0001
sub × ref × bw	373.1439	50	7.4629	0.9246	0.6086
dur × ref × bw	96.7137	5	19.3427	2.3964	0.0503
sub × dur × R × bw	403.5745	50	8.0715	1.1787	0.1948

IV. DISCUSSION

A. Role of reference bandwidth

The results of the second experiment showed that approximately the same level-bandwidth relation of equally loud signals was found for the narrow-band and the broadband references. The data seem to be inconsistent with the hypothesis by Anweiler and Verhey (2006) and Fruhmann *et al.* (2003) that loudness matching data might be slightly biased by the choice of the reference. Fruhmann *et al.* (2003) used the same loudness matching procedure as Verhey and Kollmeier (2002) with a smaller reference bandwidth (400 Hz). Their average results of eight normal-hearing subjects were similar to those of Verhey and Kollmeier (2002) for a reference bandwidth of 3200 Hz and a reference level of 65 dB. For the largest bandwidth (6400 Hz), however, they found a smaller magnitude of spectral loudness summation. Fruhmann *et al.* (2003) argued that this discrepancy might indicate that bias effects influenced the results of Verhey and Kollmeier (2002) at the largest bandwidth. Anweiler and Verhey (2006) found a similar discrepancy between their loudness matching data with a 65-dB 3200-Hz wide reference and their spectral loudness summation data derived from the loudness functions. In the present study the level difference for the 3200-Hz wide reference was much smaller than in the previous studies.

The difference between the studies may be due to the different group of subjects that participated in the studies. Verhey and Kollmeier (2002) as well as the present study showed large individual differences in the amount of spectral loudness summation. The level of the 3200-Hz wide 1000-ms reference of the second experiment was on average 58 dB SPL. Thus, the data are not directly comparable to the data in Anweiler and Verhey (2006) for the reference level of 65 dB. The difference between the loudness matching and the loudness scaling data for 3200 and 6400 Hz was smaller for the reference level of 45 dB (see also Verhey and Kollmeier, 2002). It is, thus, possible that a similar effect of reference bandwidth would have been observed with the group of subjects participating in the present study if a 10 dB higher reference level had been used.

It should be noted that, although the magnitude of spectral loudness summation differed between the present study and Anweiler and Verhey (2006), Fruhmann *et al.* (2003), and Verhey and Kollmeier (2002), all studies showed a similar increase in spectral loudness summation as the signal duration was decreased from 1000 to 10 ms. The level difference ΔL between equally loud 200- and 6400-Hz wide noise bursts was about 8 dB larger for the short signals than for the long signals (compared to 4–8 dB in the previous studies).

B. Role of starting level

Fruhmann *et al.* (2003) measured spectral loudness summation using the same reference bandwidth (400 Hz) and level (70 dB) as in the present study. For a signal duration of 1000 ms, they obtained a level difference of about 10 dB for equally loud 6400-Hz wide test signal and the reference signal. This is considerably smaller than the 19 dB found in the present study.

One procedural difference between Fruhmann *et al.* (2003) and the present study is the choice of the starting level. Fruhmann *et al.* (2003) used starting levels that corresponded to a loudness of reference signals of 25 or 35 on a categorical loudness scale from 0 (inaudible) to 50 (too loud). The equivalent level range of starting levels can be derived from Fig. 1 of Anweiler and Verhey (2006), assuming that the loudness functions for bandwidths 200 and 400 Hz are very similar. The level range depends on the subjects and ranges from less than 5 to about 20 dB. Thus, the average level range of the starting levels may have been smaller than in the present study (20 dB). Several studies on loudness recalibration and induced loudness reduction have shown that loud sounds preceding softer sounds may reduce the loudness of the softer sounds (e.g., Marks, 1994; Mapes-Riordan and Yost, 1999; Wagner and Scharf, 2006). Loudness recalibration would predict an increase in loudness difference between the reference and the test signal for a starting level considerably higher than the reference level. This effect would be smaller in Fruhmann *et al.* (2003), where it is likely that a smaller maximum starting level was used than in the present study. However, the final estimate for the level at equal loudness is derived for three different starting levels, of which only one was higher than the reference level. This averaging will considerably reduce the effect of loudness recalibration on the data. In addition, Verhey (1999) showed no systematic effect of starting level on the level of the test signal at equal loudness indicating a negligible effect of loudness recalibration on the results. Thus, it is more likely that the difference between the data in Fruhmann *et al.* (2003) and the present study results from a different set of subjects (see earlier text) rather than from a different range of starting levels.

C. Role of repetition rate

The results of the present study indicate that spectral loudness summation is similar for single noise bursts and sequences of noise bursts with high repetition rates of up to 50 Hz. For broadband signals there is no difference between the loudness of a sound with a 50-Hz repetition rate and a continuous sound (Port, 1963b). A repetition rate of 50 Hz is higher than the lower limit of pitch of about 30 Hz (Krumbholz *et al.*, 2000). It is also higher than the repetition rate of about 15 Hz, beyond which roughness starts to increase (Zwicker and Fastl, 1999). In order to elicit the sensation of roughness and pitch the noise bursts have to be perceived as part of one auditory object suggesting that, at these high rates, the noise bursts are no longer perceived as single bursts. This hypothesis is further supported by experiments on modulation-phase sensitivity (Dau, 1996). Dau (1996) found that subjects are unable to distinguish between different starting phases of a modulation for modulation frequencies of 12.5 Hz or higher. The data for repeated noise bursts of the present study indicate that, although they are probably no longer perceived as single noise bursts, their spectrum is still processed similarly to a single burst.

D. Role of compression

Verhey and Kollmeier (2002) proposed a duration-dependent compression in the spectral analysis stage of loudness models to account for the difference in the magnitude of spectral loudness summation for short and long signals. A model using a higher compression at the beginning of the signal predicts a larger amount of spectral loudness summation for short signals as found in the present data (Fig. 1) and in previous studies (Verhey and Kollmeier, 2002; Fruhmann *et al.*, 2003; Anweiler and Verhey, 2006).

A duration-dependent compression would imply different slopes of the loudness functions for stimuli with different durations. This is, however, not observed in recent data on the relation between loudness and level for short and long signals (Epstein and Florentine, 2005, 2006; Anweiler and Verhey, 2006): Using cross-modality matching and absolute loudness estimation, Epstein and Florentine (2005, 2006) found that the loudness ratio between the 5 and 200-ms 1-kHz tones was approximately constant, i.e., the two loudness functions had the same slope at the same level. In qualitative agreement with Epstein and Florentine (2005, 2006), Anweiler and Verhey (2006) found a constant vertical distance between loudness functions of bandpass noises with the same bandwidth and different durations (10 and 1000 ms). Thus, it is unlikely that compression changes with duration as was proposed in Verhey and Kollmeier (2002) to account for the duration effect in spectral loudness summation.

E. Role of frequency selectivity

Since the amount of spectral loudness summation is not only determined by the compression but also by the filter characteristics it is possible to account for the duration effect in spectral loudness summation by assuming a time-varying frequency selectivity. To predict a larger magnitude of spectral loudness summation for short signals than for long signals, a sharper filter at the beginning of the signal and a gradual increase of the filter width with increasing duration of the signal has to be assumed. Recent data on spectral selectivity in an overshoot paradigm seem to support this hypothesis of narrower filters at the beginning of a stimulus (Strickland, 2001).

To account for the data with repeated noise bursts of the present study, the change in frequency selectivity would have to be considerably faster than the time constant of 100–200 ms commonly associated with temporal integration of loudness (Munson, 1947; Stevens, 1961; Zwicker, 1966; Zwislocki, 1969) in order to account for the present data. Poulsen (1981) assumed a combination of a short (around 5–10 ms) and a long time constant (around 100–200 ms) for temporal integration of loudness. A change of the frequency selectivity with a time constant comparable to the shorter time-constant of Poulsen (1981) would be in line with the results of the present study.

F. Role of temporal integration

On the basis of their results, Anweiler and Verhey (2006) proposed a modified ELRH with a bandwidth-dependent loudness ratio between short and long signals. They showed that such a bandwidth-dependent loudness ratio could be implemented in loudness models by a duration-dependent spectral stage of the model.[4]

A possible implementation of such a bandwidth-dependent loudness ratio is a model using a bandwidth-dependent temporal integration stage. There are several studies on temporal aspects of loudness perception of repeated noise bursts that indicate such an influence of the spectrum on temporal integration (Port, 1963b; Grimm *et al.*, 2002; van Beurden and Dreschler, 2005). In agreement with the finding of the present study of a larger spectral loudness summation for repeated noise bursts than for continuous signals, these studies found a smaller temporal integration for broadband than for narrow-band signals. Different burst durations, bandwidths, and procedures are presumably the reason for the smaller influence of the spectrum on temporal integration in loudness in Grimm *et al.* (2002) and van Beurden and Dreschler (2005) than would have been expected on the basis of the spectral-loudness-summation data of the present study.

A bandwidth-dependent time window for spectral integration was proposed by van den Brink and Houtgast (1990) on the basis of their data on spectrotemporal integration in signal detection. Van den Brink and Houtgast (1990) assumed that loudness for narrow-band signals is derived from the integration over a longer duration than for broadband signals. The stimulus with the largest bandwidth in the present study is about three octaves wide. For this bandwidth (in octaves), van den Brink and Houtgast (1990) found a time window with a duration of about 30 ms. For narrow-band signals, they observed temporal integration over at least 100 ms. This change in temporal integration reduces the amount of spectral loudness summation for the 1000-ms signals whereas it does not affect the results for the 10-ms

single noise bursts because 10 ms is considerably shorter than the duration of the time window for the broad signal. For repeated noise bursts, spectral loudness summation should decrease as soon as more than one burst falls within the 100-ms time window for the narrow-band signal, i.e., for repetition rates exceeding 10 Hz. This prediction is at odds with the experimental data. Thus, a bandwidth-dependent temporal integration alone with the time constants found in van den Brink and Houtgast (1990) cannot account for the results of the present study.

V. SUMMARY AND CONCLUSIONS

The magnitude of spectral loudness summation was measured with a loudness matching procedure for sequences of 10-ms noise bursts as well as single noise bursts with a duration of 10 and 1000 ms. In agreement with previous data, spectral loudness summation was found to be on average about 8 dB larger for the single short burst than for the long burst. The results were unaffected by the choice of the reference. The magnitude of spectral loudness summation for sequences of noise bursts with repetition rates of up to 50 Hz was similar to that for 10-ms single noise bursts. Spectral loudness summation for a sequence of noise bursts of 70 Hz was smaller than for the lower repetition rates but still significantly different from that for a 1000-ms signal. The results show that, for a 70-dB narrow-band reference signal, spectral loudness summation is dominated by the duration of single noise bursts even when the noise bursts are no longer perceived as single events. It remains to be seen if the effect is similar for higher and lower reference levels than that used in the present study.

ACKNOWLEDGMENTS

We would like to thank the editor and the two reviewers for their many helpful comments on a previous version of the manuscript. This work was partly supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich SFB/TRR 31).

[1]Since low-noise noise was only generated with a single iteration it is perceptually very similar to Gaussian noise. In addition, the processing of the auditory periphery changes phase and amplitude characteristics of the noise, i.e., for the bandwidths used in the present study, low-noise noise is no longer low-noise noise at the output of the auditory filters (Kohlrausch *et al.*, 1997).

[2]The level refers to the level of the 45 500-sample buffer of a bandlimited noise prior to the multiplication with the desired temporal envelope. Thus, the overall level of the sequence of noise bursts—calculated as the root mean square over the entire signal interval including the interburst intervals—decreases as the repetition rate decreases. The level of each noise burst is, on average, equal to 70 dB SPL. This choice of reference level facilitates the comparison with the results to data on spectral loudness summation of single noise bursts with different durations (Verhey and Kollmeier, 2002; Anweiler and Verhey, 2006). Verhey and Kollmeier (2002) motivated a comparison at the same reference level with predictions of the equal-loudness-ratio hypothesis [(ELRH), Florentine *et al.*, 1996; Buus *et al.*, 1997]: for the same level, the ELRH predicts the same compression for signals with the same spectrum and different durations and hence, if the loudness ratio is independent of the spectral content as assumed in the original version of the ELRH, the same spectral loudness summation. The current choice of reference level does not allow one to test other hypotheses, e.g., that the compression for narrow-band signals and thus the magnitude of spectral loudness summation depends on the overall loudness of the reference.

[3]Several other studies also used a number of bursts per interval which was equal to the repetition rate (Garner, 1948; Zwicker, 1966; van Beurden and Dreschler, 2005). In contrast to the present study, they used 1000-ms intervals, irrespective of the repetition rate, whereas in the present study the last burst determined the duration of the interval. The choice of the interval duration of the present study ensures a perceptually well defined end of each interval. Constant interval durations, as used in the other studies, usually result in a period of silence after the last burst of an interval which is indistinguishable from the silent interval separating the signal intervals.

[4]Figure 4 in Anweiler and Verhey (2006) shows level difference between 200- and 3200-Hz wide equally loud noise bursts together with predictions assuming a bandwidth-dependent loudness ratio between short and long signals. Although a duration effect in spectral loudness summation is predicted, there are some aspects of the data that are not accounted for by the assumption of a bandwidth-dependent loudness ratio. For example, the predicted effect of duration is largest where the spectral loudness summation for the long signal is largest and decreases toward higher and lower levels. In contrast, the experimental data show approximately the same duration effect for medium to high reference levels, although the spectral loudness summation for the long signals decreases toward higher levels. This suggests that a bandwidth-dependent equal-loudness ratio is probably not sufficient to account for all aspects of the effect of duration on spectral loudness summation.

Anweiler, A. K. (**2005**). "Duration dependence of spectral loudness summation in normal-hearing and hearing-impaired listeners," Master's thesis, Universität Oldenburg, Oldenburg, Germany.

Anweiler, A. K., and Verhey, J. L. (**2006**). "Spectral loudness summation for short and long signals as a function of level," J. Acoust. Soc. Am. **119**, 2919–2928.

Buus, S., Florentine, M., and Poulsen, T. (**1997**). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am. **101**, 669–680.

Buus, S., Florentine, M., and Poulsen, T. (**1999**). "Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin," J. Acoust. Soc. Am. **105**, 3464–3480.

Cacace, A. T., and Margolis, R. H. (**1985**). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am. **78**, 1568–1573.

Dau, T. (**1996**). *Modeling Auditory Processing of Amplitude Modulation* (BIS, Oldenburg), an online version is available at the Internet site ⟨http://www.bis.uni-oldenburg.de/bisverlag/daumod96/daumod96.html⟩ last viewed in August 2007.

Epstein, M., and Florentine, M. (**2005**). "A test of the equal-loudness-ratio hypothesis using cross-modality matching functions," J. Acoust. Soc. Am. **118**, 907–913.

Epstein, M., and Florentine, M. (**2006**). "Loudness of brief tones measured by magnitude estimation and loudness matching (L)," J. Acoust. Soc. Am. **119**, 1943–1945.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, H., and Steinberg, J. (**1924**). "The dependance of the loudness of a complex sound upon the energy in the various frequency regions of the sound," Phys. Rev. **24**, 306–317.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.

Fruhmann, M., Chaluppper, J., and Fastl, H. (**2003**). "Zum Einfluss von Innenohrschwerhörigkeit auf die Lautheitssummation" ("Influence of hearing impairment on spectral loudness summation)," Fortschritte der Akustik - DAGA 2003, Oldenburg, Deutsche Gesellschaft für Akustik e.V., pp. 253–254.

Garner, W. R. (**1948**). "The loudness of repeated tones," J. Acoust. Soc. Am. **20**, 513–527.

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.

Grimm, G., Hohmann, V., and Verhey, J. L. (**2002**). "Loudness of fluctuating sounds," Acust. Acta Acust. **88**, 359–368.

Hübner, R., and Ellermeier, W. (**1993**). "Additivity of loudness across critical bands: A critical test," Percept. Psychophys. **54**, 185–189.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-

noise noise: Further evidence of the role of envelope fluctuations," Acust. Acta Acust. **83**, 659–669.
Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1070–1080.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.
Marks, L. E. (**1994**). "Recalibrating the auditory system: The perception of loudness," J. Exp. Psychol. **20**, 382–396.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model of the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–239.
Munson, W. (**1947**). "The growth of auditory sensation," J. Acoust. Soc. Am. **19**, 584–591.
Nelken, I., Rotman, Y., and Yosef, O. B. (**1999**). "Responses of auditory-cortex neurons to structural features of natural sounds," Nature (London) **397**, 154–157.
Niese, H. (**1959**). "Die Trägheit der Lautstärke in Abhängigkeit vom Schallpegel" ("Sluggishness of loudness as a function of level"), Hochfrequenzt und Elektroakustik **68**, 143–152.
Ogura, Y., Suzuki, Y., and Sone, T. (**1991**). "A temporal integration model for loudness perception of repeated impulsive sounds," J. Acoust. Soc. Jpn. **12**, 1–11.
Port, E. (**1963a**). "Über die Lautstärke einzelner kurzer Schallimpulse" ("On the loudness of single sound bursts"), Acustica **13**, 212–223.
Port, E. (**1963b**)."Zur Lautstärkeempfindung und Lautstärkemessung von pulsierenden Geräuschen" ("On the sensation of loudness and the measurement of loudness of repeated sound bursts"), Acustica **13**, 224–233.
Poulsen, T. (**1981**). "Loudness of tones in a free field," J. Acoust. Soc. Am. **69**, 1786–1790.
Reichardt, W. (**1965**). "Zur Trägheit der Lautstärkebildung" ("Sluggishness of loudness perception"), Acustica **15**, 345–354.
Reichardt, W. (**1970**). "Subjective and objective measurement of loudness level of single and repeated impulses," J. Acoust. Soc. Am. **47**, 1557–1562.
Reichardt, W., and Niese, H. (**1970**). "Choice of sound duration and silent intervals for test and comparison signals in the subjective measurement of loudness level," J. Acoust. Soc. Am. **47**, 1083–1090.
Scharf, B. (**1959**). "Loudness of complex sounds as a function of the number of components," J. Acoust. Soc. Am. **31**, 783–785.
Scharf, B. (**1962**). "Loudness summation and spectrum shape," J. Acoust. Soc. Am. **34**, 228–233.
Schneider, B. (**1988**). "The additivity of loudness across critical bands: A conjoint measurement procedure," Percept. Psychophys. **43**, 211–222.
Stevens, S. S. (**1961**). "Procedure for calculating loudness: Mark VI," J. Acoust. Soc. Am. **33**, 1577–1585.
Strickland, E. A. (**2001**). "The relationship between frequency selectivity and overshoot," J. Acoust. Soc. Am. **109**, 2062–2073.
van Beurden, M. F. B., and Dreschler, W. A. (**2005**). "Bandwidth dependency of loudness in series of short noise bursts," Acust. Acta Acust. **91**, 1020–1024.
van den Brink, W. A. C., and Houtgast, T. (**1990**). "Spectro-temporal integration in signal detection," J. Acoust. Soc. Am. **88**, 1703–1711.
Verhey, J. L. (**1999**). *Psychoacoustic of Spectro-Temporal Effects in Masking and Loudness Perception* (BIS, Oldenburg), an online version is available at the Internet site ⟨http://www.bis.uni-oldenburg.de/bisverlag/verpsy99/verpsy99.html⟩ last viewed in August 2007.
Verhey, J. L., and Kollmeier, B. (**2002**). "Spectral loudness summation as a function of duration," J. Acoust. Soc. Am. **111**, 1349–1358.
Wagner, E., and Scharf, B. (**2006**). "Induced loudness reduction as a function of exposure time and signal frequency," J. Acoust. Soc. Am. **119**, 1012–1020.
Zwicker, E. (**1966**). "Ein Beitrag zur Lautstärkemessung impulshaltiger Schalle" ("A contribution to the loudness of impulsive sounds"), Acustica **17**, 11–22.
Zwicker, E. (**1969**). "Der Einfluß der zeitlichen Struktur von Tönen auf die Addition von Teillautheiten" ("Influence of the temporal structure of tones on the summation of partial loudnesses"), Acustica **21**, 16–25.
Zwicker, E. (**1977**). "Procedure of calculating loudness of temporally variable sounds," J. Acoust. Soc. Am. **62**, 675–682.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics*, Springer Series in Information Sciences, 2nd ed. (Springer, Berlin).
Zwicker, E., and Feldtkeller, R. (**1955**). "Über die Lautstärke von gleichförmigen Geräuschen" ("On the loudness of uniform sounds"), Acustica **5**, 303–316.
Zwicker, E., Flottorp, G., and Stevens, S. (**1957**). "Critical bandwidth and loudness summation," J. Acoust. Soc. Am. **29**, 548–557.
Zwicker, E., and Scharf, B. (**1965**). "A model of loudness summation," Psychol. Rev. **72**, 3–26.
Zwislocki, J. J. (**1969**). "Temporal summation of loudness: An analysis," J. Acoust. Soc. Am. **46**, 431–441.

The pulse-train auditory aftereffect and the perception of rapid amplitude modulations

Alexander Gutschalk[a)]
Department of Neurology, University of Heidelberg, 69120 Heidelberg, Germany

Christophe Micheyl and Andrew J. Oxenham
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 27 February 2007; revised 23 November 2007; accepted 29 November 2007)

Prolonged listening to a pulse train with repetition rates around 100 Hz induces a striking aftereffect, whereby subsequently presented sounds are heard with an unusually "metallic" timbre [Rosenblith *et al.*, Science **106**, 333–335 (1947)]. The mechanisms responsible for this auditory aftereffect are currently unknown. Whether the aftereffect is related to an alteration of the perception of temporal envelope fluctuations was evaluated. Detection thresholds for sinusoidal amplitude modulation (AM) imposed onto noise-burst carriers were measured for different AM frequencies (50–500 Hz), following the continuous presentation of a periodic pulse train, a temporally jittered pulse train, or an unmodulated noise. AM detection thresholds for AM frequencies of 100 Hz and above were significantly elevated compared to thresholds in quiet, following the presentation of the pulse-train inducers, and both induced a subjective auditory aftereffect. Unmodulated noise, which produced no audible aftereffect, left AM detection thresholds unchanged. Additional experiments revealed that, like the Rosenblith *et al.* aftereffect, the effect on AM thresholds does not transfer across ears, is not eliminated by protracted training, and can last several tens of seconds. The results suggest that the Rosenblith *et al.* aftereffect is related to a temporary alteration in the perception of fast temporal envelope fluctuations in sounds.
 [DOI: 10.1121/1.2828057]

PACS number(s): 43.66.Dc, 43.66.Lj [JHG] Pages: 935–945

I. INTRODUCTION

Prolonged exposure to a constant or repeating stimulus can induce a transient alteration in the perception of subsequent stimuli, which is commonly referred to as an "aftereffect" in the psychophysical literature. In some cases from the visual domain, the aftereffect manifests itself as an "afterimage." Perhaps the simplest and most familiar visual aftereffect is experienced when closing one's eyes after staring at a bright light. Another famous example is the "waterfall" illusion in which, after watching a waterfall for several seconds, one sees fixed objects (such as nearby rocks) "as if in upward motion" (Addams, 1834). The waterfall illusion is just one example of a large set of motion aftereffects, wherein exposure to movement in a certain direction causes a following stationary stimulus to be perceived as moving in the opposite direction [see Wade (1994) for a historical review of motion aftereffects]. Other types of aftereffects are produced using stationary stimuli. These include tilt aftereffects, wherein an oriented stimulus appears to be rotated away from the orientation of a prior stimulus (Gibson and Radner, 1937; Mitchell and Muir, 1976; Magnussen and Johnsen, 1986; He and McLeod, 2001). In fact, aftereffects have been identified, not just for motion and orientation, but for almost all features of visual perception, including spatial frequency, contrast, color, stereoscopic depth, size, and others (e.g., MacKay, 1964; Blakemore and Sutton, 1969; Blakemore and Campbell, 1969; Blakemore and Julesz, 1971).

Common explanations for aftereffects are couched either in terms of sensory persistence, or in terms of neural adaptation. According to the latter type of explanation, prolonged exposure to a stimulus causes a reduction in the responsiveness of neurons that are specifically activated by certain features of that stimulus—a view supported by physiological findings (Barlow and Hill, 1963; Movshon and Lennie, 1979; Maffei *et al.*, 1973; Barlow, 1990). This selective "adaptation" biases subsequent responses of the corresponding array of feature detectors (which are systematically tuned to different values of the stimulus parameter) toward activation patterns shifted away from that previously evoked by the adapting stimulus (Clifford *et al.*, 2000). Alternative and more elaborate explanations involve release from inhibition (Sekuler and Pantle, 1967; Mather *et al.*, 1998), shifts in tuning (Jin *et al.*, 2005), or other neural mechanisms. Perceptual aftereffects provide a unique psychophysical tool as evidence for the existence of specific feature detectors in a sensory system.

While the visual psychophysics literature abounds with examples of aftereffects, examples of analogous phenomena in the auditory modality are much less common. Compared to their visual counterparts, auditory aftereffects remain rather elusive and often require very specific and somewhat unnatural test conditions in order to reveal themselves (e.g., Shu *et al.*, 1993). One example is the so-called "Zwicker

[a)]Author to whom correspondence should be addressed. Electronic mail: alexander_gutschalk@med.uni-heidelberg.de

tone," in which an illusory tonal sensation is heard for a few seconds following the presentation of a broadband noise containing a spectral notch about one-third octave in width, with relatively sharp edges (Zwicker, 1964; Lummis and Guttman, 1972; Wiegrebe *et al.*, 1996; Norena *et al.*, 2000, 2002). The Zwicker tone has been described as a "negative auditory afterimage," because the pitch of the transient illusory tone corresponds roughly to the center frequency of the spectral notch in the preceding noise. While the mechanisms responsible for the generation of the Zwicker tone are still not entirely clear, neurophysiologically inspired models (e.g., Norena *et al.*, 2000; Franosch *et al.*, 2003) have been offered, and potential neural correlates for the phenomenon have been identified at the level of the auditory cortex (Hoke *et al.*, 1996; Norena and Eggermont, 2003). It appears that the aftereffect is related to a temporary enhancement of responsiveness, possibly related to a release from inhibition, in central auditory neurons with best frequencies within the spectral notch, which were least stimulated during the presentation of the inducer.

Another example of an auditory aftereffect is when a stimulus with a uniform (or "flat") spectrum acquires a timbre that is related to the complement (or "negative") spectrum of a preceding stimulus. A compelling demonstration of this type of aftereffect was provided by Summerfield *et al.* (1984). By removing from a series of equal-amplitude harmonics three harmonics, the frequencies of which were close to those of the first three formants in a vowel, Summerfield *et al.* generated precursors that resembled spectral complements of the vowel. When listeners were presented with such precursors followed by the whole series of equal-amplitude harmonics, they heard the latter as an identifiable vowel, despite its physically flat spectrum. A simpler demonstration of a potentially related effect involves removing just one component in a complex tone; when that component is later reintroduced, it stands out perceptually (Wilson, 1970; Viemeister, 1980; Viemeister and Bacon, 1982). The mechanisms underlying such "auditory enhancement" effects remain unclear. The most common explanation involves adaptation, such that the neural responses to subsequently presented stimulus components that were not part of the precursor are enhanced relative to the (adapted) responses to components that were in the precursor.

Another interesting auditory aftereffect appears to have been forgotten soon after its initial description by Rosenblith *et al.* (1947). These authors discovered that, after listening for 1 to 2 min to a train of rectangular pulses repeating at a relatively fast rate (e.g., 100 Hz), listeners experienced various environmental sounds, such as their own voice, a typewriter, a handclap, or the sound of rubbing sandpaper as having an unusually "metallic" timbre—also described as an added "jangly," "twangy," or "like a rasping file" quality. Rosenblith *et al.* explored the influence of various stimulus parameters, including the pulse rate, duration, and level of the inducer. They found that the strength of the aftereffect increased with inducer level and that inducer pulse rates between 30 and 200 Hz were most effective. Using inducer durations ranging from 5 to 240 s, they showed that the duration of the aftereffect increased as a function of exposure time. Based on these results, they suggested exposure times of 20–30 s as a "convenient compromise between the listener's impatience and the experimenter's desire to produce a measurable effect."

Initially unaware of the Rosenblith *et al.* aftereffect, we were recently led to rediscover it during a series of magnetoencephalography experiments (Gutschalk *et al.*, 2007), which involved the continuous presentation of pulses repeating at 80 Hz over tens of minutes. At the end of such experiments, many listeners spontaneously reported experiencing a noticeable change in the perceived quality of sounds, which invariably subsided within a few minutes. Listening to these sounds, we also noted a second feature, which was not reported by Rosenblith *et al.* (1947): When the pulse train is played for longer than 1 min, it appears to serve as its own test stimulus, and prominently changes its sound character. At its beginning, the pulse train sounds like one coherent source, with a buzzing pitch and a certain roughness. After about 20–30 s, however, the roughness appears to slowly die away, while a buzzing, which was previously only a minor element of the coherent percept, becomes increasingly prominent and segregated from the first. In informal listening, the latter percept has been likened to the sound of cicada or midges. Similar to the afterimage, this phenomenon is more prominent when the pulse train is high-pass filtered above 2000–4000 Hz, while low-pass filtering attenuates it, such that the effect is completely abolished for low-pass cutoff frequencies below about 2000 Hz.

Unlike the Zwicker tone, the Rosenblith *et al.* (1947) auditory aftereffect has been the object of very little research up to now, and the perceptual and neural mechanisms underlying it remain essentially unknown. Based on the observations of Rosenblith *et al.* (confirmed by informal listening experiments on ourselves) that the aftereffect was largest for inducer pulse rates between 30 and 200 Hz, and could be eliminated or greatly reduced by low-pass filtering, we hypothesized that the effect was (a) dependent upon the presence of relatively fast and marked temporal envelope fluctuations in the outputs of peripheral auditory filters stimulated by the inducer, and (b) related to an alteration of the perception of such fast temporal envelope fluctuations in subsequently presented sounds. In order to test this hypothesis, we measured how amplitude modulation (AM) detection thresholds for probe noise bursts were influenced by the prior presentation of three types of inducers: (1) A high-pass-filtered 100-Hz harmonic complex with components in sine phase, the temporal waveform and spectrum of which are similar to those of a pulse train with a corresponding rate, (2) a 100-Hz "jittered" pulse train, wherein the timing of each pulse was randomly shifted forward or backward relative to its nominal position, resulting in a stimulus that was aperiodic, but still had marked temporal envelope fluctuations, and (3) an unmodulated noise inducer, which was expected to produce no aftereffect, and so served as a control.

In addition to a main experiment, in which we compared the influence of these three types of inducers on thresholds for the detection of AM imposed onto short noise-burst carriers for different AM frequencies, we performed three further experiments. Experiment 2 measured the time course of

the threshold recovery. Experiment 3 was sparked by the informal observation of Rosenblith *et al.* (1947) that the aftereffect reported in their study was not elicited when the test stimulus was not presented to the same ear as the inducer; this prompted us to test whether the aftereffect on AM detection thresholds in the present study would also not be present under such listening conditions. Experiment 4 was motivated by a recent report that AM threshold adaptation effects can disappear with protracted task practice (Bruckert *et al.*, 2006); in order to test whether the effects observed in the present study were also susceptible to training, some of the listeners who had already taken part in the previous three experiments were tested further.

II. GENERAL METHODS

A. Listeners

Four listeners (one female, three male; age 24–32) participated in all experiments, except that listener 1 did not participate in experiment 4. They had normal hearing, defined here as pure-tone thresholds below 20 dB HL at octave frequencies between 250 and 8000 Hz, and reported no history of peripheral or central hearing disorders. The listeners were tested individually in a double-walled sound-attenuating chamber. They were paid an hourly wage for their participation. The study protocol was approved by the institutional review board of the Massachusetts Institute of Technology.

B. Apparatus

The stimuli were generated digitally under MATLAB (The MathWorks, MA), stored onto the computer hard disk, and played out at a 48-kHz sampling rate using the 24-bit digital-to-analog converter of a LynxStudio LynxOne soundcard. They were delivered diotically (except in experiment 3) to the listener via HD580 circumaural headphones (Sennheiser, Old Lyme, CT). The overall sound intensity for inducers as well as probes was set to 46 dB SPL. This moderate level was chosen based on preliminary listening tests, indicating that this level was sufficient to induce a strong aftereffect while still falling well below levels that could cause some listeners discomfort during protracted listening.

III. EXPERIMENT 1: INFLUENCE OF INDUCER TYPE AND AM FREQUENCY TUNING

A. Stimuli and procedure

In this experiment, modulation detection thresholds for probe noise bursts were measured in the absence and in the presence of an "inducer" or "adaptor." Three different inducers were tested in separate conditions: (1) A 100-Hz F0 harmonic complex with all harmonics starting in sine (i.e., 0°) phase, which approximates a regular pulse train with a repetition rate of 100 Hz; (2) a train of temporally jittered pulses with a 100-Hz average rate, which was obtained by shifting randomly and independently the timing of each pulse in an original 100-Hz pulse train over a 10-ms range (i.e., from −5 to +5 ms around the pulse's nominal temporal position with uniform distribution); and (3) a Gaussian noise. The probe stimuli were three 150-ms noise bursts, including 20-ms on and off raised-cosine ramps. One of the three bursts, chosen at random with equal probability on each trial, was sinusoidally modulated in amplitude. Both the inducer and probe stimuli were bandpass filtered using a sixth-order, zero phase-shift Butterworth filter with 6-dB cutoff frequencies of 4 and 16 kHz.

An adaptive tracking procedure was used to estimate the detection threshold for the sinusoidal amplitude modulation. At the beginning of each run, the adaptor was played for 60 s. Two seconds before the end of the adaptor, listeners were visually alerted that the first test trial was about to begin. Following a 200-ms silent interval after the offset of the adaptor, the three probe bursts were presented, separated from each other by 200 ms. The amplitude of one of the three bursts was modulated with a modulation index (m) of −2 dB (i.e., the modulator was 2 dB below the level required for 100% sinusoidal amplitude modulation). The third probe burst was followed by a 200-ms silent interval, after which the inducer was resumed for 4 s and the next trial (i.e., series of probe bursts) began. This alternation of 4-s inducer and probe tones continued until the termination of the threshold-tracking procedure, which occurred after the tenth reversal in the direction of the changes in AM depth. The step size, by which the modulation index was changed, was initially set to 4 dB; it was reduced to 2 dB after the second reversal, and to 1 dB after the fourth reversal. The threshold was computed as the mean of the modulation index (in decibels) across the last six reversals. Listeners had a time window of 2 s after the end of the third probe burst to indicate which of the three probe bursts they thought was amplitude modulated, and they were instructed to respond before the end of that time period, as far as possible. In rare cases where they failed to respond within this time window, the AM depth was left unchanged for the next trial. Otherwise, the AM depth was reduced after any two consecutive correct responses, and increased after any incorrect response.

The following five AM frequencies for the probe bursts were tested in separate conditions: 50, 100, 150, 250, and 500 Hz. These five AM rate conditions, combined with the three inducer conditions and the baseline (no-inducer) condition, yielded a total of 20 different test conditions. The no-inducer condition was tested first. When data collection in this condition was completed, the inducer was introduced. Listeners were given the opportunity to practice the task with the inducer present before data collection for the different inducer conditions started. The different AM rate conditions with the inducer present were tested in randomized order, but the inducer was always the same within one session. Each listener performed a minimum of four runs in each condition; the thresholds measured on the last four runs were averaged.

B. Results

Figure 1 shows how the AM depth (AMD) at threshold, defined in terms of the modulation index (m), varied as a function of the probe-burst modulation rate in the different test conditions of experiment 1. The four panels of Fig. 1(a) show individual data; Fig. 1(b) shows the average across the

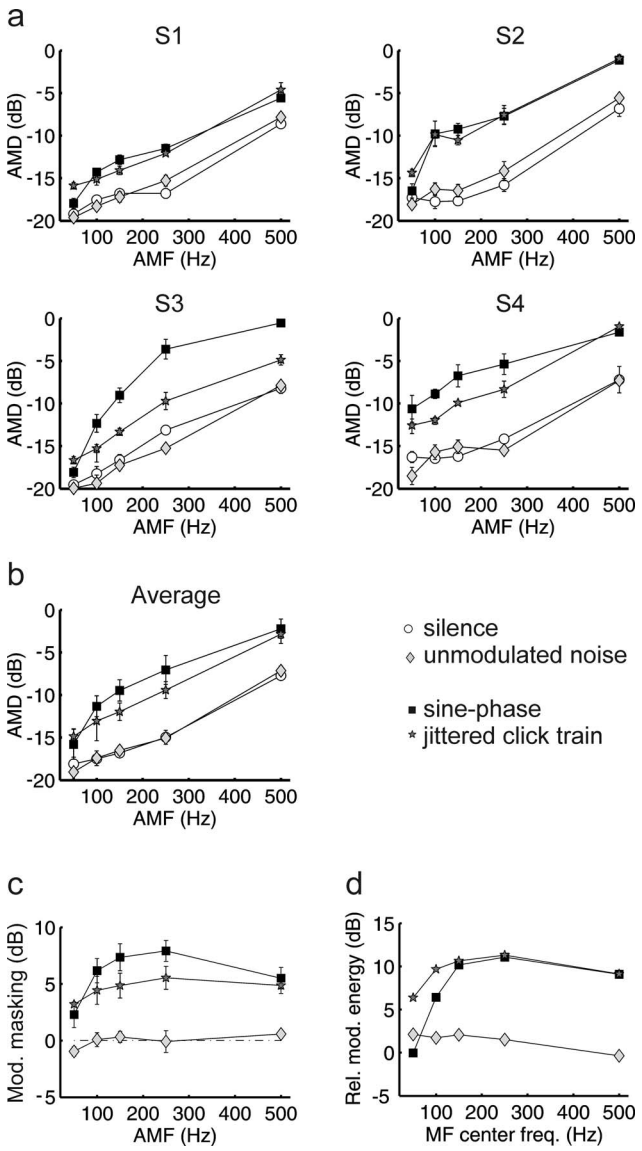


FIG. 1. (a) Individual thresholds for detecting sinusoidal AM imposed on a bandpass noise carrier in silence (open circles), in the presence of the two pulse-train inducers (sine-phase pulse train, closed squares; jittered pulse train, closed stars), and in the presence of unmodulated bandpass noise (closed diamonds) are plotted in terms of $20\log_{10}(m)$, where m is the modulation index, as a function of modulation frequency (AMF). Error bars represent ± 1 standard error of the mean. (b) Mean of the individual results; error bars represent ± 1 standard error across listeners. (c) Elevation of AM detection thresholds in the presence of the three inducers compared to silence (averaged across listeners $\pm$ standard error). The data represent the difference between the conditions shown in (b) and thresholds in silence. (d) Output of a modulation filterbank in response to the inducers of experiment 1, based on the filterbank parameters of Ewert and Dau (2000). Filtered modulation power is considerably higher for the periodic and jittered pulse trains than for bandpass noise at most frequencies, and shows a broad peak at modulation-filter center frequencies around 200 Hz, in general agreement with the psychophysical data.

four listeners. The threshold AMD is expressed in decibels, as $20\log_{10}(m)$, such that lower values correspond to better detection. In all conditions, AM detection thresholds increased with the AM frequency (AMF). The thresholds measured in the absence of any inducer (silence condition, open circles) and those measured following the unmodulated noise inducer (gray diamonds) were usually the lowest and were not significantly different from each other ($F_{1,3}=0.00$, $p=0.975$). Thresholds measured following the sine-phase complex tone inducer (closed squares) were usually the highest, and were significantly higher than those measured in quiet ($F_{1,3}=45.05$, $p=0.0068$). Thresholds measured following the jittered pulse train inducers were slightly, but not significantly different from the regular pulse train ($F_{1,3}=2.24$, $p=0.232$), but were significantly higher than those found in the quiet condition ($F_{1,3}=35.25$, $p=0.0095$).

Figure 1(c) illustrates the "tuning" of the inducer effect with respect to the AMF. The data shown in Fig. 1(c) were obtained by subtracting the thresholds measured in the silent condition from those measured in the different inducer conditions. Accordingly, higher values indicate larger increases in threshold caused by the inducer. The mean increase in AM depth at threshold across listeners for the sine-phase-complex inducer was 2.3 dB (range: 0.8–5.7 dB) at 50 Hz, 6.2 dB (3.3–8.0 dB) at 100 Hz, 7.4 dB (4.0–9.5 dB) at 150 Hz, 7.9 dB (5.2–9.5 dB) at 250 Hz, and 5.5 dB (3.0–7.7 dB) at 500 Hz. For the jittered-pulse-train inducer, the mean increase was 3.2 dB (range 2.9–3.7 dB) at 50 Hz, 4.4 dB (2.4–7.9 dB) at 100 Hz, 4.9 dB (2.7–7.1 dB) at 150 Hz, 5.5 dB (3.4–8.2 dB) at 250 Hz, and 4.9 dB (3.3–6.3 dB) at 500 Hz. As can be seen, for those inducers that had a significant effect (i.e., the complex tones and jittered pulse trains), the effect was broadly tuned, with a peak usually corresponding to around 250 Hz AMF—higher than the 100-Hz AMF of the probe stimulus. As indicated by significant interactions between the "AMF" and "condition" factors in two-way analyses of variance (ANOVA) on the data from the silent and inducer conditions, the influence of the periodic, but not the irregular pulse train, depended on the AMF of the probe [sine phase inducer: $F_{4,12}=10.70$, $p=0.0125$ (including a Greenhouse–Geisser, GG, correction for lack of sphericity where appropriate); irregular pulse train: $F_{4,12}=2.30$, $p=0.2012$]. Student's t-tests for the 50-Hz AMF were only significant for the jittered pulse train compared to silence ($t=16.35$, $p=0.0005$). For all other AMF conditions, the difference between inducers and the silent condition was significant in paired t-tests for both the periodic and jittered pulse trains. No significant interaction with AMF in the contrast of periodic pulse train and silence was observed when the 50-Hz-AMF condition was not included in the ANOVA, indicating that the observed tuning of the periodic pulse train inducer shows mainly a high-pass characteristic above the F0. However, the frequency interaction between periodic and aperiodic pulse trains missed significance in the two-way ANOVA when a GG sphericity correction was applied [$F_{4,12}=4.08$, $p=0.0259$ (uncorrected); $p=0.1146$].

The high-pass nature of the adaptation pattern is similar to what might be expected, given what is currently known about modulation masking and modulation frequency selectivity (e.g., Bacon and Grantham, 1989; Houtgast, 1989; Ewert and Dau, 2000). The modulation spectrum of the 100-Hz pulse train has components at multiples of 100 Hz. Given the broad tuning of the hypothesized modulation filters, with postulated Q values of 1 or less (e.g., Ewert and Dau, 2000), one would not expect to see distinct masking peaks at 100, 200, and 300 Hz, but rather a broadly tuned

response. This is illustrated in Fig. 1(d), which shows the time-averaged output of a modulation filterbank, generated using the parameters proposed by Ewert and Dau (2000), operating on the different inducer stimulus waveforms, after processing designed to simulate the effects of the peripheral auditory system. This included passing the stimuli through a gammachirp auditory filter (Irino and Patterson, 1997) with a characteristic frequency of 5 kHz (i.e., within the stimulus passband) and phase response modified according to Oxenham and Dau (2001a, b), and extracting the envelope through halfwave rectification. Additionally, a low-pass filter (first order) with a cutoff frequency of 150 Hz was applied, to account for the relative reduction of sensitivity at higher modulation frequencies (Kohlrausch *et al.*, 2000; Ewert and Dau, 2000). The center frequencies of the modulation filters (Ewert and Dau, 2000; Q value=1) were chosen to coincide with the AM frequencies of the probe tones. Finally, the energy at the output of these filters was computed. The resulting modulation-filter responses are broadly consistent with the psychophysical findings shown in Fig. 1(c) in that both the periodic and jittered pulse trains exhibit a broadly tuned bandpass shape peaking around 200–300 Hz, although some discrepancies between the model predictions and the data are apparent for the lowest probe modulation frequencies.

IV. EXPERIMENT 2: TIME COURSE OF THE EFFECT

A. Rationale, stimuli, and procedure

This experiment sought to measure the time course of the change in AM detection thresholds following the offset of a 60-s, 100-Hz sine-phase harmonic complex inducer, similar to that used in experiment 1. The bandwidths of the inducer and probe were the same as in experiment 1. Each run started with the presentation of this inducer, followed after a 200-ms silent interval by a first triplet of probe bursts. As in experiment 1, one of the three probe bursts (the target burst) was modulated in amplitude, and the listener's task was to indicate which one. Based on the results of experiment 1, which showed the largest effect at an AM frequency of 250 Hz, the target probe was modulated at a rate of 250 Hz here. Also based on the results of experiment 1, the following four modulation indices were selected: −6, −10, −14, and −18 dB. On each run, one of these four depths was randomly selected, and applied to all target bursts. Following a 3-s silent interval after the end of the first triplet, another triplet was presented, again with the position of the target probe randomized. This was done repeatedly over a period of 60 s, during which a total of 20 probe triplets was presented. Each run was followed by a 30-s silent interval. Following this fixed period of silence, listeners could initiate the next run with a button press. To establish a baseline, the same four modulation indices were also tested in silence at the beginning of the experiment, in the same random order.

Listeners had a time window of 2.5 s after the offset of the last stimulus in each triplet to respond. Rare trials on which listeners failed to respond within the time window were discarded. Each listener performed 50 runs at each of the five AM depths. For each of the 20 triplet positions, the estimated correct-response probabilities were plotted as a function of the AM depth and fitted with a logistic function using a maximum likelihood procedure. This permitted the estimation of the 70%-correct AM detection thresholds as a function of the delay between the offset of the 30-s inducer and the onset time of each probe triplet.

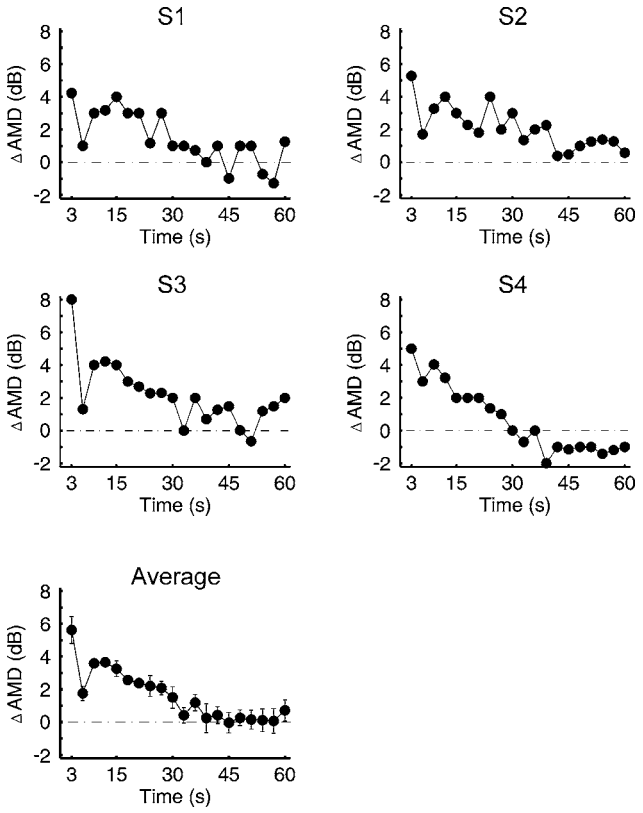


FIG. 2. Recovery of AM detection thresholds over time. Each data point represents the difference of the AM threshold determined after the inducer and the AM threshold in silence (ΔAMD), as determined at the beginning of experiment 2. The inducer was 60 s long, and adapted thresholds were determined in 3-s steps after the end of the inducer (a sine-phase pulse train). The AM depth of the probe was −6, −10, −14, or −18 dB. Thresholds corresponding to 70% correct on the psychometric function were calculated using a maximum likelihood procedure. The bottom panel shows the mean across listeners, and the error bars in that panel correspond to 1 standard error of the mean across listeners.

B. Results

The results of experiment 2 are shown in Fig. 2, where the upper four panels show individual data, and the lower panel shows the mean across listeners. The data points in these plots were obtained by subtracting the AM detection threshold in the absence of the inducer, as determined by the same procedure at the beginning of the experiment, from that measured in the presence of the inducer at each probe-triplet position, across all triplet positions. Thus, these data points reflect the influence of the sine-phase inducer on AM detection thresholds at different times after the offset of the inducer: Positive values along the Y axis indicate an increase in threshold compared to the reference (quiet) condition; the zero point is marked by a horizontal dashed-dotted line.[1]

As can be seen, the effect of the inducer usually decreased over the first 30 s following the offset of the inducer, after which it remained roughly constant on average, and not

significantly different from zero [main effect over the whole 1- to 60-s period: $F_{19,57}=12.03$, $p=0.0020$ (GG); main effect over the 1- to 30-s period: $F_{9,27}=7.05$, $p=0.0102$ (GG); linear contrast over the same period: $F_{1,3}=12.10$, $p=0.0401$; (third- and fifth-order contrasts, not reported here, were also significant); main effect over the 31- to 60-s period: $F_{9,27}=0.85$, $p=0.4769$ (GG); linear contrast over the same period: $F_{1,3}=1.34$, $p=0.3302$ (second-order contrast significant); main effect of interval, i.e., contrast between 1- to 30-s and 31- to 60-s periods: $F_{1,3}=58.61$, $p=0.0046$].

One unexpected feature of the results, which is apparent in both the individual and the average data in Fig. 2, relates to the presence of a "kink," indicating a decrease of the inducer effect, at the second probe-triplet position. As revealed by the results of *t*-tests contrasting the effect at the second triplet-position with those at the surrounding, first and third positions, this effect was significant (First versus second triplet: $t=3.88$, $p=0.0304$; second versus third triplets: $t=5.23$, $p=0.0136$). However, we have no explanation for the effect at this time.

V. EXPERIMENT 3: IS THE EFFECT EAR SPECIFIC?

A. Rationale

Rosenblith *et al.* (1947) mentioned in passing near the end of their article that the aftereffect measured in their study did not transfer across ears. This anecdotal observation, which can easily be verified by listening, is particularly interesting, not only for its potential implications regarding the neural substrate of the aftereffect, but also because it can be used as a tool to further investigate the relationship between the aftereffect described by Rosenblith *et al.* (1947) and the elevation in AM detection thresholds measured in the present study. If the two are related, the induced increase in AM detection thresholds should not be observed when the inducer and probe stimuli are presented to opposite ears.

B. Methods

This experiment used a procedure similar to that of experiment 1, with probe-burst triplets presented at regular time intervals between 4-s bursts of a 100-Hz F0 sine-phase harmonic complex tone or unmodulated noise inducer. Each run started with the uninterrupted presentation of the inducer for 60 s. The unmodulated noise inducer was used as a control. The sine-phase complex inducer and a probe AMF of 250 Hz were selected because this combination yielded the largest effects in experiment 1. The main difference between this and previous experiments was that instead of presenting the inducer and probe diotically, the probe was presented monaurally to the right ear and the inducer was presented either to the same ear (ipsilateral-inducer condition) or to the opposite ear (contralateral-inducer condition).

C. Results

As can be seen in Fig. 3, the results of experiment 3 were consistent across listeners. There is a clear interaction between the inducer type and stimulation mode ($F_{1,3}=620.13$, $p=0.0001$): For the sine-phase complex inducer

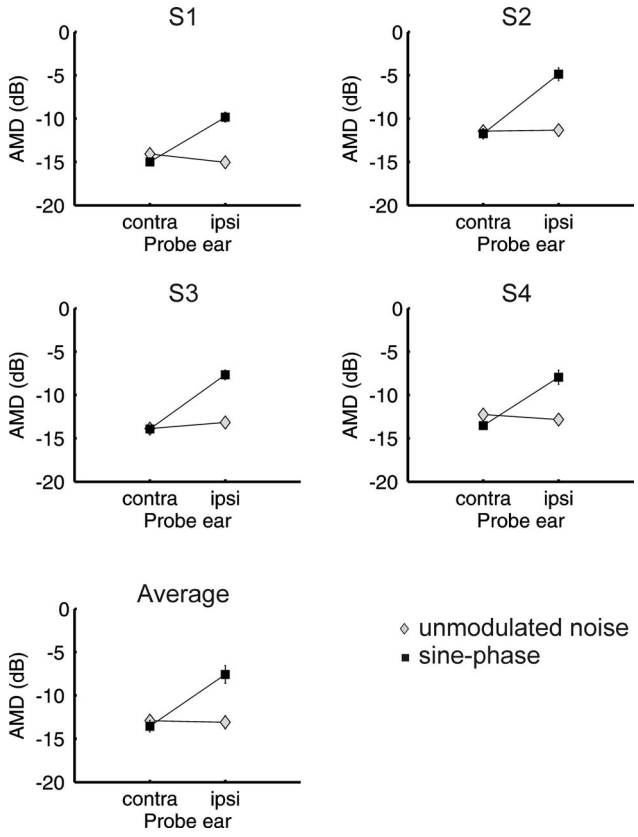


FIG. 3. Amplitude modulation detection thresholds for the right ear, with the sine-phase complex tone (squares) or unmodulated noise (diamonds) inducer presented either to the ipsilateral (right) or the contralateral (left) ear. The probe was sinusoidally amplitude modulated at 250 Hz. Each data point in the four upper panels is an average across four runs of the adaptive procedure, and the error bars in those panels indicate the corresponding standard errors. The bottom panel shows the mean across listeners, and the error bars in that panel correspond to 1 standard error of the mean across listeners.

(closed squares), AM detection thresholds were significantly larger in the ipsilateral-inducer condition than in the contralateral-inducer condition ($F_{1,3}=252.63$, $p=0.0005$). In the contralateral condition, they were not significantly different from those measured with the unmodulated noise inducer ($F_{1,3}=4.94$, $p=0.1127$), which were not different between the ipsilateral- and contralateral-inducer conditions ($F_{1,3}=0.23$, $p=0.6657$). In contrast, in the ipsilateral-inducer condition, the thresholds measured in the presence of the sine-phase inducer were significantly larger than those measured with the unmodulated noise inducer ($F_{1,3}=261.45$, $p=0.0005$). Thus, like the aftereffect discovered by Rosenblith *et al.*, the modulation-adaptation effect measured in the present study does not transfer across ears.

VI. EXPERIMENT 4: DOES THE EFFECT DISAPPEAR WITH PRACTICE?

A. Rationale and procedure

Although numerous investigators have reported significant adaptation effects in the detection of amplitude or frequency modulation (Kay and Matthews, 1972; Green and Kay, 1973, 1974; Regan and Tansley, 1979; Gardner and Wilson, 1979; Davidson *et al.*, 1981; Cole *et al.*, 1981; Tans-

ley and Suffield, 1983; Wojtczak and Viemeister, 2003, 2005), some results suggest that these effects can substantially diminish (Moody *et al.*, 1984; Wakefield and Viemeister, 1984) and even completely vanish (Bruckert *et al.*, 2006) within the course of a few to several hours of practice. For instance, Bruckert *et al.* (2006) found that a 5-kHz pure-tone carrier inducer modulated at a 16-Hz rate, which initially caused a substantial increase in AM detection thresholds measured with probe tones of the same carrier frequency and modulation rates between 4 and 64 Hz, completely lost its effect during the course of 10–12 h of practice. Although they were observed under testing conditions substantially different from those used in the present study, these findings prompted us to add one additional experiment at the end of this study, in order to check whether the inducer still had a significant effect following the many hours of testing accumulated by the listeners during the course of the previous three experiments.

B. Stimuli and procedure

This experiment involved a single 2-h session, during which AM detection thresholds for 250-Hz AMF probe bursts were measured after exposure to a 100-Hz F0 sine-phase complex tone, using the same procedure as in experiment 1. At the beginning of the session, thresholds were measured four times in the absence of the inducer, in order to obtain a baseline. The stimuli in this experiment were presented diotically. Three of the four listeners who had taken part in experiment 3 could be tested in this final experiment. Thus, prior to beginning this final experiment, all listeners had accumulated at least 14 h of experience with the task (18 h when including the control conditions).

C. Results

The results of experiment 4 are shown in Fig. 4. The first three panels show individual data. Listener 2 performed only 9 runs; listener 3 performed 15, and listener 4 performed 20. The average data, including only the first 9 runs in which data from all subjects are available, are shown in the lower right-hand corner. The data points that are shown on the right part, corresponding to the "avr" mark on the *X* axis, represent the mean thresholds (across runs) measured in the presence of the sine-phase inducer (closed square) or in quiet (open circle) in this experiment. The rightmost data points, which correspond to the "E1" mark, are replotted from the 250-Hz condition of experiment 1.

When averaged across listeners, thresholds remained fairly constant across repetitions, and no statistically significant variation across repetitions was observed ($F_{8,16}=0.73$, $p=0.5225$). However, some interindividual differences were apparent in the data. While listener 3's thresholds tended to decrease across repetitions, listener 4's thresholds showed a trend in the opposite direction.

When comparing these data to those of the same three listeners in experiment 1 ("E1"), a reduction in effect size is evident. Whereas the average threshold increase in experiment 1 was 8.8 dB (8.1, 9.5, and 8.8 dB for listeners 2–4), in the current experiment, the average attenuation was only

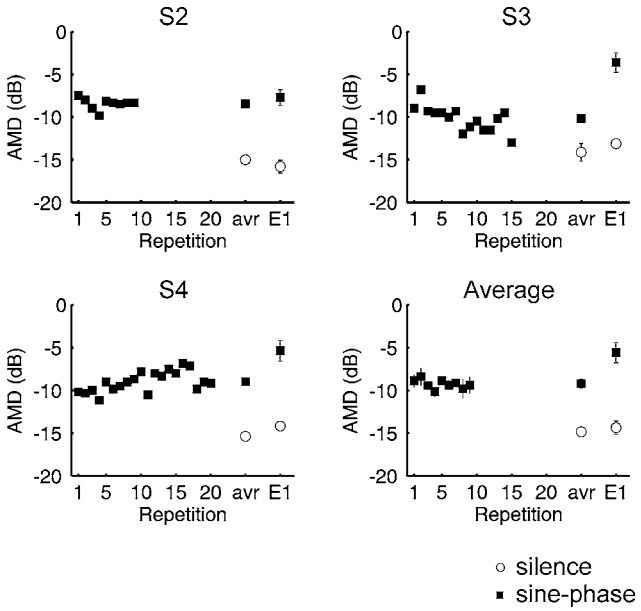


FIG. 4. Amplitude modulation detection thresholds measured in three listeners (2–4) in the presence of a sine-phase pulse train. The data points situated second from the right in each panel correspond to the mean thresholds (across runs) measured in the presence of the sine-phase inducer (closed squares) and in quiet (open circles) in this experiment ("avr"). The rightmost data points replot the two corresponding conditions from experiment 1 ("E1"), for comparison. The lower right panel shows the average across listeners; error bars denote the standard error of the mean across the three listeners.

5.6 dB (6.6, 3.9, and 6.3 dB for listeners 2–4); this corresponds to an average reduction of the effect of approximately 3 dB between the two experiments. This difference was statistically significant ($t=6.65$, $p=0.0219$). Thus, although some effects of learning are evident, the effect remains robust even after 16 h or more of practice.

VII. GENERAL DISCUSSION

A. Summary of results

The main results of this study can be summarized as follows: (1) Following the presentation of a 60-s long pulse train with an average rate of 100 Hz, sinusoidal AM detection thresholds for noise bursts with AM frequencies between 100 and 500 Hz were significantly elevated. (2) Unmodulated noise had no significant effect as an inducer. (3) The elevation in AM detection thresholds induced by a 60-s 100-Hz pulse train decreased over approximately 30 s following the offset of the inducer, after which it was no longer significant. (5) When the inducer and probe were presented to opposite ears, no significant threshold elevation was observed. (6) A significant effect of the inducer was still present in listeners who had received about 16 h practice in the task.

B. Temporal amplitude fluctuations as an essential determinant of the aftereffect

Our use of a relatively low stimulus repetition rate (100 Hz) and a high-pass filter with a relatively high lower cutoff frequency (4 kHz) imposes strong constraints on the factors and mechanisms responsible for the perceptual effects observed in this study. Because the bandwidths of au-

ditory filters with center frequencies above 4 kHz exceed the spacing between consecutive spectral components in a 100-Hz pulse train, these spectral components are not individually resolved in the auditory periphery. This has two important consequences. First, it makes it very unlikely that the results obtained in this study were mediated by spectral cues. Second, the interaction between multiple spectral components within the auditory-filter bandwidths resulted in the auditory-filter outputs fluctuating in amplitude at a rate corresponding to the stimulus F0 or pulse rate. In the following two sections, we discuss how these amplitude modulations may have interfered with the external amplitude modulations imposed on the target probe burst, which the listeners had to detect, and how this may relate to the aftereffect first described by Rosenblith *et al.* (1947).

The finding that the temporally jittered pulse train yielded a significant increase in AM detection thresholds reveals that precise temporal periodicity is not a prerequisite to elicit the aftereffect. The jittered pulse train did not show the same marked high-pass characteristic observed with the periodic inducers, which is compatible with the predictions of an auditory modulation filterbank [see Figs. 1(c) and 1(d)]. On the other hand, in contrast to the predictions of the modulation filterbank, the effect of the jittered pulse train appeared to be somewhat less than that of the periodic sine-phase inducer, although this difference was not statistically significant in our small sample of four subjects. Based on these considerations, we suggest that the effects observed in the present study were mediated by mechanisms, the operation of which was based on relatively fast (100–500 Hz), but not necessarily periodic, amplitude fluctuations in the peripheral auditory responses to the inducer and probe stimuli. Of course, we cannot rule out the possibility that the inducer affects other aspects of perception, which may not involve the coding of rapid temporal fluctuations. Clearly, an exhaustive exploration of the many possible perceptual effects of inducers such as those used here and in the study of Rosenblith *et al.* will require many additional experiments. Exploratory experiments, which we performed in a single listener prior to those described here, showed no evidence for an effect of the pulse-train inducer on either pitch discrimination measured using pulse trains with an F0 close to the inducer, or on detection of identically bandpass filtered noise bursts.

C. Can the effect be explained by AM adaptation?

A possible explanation for the finding of elevated AM detection thresholds following the prolonged presentation of temporally modulated stimuli is in terms of modulation adaptation. Several investigators have found that prolonged exposure to an amplitude- or frequency-modulated tone could cause a temporary impairment in the ability to detect subsequent amplitude or frequency modulation (Kay and Matthews, 1972; Green and Kay, 1973, 1974; Regan and Tansley, 1979; Gardner and Wilson, 1979; Davidson *et al.*, 1981; Cole *et al.*, 1981; Tansley and Suffield, 1983; Moody *et al.*, 1984; Wakefield and Viemeister, 1984; Wojtczak and Viemeister, 2003, 2005). Although this adaptation to modulation has traditionally been studied using pure-tone stimuli, the same phenomenon may have mediated the elevation in AM detection thresholds that was observed here using a noise carrier as a target, and pulse-train inducers, which have significant inherent amplitude modulations. However, there are several differences between the present study and earlier studies on AM adaptation. First, the inherent modulation frequencies in our inducer were all at rates of 100 Hz and above. These rates are substantially higher than those used in earlier studies on AM adaptation, which were usually below 30 Hz, and adaptation effects have been reported to be most prominent around 16 Hz (Tansley and Suffield, 1983). From a phenomenological point of view, amplitude modulations with rates below 30 Hz usually evoke a sensation of flutter; they are effectively perceived as fluctuations in the level of the sound. In contrast, modulations with rates between about 100 and 500 Hz, the range over which the inducer was found to produce a significant elevation in thresholds here, usually evoke a sensation of pitch and/or roughness (Burns and Viemeister, 1976). These phenomenological differences may also reflect, or result in, differences in the underlying adaptation mechanisms.

A second important difference between this study and earlier studies on AM adaptation is that the adaptation observed in our experiments does not appear to transfer across ears. Using frequency-modulated tones, Kay and Matthews (1972) and Green and Kay (1973) found that the adaptation effect transferred largely between the two ears: When the inducer was presented to one ear and the probe to the contralateral ear, the size of the effect was still around 80% of that observed in a condition where inducer and probe were presented to the same ear. This may be a difference between AM and FM adaptation, or may be a result of the different range of modulation frequencies tested. In any case, it suggests the possibility that the effects are mediated by different neural mechanisms, which take place at different levels of the auditory system.

Third, our study showed that adaptation persisted despite considerable exposure and training. This appears to be consistent with early investigations on modulation adaptation, which used subjective measures and reported large and durable effects (e.g., Kay and Matthews, 1972; Green and Kay, 1974; Tansley and Suffield, 1983). However, more recent studies using forced-choice procedures (Moody *et al.*, 1984; Wakefield and Viemeister, 1984; Bruckert *et al.*, 2006) found that the adaptation effect decreased in an orderly fashion across practice sessions; in some cases, after approximately 10–12 h of testing, the effect had completely disappeared (Moody *et al.*, 1984; Bruckert *et al.*, 2006). In contrast, in the present study, despite the different experiments involving approximately 16 h of testing and one of the listeners having even more extensive experience listening to the stimuli, the effect was still present at the end of the study. Although the effect observed in the final experiment was reduced compared to that measured using comparable stimulus conditions in the same three listeners in the first experiment, the finding that the effect is still present after 16 h of training indicates that the effect observed here is less suscep-

tible to practice than that observed in earlier studies on AM adaptation using pure tones at much lower modulation frequencies.

In summary, differences in stimulus characteristics limit direct comparisons between the adaptation effect observed in the present study and those observed in earlier studies on AM adaptation. However, dissimilarities in the characteristics of the adaptation effect observed in this study compared to earlier studies suggest that the form of AM adaptation suggested by the present results is functionally dissimilar from the AM adaptation observed at lower frequencies in earlier studies (Kay and Matthews, 1972; Green and Kay, 1974; Tansley and Suffield, 1983; Wakefield and Viemeister, 1984; Bruckert *et al.*, 2006). The existence of different mechanisms for AM perception at low and high rates has been suggested in different contexts (e.g., Wright and Dai, 1998; Sheft and Yost, 2005). Possibly related neural phenomena could be the decrease of phase locking along the ascending auditory system (Creutzfeld *et al.*, 1980), or the more recent evidence for two separate temporal codes in monkey auditory cortex (Lu *et al.*, 2001) for pulse rates above and below about 20 Hz. It may, for instance, be that modulation rates coded primarily by temporal mechanisms in cortex exhibit different adaptation characteristics from those that are coded by a cortical rate code.

D. Relationship to the Rosenblith *et al.* aftereffect

The stimuli that were found to elevate AM detection thresholds in the present study are similar to those that Rosenblith *et al.* (1947) found to induce a temporary change in the timbre of sounds, and informal listening tests confirmed that they exerted a similar subjective aftereffect. Like Rosenblith *et al.* (1947), we noticed that if the inducer was sufficiently long, the subjective aftereffect could persist for over 10 s; experiment 2 revealed that the aftereffect on AM detection thresholds also took over 10 s to subside. Finally, we confirmed the Rosenblith *et al.* (1947) finding that the subjective aftereffect was not elicited when the inducer was presented to a different ear than the subsequent probe sounds; experiment 3 revealed that the aftereffect on AM detection thresholds did not transfer across ears. Therefore, it seems reasonable to suggest that the subjective aftereffect discovered by Rosenblith *et al.* (1947) is related to the "objective" aftereffect on AM detection thresholds characterized in the present study.

The use of stimuli with specific spectral and temporal characteristics in the present study imposes some new constraints on possible explanations of the timbre modification, which Rosenblith *et al.* (1947) described as an added "metallic" quality. In particular, the present findings suggest that both effects have their origin in the altered perception of temporal envelope fluctuations between about 100 and at least 500 Hz. One possible explanation is that the change in timbre and the elevation in AM detection thresholds are both due to the inducer causing a decrease in AM sensitivity over that range of AM frequencies. It is possible that a subjective attenuation of fast amplitude fluctuations could change the timbre to a metallic quality, by changing the natural roughness of sound to an unnatural quality, which is then perceived as metallic. This might be similar to the slight timbre change sometimes caused by artificial reverberation, which tends to smear the temporal envelope of sounds, thereby effectively low-pass filtering the envelope spectrum. The perception of some listeners, which referred to the aftereffect as "spatial," or "like stepping into another room," might be attributed to a similar association. On the other hand, most sounds take on an unnatural, unpleasant quality, which can be described as "rough," "chopped," or "metallic," when they are amplitude modulated at relatively fast rates. In fact, we noticed in informal listening experiments that recorded speech or environmental sounds with added amplitude modulations in the 100- to 500-Hz range sounded more similar to what the original sounded like when presented immediately after a 100-Hz pulse train inducer. Based on these observations, the auditory aftereffect discovered by Rosenblith *et al.* (1947) may alternatively be due to a form of "persistence," rather than adaptation, in the perception of AM. According to this interpretation, the reason that AM detection thresholds were elevated after the presentation of the inducer is not that listeners were less able to hear the AM imposed onto the target probe burst (as assumed by the "AM adaptation" explanation), but rather that they heard at least one of the other two probe bursts as modulated too, because some of the modulations present in the inducer persisted beyond its termination. The interpretation in terms of persistence is consistent with reports from some listeners, who reported having sometimes had the impression that more than one probe sound in the trial was modulated. Further study is required in order to determine whether adaptation or persistence is actually responsible for the aftereffect.

E. Neural locus?

The neural substrates of visual aftereffects have received substantially more attention than those of auditory aftereffects. The results of several studies in the visual modality indicate selective adaptation at the cortical level as a key mechanism behind various aftereffects (Movshon and Lennie, 1979). For example, the motion aftereffect has been attributed to the adaptation of motion-selective neurons in area V5 (Kohn and Movshon, 2004). However, it should be noted that some visual aftereffects may be explained by adaptation at the retinal level (Barlow and Sparrock, 1964). As mentioned earlier, neural correlates of the Zwicker tone have been proposed at the cortical level in both humans (Hoke *et al.*, 1996, 1998) and cats (Norena and Eggermont, 2003). Psychophysical results alone cannot precisely pin down the neural locus of the Rosenblith *et al.* aftereffect, or that of AM adaptation. The observation that the effect of the inducer in both the present study and that of Rosenblith *et al.* was strongly ear-specific might suggest a fairly peripheral neural locus, prior to binaural integration. Neurons that exhibit bandpass tuning to AM have been described as early as in the ventral cochlear nucleus (Møller, 1972; Frisina *et al.*, 1990). Such neurons might also mediate selective adaptation to AM. On the other hand, we cannot rule out the possibility that selective adaptation effects in the AM domain originate at a

higher level of processing in the auditory system, where neural responses may be selective to spatial location, rather than simply ear of entry. For example, selective adaptation to AM has been demonstrated in the auditory cortex (Barlett and Wang, 2005). At this point, further insights may be gained from physiological studies into the underlying mechanisms of those effects.

ACKNOWLEDGMENTS

This work was carried out while the authors were at the Massachusetts Institute of Technology's Research Laboratory of Electronics, Cambridge, MA. We thank Daniel Pressnitzer for pointing us to the publication of Rosenblith *et al.* (1947), and Torsten Dau and Stephan Ewert for generously providing us the code of the modulation filterbank. This research was supported by National Institutes of Health Grant No. R01 DC 03909, Deutsche Forschungsgemeinschaft Grant GU 593/2-1, and the Dietmar-Hopp-Stiftung.

[1]On the first triplet, listener 3 was unable to detect the AM above chance level in any of the AM depth conditions tested. The corresponding data point was artificially set to an AMD of 6 dB, the highest modulation depth tested. The threshold AMD measured in the absence of the inducer in this listener was 14 dB, yielding an estimated attenuation of 8 dB, as plotted in the graph.

Addams, R. (**1834**). "An account of a peculiar optical phenomenon seen after having looked at a moving body," London and Edinburgh Philosophical Magazine and Journal of Science **5**, 373–374.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Barlett, E. L., and Wang, X. (**2005**). "Long-lasting modulation by stimulus context in primate auditory cortex," J. Neurophysiol. **94**, 83–104.

Barlow, H. B. (**1990**). "A theory about the functional role and synaptic mechanisms of visual after-effects," in *Vision: Coding and Efficiency*, edited by C. Blakemore (Cambridge University Press, Cambridge), pp. 363–375.

Barlow, H. B., and Hill, R. M. (**1963**). "Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects," Nature (London) **28**, 1345–1347.

Barlow, H. B., and Sparrock, J. M. B. (**1964**). "The role of afterimages in dark adaptation," Science **144**, 1309–1314.

Blakemore, C., and Campbell, F. W. (**1969**). "Adaptation to spatial stimuli," J. Physiol. Paris **200**, 11–13.

Blakemore, C., and Julesz, B. (**1971**). "Stereoscopic depth aftereffect produced without monocular cues," Science **171**, 286–288.

Blakemore, C., and Sutton, P. (**1969**). "Size adaptation: A new aftereffect," Science **166**, 245–247.

Bruckert, L., Herrmann, M., and Lorenzi, C. (**2006**). "No adaptation in the AM domain in trained listeners," J. Acoust. Soc. Am. **199**, 3542–3545.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Clifford, C. W., Wenderoth, P., and Spehar, B. (**2000**). "A functional angle on some after-effects in cortical vision," Proc. Biol. Sci. **267**, 1705–1710.

Cole, D., Moody, D. B., and Steebens, W. C. (**1981**). "On the existence and evanescence of FM channels," J. Acoust. Soc. Am. **70**, Suppl. 1, S88.

Creutzfeldt, O., Hellweg, F. C., and Schreiner, C. (**1980**). "Thalamocorticall transformation of responses to complex auditory stimuli," Exp. Brain Res. **39**, 87–104.

Davidson, L. M., Stebbens, W. C., Moody, D. B., and Cole, D. M. (**1981**). "Detection of frequency modulation (FM) following FM adaptation," J. Acoust. Soc. Am. **69**, Suppl. 1, S105.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Franosch, J. M., Kempter, R., Fastl, H., and van Hemmen, J. L. (**2003**). "Zwicker tone illusion and noise reduction in the auditory system," Phys. Rev. Lett. **90**, 178103.

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (**1990**). "Endcoding of amplitude modulation in the gerbil cochlea nucleus. I. A hierarchy of enhancement," Hear. Res. **44**, 99–122.

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.

Gibson, J. J., and Radner, M. (**1937**). "Adaptation, after-effect, and contrast in the perception of tilted lines. I. Quantitative studies," J. Exp. Psychol. **20**, 453–467.

Green, G. G. R., and Kay, R. H. (**1973**). "The adequate stimuli for channels in the human auditory pathways concerned with the modulation present in frequency-modulated tones," J. Physiol. (London) **234**, 50–52P.

Green, G. G. R., and Kay, R. H. (**1974**). "Channels in the human auditory pathways concerned with the waveform of the modulation present in amplitude- and frequency-modulated tones," J. Physiol. (London) **241**, 29–30P.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., and Rupp, A. (**2007**). "The effect of temporal context on the sustained pitch response in human auditory cortex," Cereb. Cortex **17**, 552–561.

He, S., and MacLeod, D. I. (**2001**). "Orientation-selective adaptation and tilt after-effect from invisible patterns," Nature (London) **411**, 473–476.

Hoke, E. S., Hoke, M., and Ross, B. (**1996**). "Neurophysiological correlate of the auditory after-image ('Zwicker tone')," Audiol. Neuro-Otol. **1**, 161–174.

Hoke, E. S., Ross, B., and Hoke, M. (**1998**). "Auditory afterimage: Tonotopic representation in the auditory cortex," NeuroReport **9**, 3065–3068.

Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.

Irino, T., and Patterson, R. D. (**1997**). "A time-domain, level-dependent auditory filter: The gammachirp," J. Acoust. Soc. Am. **101**, 412–419.

Jin, D. Z., Dragoi, V., Sur, M., and Seung, H. S. (**2005**). "Tilt aftereffect and adaptation-induced changes in orientation tuning in visual cortex." J. Neurophysiol. **94**, 4038–4050.

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.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.

Kohn, A., and Movshon, J. A. (**2004**). "Adaptation changes the direction tuning of macaque MT neurons," Nat. Neurosci. **7**, 764–772.

Lu, T., Liang, L., and Wang, X. (**2001**). "Temporal and rate representation of time-varying signals in the auditory cortex of awake primates," Nat. Neurosci. **85**, 2364–2380.

Lummis, R. C., and Guttman, N. (**1972**). "Exploratory studies of Zwicker's 'negative afterimage' in hearing," J. Acoust. Soc. Am. **51**, 1930–1944.

MacKay, D. M. (**1964**). "Central adaptation in mechanisms of form vision," Nature (London) **203**, 993–994.

Maffei, L., Fiorentini, A., and Bisti, S. (**1973**). "Neural correlates of perceptual adaptation to gratings," Science **182**, 1036–1038.

Magnussen, S., and Johnsen, T. (**1986**). "Temporal aspects of spatial adaptation. A study of the tilt aftereffect," Vision Res. **26**, 661–672.

Mather, G., Verstraten, F., and Anstis, S. (**1998**). *The Motion Aftereffect: A Modern Perspective* (MIT, Cambridge, MA).

Mitchell, D. E., and Muir, D. W. (**1976**). "Does the tilt after-effect occur in the oblique meridian?," Vision Res. **16**, 609–613.

Møller, A. R. (**1972**). "Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat," Acta Physiol. Scand. **86**, 223–238.

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.

Movshon, J. A., and Lennie, P. (**1979**). "Pattern-selective adaptation in visual cortical neurons," Nature (London) **278**, 850–852.

Norena, A., Micheyl, C., and Chery-Croze, S. (**2000**). "An auditory negative after-image as a human model of tinnitus," Hear. Res. **149**, 24–32.

Norena, A., Micheyl, C., Garnier, S., and Chery-Croze, S. (**2002**). "Loudness changes associated with the perception of an auditory after-image," Int. J. Audiol. **41**, 202–207.

Norena, A. J., and Eggermont, J. J. (**2003**). "Neural correlates of an auditory afterimage in primary auditory cortex," J. Assoc. Res. Otolaryngol. **4**, 312–328.

Oxenham, A. J., and Dau, T. (**2001a**). "Reconciling frequency selectivity and phase effects in masking," J. Acoust. Soc. Am. **110**, 1525–1538.

Oxenham, A. J., and Dau, T. (**2001b**). "Towards a measure of auditory-filter phase response," J. Acoust. Soc. Am. **110**, 3169–3178.
Regan, D., and Tansley, B. (**1979**). "Selective adaptation to frequency-modulated tones: Evidence for an information-processing channel selective to frequency changes," J. Acoust. Soc. Am. **65**, 1249–1257.
Rosenblith, W. A., Miller, G. A., Egan, J. P., Hirsh, I. J., and Thomas, G. J. (**1947**). "An auditory afterimage?" Science **106**, 333–335.
Sekuler, R., and Pantle, A. (**1967**). "A model for after-effects of seen movement," Vision Res. **7**, 427–439.
Sheft, S., and Yost, W. A. (**2005**). "Minimum integration times for processing of amplitude modulation," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveign, S. McAdams, and L. Collet (Springer, New York), pp. 244–250.
Shu, Z. J., Swindale, N. V., and Cynader, M. S. (**1993**). "Spectral motion produces an auditory after-effect," Nature (London) **19**, 721–723.
Summerfield, Q., Haggard, M., Foster, J., and Gray, S. (**1984**). "Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect," Percept. Psychophys. **35**, 203–213.
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.
Viemeister, N. F. (**1980**). "Adaptation of masking," in *Psychophysical, Physiological, and Behavioural Studies in Hearing*, edited by G. Van den Brink and F. A. Bilsen (Delft University Press, Delft, The Netherlands), pp. 190–199.
Viemeister, N. F., and Bacon, S. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**, 1502–1507.
Wade, N. J. (**1994**). "A selective history of the study of visual motion aftereffects." Perception **23**, 1111–1134.
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.
Wiegrebe, L., Kossl, M., and Schmidt, S. (**1996**). "Auditory enhancement at the absolute threshold of hearing and its relationship to the Zwicker tone," Hear. Res. **100**, 171–180.
Wilson, J. P. (**1970**). "An auditory after-image," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden, The Netherlands), pp. 303–318.
Wojtczak, M., and Viemeister, N. F. (**2003**). "Suprathreshhold effects of adaptation produced by amplitude modulation," J. Acoust. Soc. Am. **114**, 991–997.
Wojtczak, M., and Viemeister, N. F. (**2005**). "Forward masking of amplitude modulation: Basic characteristics," J. Acoust. Soc. Am. **118**, 3198–3210.
Wright, B. A., and Dai, H. (**1998**). "Detection of sinusoidal amplitude modulation at unexpected rates," J. Acoust. Soc. Am. **104**, 2991–2996.
Zwicker, E. (**1964**). "'Negative afterimage' in hearing," J. Acoust. Soc. Am. **36**, 2413–2415.

The relationship between precursor level and the temporal effect[a)]

Elizabeth A. Strickland[b)]
Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907-2038, USA

(Received 3 August 2007; accepted 13 November 2007)

Previous studies have suggested that temporal effects in masking may be consistent with a decrease in cochlear gain. One paradigm used to show this is to measure the level of a long-duration masker required to just mask a short-duration tone that occurs near masker onset. The temporal effect is revealed when the signal is detected at a lower signal-to-noise ratio following preceding stimulation (either an extension of the masker or a separate precursor). The present study examined whether this effect depends on precursor level. The signal was a 10-ms, 4-kHz tone. The masker was 200 ms. A fixed-level precursor had the same frequency characteristics as the masker, and was 205 ms. The masker and precursor had either no notch or a wide notch about the signal frequency. For a given precursor level, the growth of masker level with signal level was determined. These data were used to estimate input–output functions. The results are consistent with a graded decrease in gain at the signal frequency when there is no notch in the masker and precursor, and a graded decrease in suppression when there is a large notch. These results could be consistent with the action of the medial olivocochlear reflex. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821977]

PACS number(s): 43.66.Dc, 43.66.Mk, 43.66.Ba [MW] Pages: 946–954

I. INTRODUCTION

A recent focus in psychoacoustic research has been the measurement of compression in the auditory system. The input–output function of the cochlea shows a compressive response for midlevel inputs, which is due to level-dependent amplification by the outer hair cells, an effect called the "active process" (for a review, see Robles and Ruggero, 2001). Many psychoacoustic effects are well explained by taking into account the transformation of stimuli by the input–output function of the cochlea (for a review, see Oxenham and Bacon, 2004). Although this input–output function is often assumed to be static, this paper provides evidence that it may change during the course of acoustic stimulation.

Evidence for a dynamic change in cochlear amplification comes from the fact that a short-duration signal presented at the onset of a masker (onset condition) is harder to hear than one delayed from the onset of the masker, or one preceded by another sound (precursor condition) (e.g. Scholl, 1962). This effect has been called overshoot (Zwicker, 1965) or, as in this paper, the temporal effect (Hicks and Bacon, 1992). Although the temporal effect may depend on effects at multiple levels of the auditory system, there is a growing body of evidence that it is consistent with a change in amplification at the level of the cochlea. Factors which affect the active process in the cochlea also tend to decrease the temporal effect. These factors include temporary threshold shift (Champlin and McFadden, 1989), permanent cochlear hearing loss (Bacon *et al.*, 1988; Bacon and Takahashi, 1992; Kimberley *et al.*, 1989; Strickland and Krishnan, 2005), and the ingestion of large amounts of aspirin (McFadden and Champlin, 1990). In these cases, the temporal effect is reduced because thresholds *improve* in the onset condition; that is, the onset condition thresholds move toward the precursor condition thresholds. Since factors which affect the active process produce thresholds that look more like the precursor condition, this suggests that the active process is less important, or the gain is reduced, in the precursor condition. This would imply that for normal-hearing listeners, gain decreases between the onset and precursor conditions, an idea proposed by Schmidt and Zwicker (1991) and developed in detail by von Klitzing and Kohlrausch (1994). In a series of papers, the author has shown that input–output functions derived from temporal effect data are consistent with a decrease in gain at the frequency of the stimulation preceding the signal (Strickland, 2001, 2004; Strickland and Krishnan, 2005). This results in a decrease in gain at the signal frequency, if preceding stimulation is at the signal frequency, or an apparent decrease in suppression, if the preceding stimulation is away from the signal frequency.

This dynamic response to sound could be mediated by the medial olivocochlear reflex (MOCR). The MOCR is a frequency-specific decrease in amplification caused by stimulation of the medial olivocochlear bundle, a pathway that feeds back to the outer hair cells of the cochlea from the level of the superior olivary complex (Warr and Guinan, 1979; Warr 1980). If the temporal effect is due to the action of the MOCR, the amount of decrease in gain should depend on the level of the preceding stimulation. The firing of MOC neurons has been shown to increase with sound level (Liberman, 1988). The effect of a contralateral sound on stimulus frequency otoacoustic emission level also increases with the level of the contralateral sound (Backus and Guinan, 2006).

[a)]Portions of this research were presented at the 141st Meeting of the Acoustical Society of America, Chicago, IL.
[b)]Electronic mail: estrick@purdue.edu.

 0001-4966/2008/123(2)/946/9/$23.00

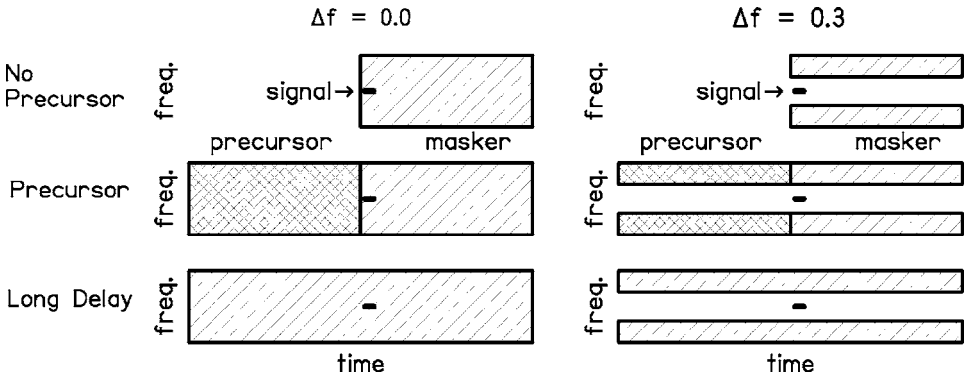


FIG. 1. Schematic showing the spectral and temporal characteristics of the signal, masker, and precursor for the no notch ($\Delta f=0.0$) and notch ($\Delta f=0.3$) conditions, for the no-precursor, precursor, and long-delay conditions.

Several psychoacoustic studies of the temporal effect have examined the role of level, either by using a fixed precursor before the signal and masker (Zwicker, 1965; Carlyon, 1989; Hicks and Bacon, 1992; Strickland, 2001), or presenting a continuous noise band in addition to the masker and signal (Carlyon, 1987). These studies have shown that the temporal effect does change with precursor or continuous noise level, and that the most effective precursor level is at or just below the masker level. The purpose of the present study was to measure the effects of precursor level for multiple signal levels, so that input–output functions could be derived and the decrease in gain measured. The effect of precursor level was measured for a broadband masker and a masker with a large notch around the signal frequency, to measure effects of a decrease in gain at excitatory and suppressive masker frequencies.

II. METHODS

A. Stimuli

The signal was a 4 kHz sinusoid. This signal frequency was chosen because previous studies have shown temporal effects for noise maskers to be larger at higher frequencies (Carlyon, 1987; Bacon and Takahashi, 1992; Strickland 2001, 2004). The masker was a noise centered about the signal frequency. The outer spectral edges of the masker were fixed at $0.2f_s$ and $1.8f_s$ (0.8 and 7.2 kHz), where f_s is the signal frequency. The masker was either broadband, or had a notch centered about the signal frequency. The relative notch edges are in units of $\Delta f=|f\text{-}f_s|/f_s$, where f is the frequency of the notch edge. The spectral edges of the notch were set at $\Delta f=0.0$ (no notch) or $\Delta f=0.3$ (2.8 and 5.2 kHz). The signal level was fixed, and the masker level varied to determine threshold. To measure the temporal effect, thresholds were also measured with a precursor before the signal and masker. The precursor was a noise with the same spectral characteristics as the masker, except that it was fixed in level for an entire range of signal levels.

The signal duration was 10 ms, including 5-ms $\cos^2$ onset and offset ramps (no steady state). This duration was chosen to be short enough to show a temporal effect, but long enough to avoid effects of spectral splatter (Bacon and Viemeister, 1985). In the no-precursor and precursor conditions, the masker duration was 200 ms, including 8.5-ms $\cos^2$ onset and offset ramps.[1] The signal onset was delayed 2 ms from the masker onset. The precursor duration was 205 ms, including 8.5-ms $\cos^2$ onset and offset ramps. The precursor offset overlapped the masker onset by 5 ms (and thus overlapped the signal by 3 ms). Signal thresholds were also measured in the presence of the precursor with no simultaneous masker present. A long-delay condition was included for comparison with the precursor condition and with previous studies. In the long-delay condition, the masker was 400 ms, and the signal onset was delayed 202 ms from the masker onset. Conditions are shown schematically in Fig. 1.

The signal, masker, and precursor were digitally produced in the frequency domain at a sampling rate of 25 kHz, and were output through three separate D/A channels (TDT DA3-4). They were low-pass filtered at 10 kHz (TDT FT5 and FT6-2). The level was controlled by programmable attenuators (TDT PA4). In addition, for low signal levels, additional attenuation of the masker was provided by a manual attenuator (Leader LAT-45). The stimuli were mixed (TDT SM3), led to a headphone buffer (TDT HB6), and presented to one of two ER-2 insert earphones. These earphones have a flat frequency response from 250 to 8000 Hz.

B. Procedures

Listeners were tested in a double-walled sound-attenuating booth. Thresholds were measured using a three-interval forced-choice adaptive tracking procedure with a two-up, one-down stepping rule. This estimates the 71% correct point on the psychometric function (Levitt, 1971). Temporal intervals were marked visually on a computer monitor, and listeners responded via a computer keyboard. Visual feedback was provided. The initial step size was 5 dB, and was decreased to 2 dB after the second reversal. Thresholds were taken as the average of the last even number of reversals at the smaller step size in a set of 50 trials. Blocks for which the standard deviation was 5 dB or greater were discarded. At least two thresholds were averaged for each data point. Each listener was tested with at least two precursor levels. The number of levels used depended on the amount of time the subject was available for testing.

C. Subjects

Five listeners participated in the experiment, two males and three females. All had hearing thresholds within laboratory norms for long-duration signals at octave frequencies from 250 to 8000 Hz. The age range was from 18 to 43 years, with a median of 21 years.

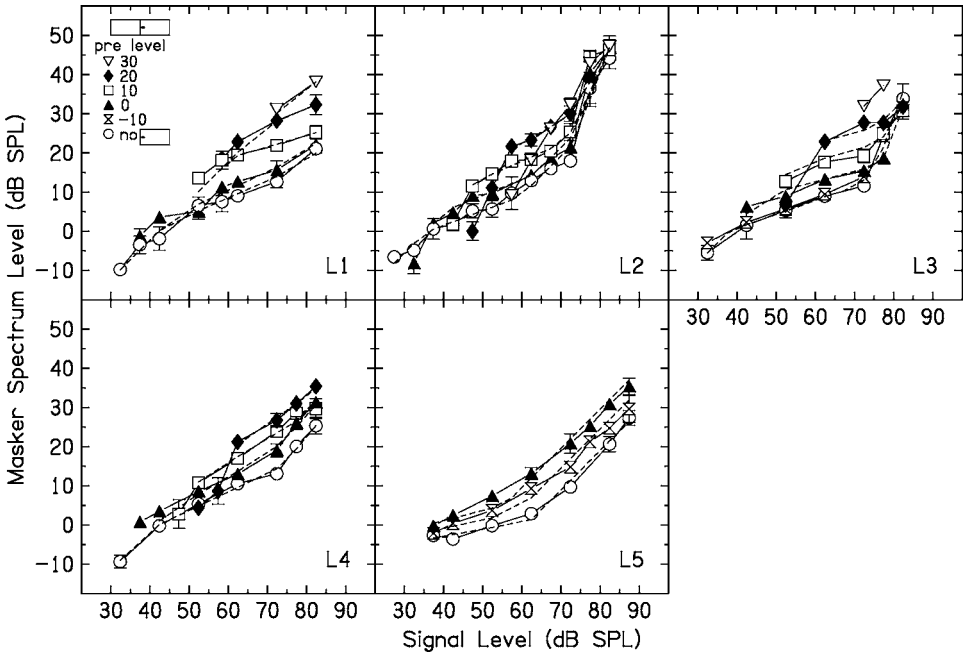


FIG. 2. No-precursor (open circles) and precursor (other symbols, see the legend) data for five listeners when the masker and precursor $\Delta f=0.0$ (no notch). Dashed lines are predictions from a model (see the text for details).

III. RESULTS

A. On-frequency

Results for individual listeners when $\Delta f=0.0$ are shown in Fig. 2. Open circles are masker thresholds with no precursor. The other symbols are masker thresholds following fixed-level precursors, with precursor levels shown in the legend in the upper left corner. Dashed lines are predictions from a model which will be discussed later. For a given signal level, the temporal effect is the difference between the open circles and each of the other symbols. The precursors increase quiet thresholds for the signal, so that the minimum signal level that could be tested increased with precursor level. At signal levels above threshold, the precursor increases the masker level at threshold, and thus causes a temporal effect. This effect tends to be greatest for midlevel signals. The temporal effect is graded, so that in general masker thresholds increase with increasing precursor level. Note that a precursor can have an effect even if it is below the level of the masker. For example, for L1, in the 10-dB precursor condition (open squares), the temporal effect was more than 10 dB, and a 20-dB masker level was required to mask the 60-dB SPL signal.

The temporal effect was measured across multiple signal levels so that input–output functions could be derived for each of the precursor conditions. This was done to determine whether the gain of the derived input–output functions decreases as precursor level increases, which would be consistent with activation of the MOCR. As in previous studies, it was assumed that the thresholds represent a constant signal-plus-masker to masker ratio in a filter centered at the signal frequency after transformation by an approximation to the cochlear input–output function (Strickland, 2001, 2004; Strickland and Krishnan, 2005). This will be called the criterion ratio. The masker level was the level estimated to pass through a filter centered at the signal frequency. The equivalent rectangular bandwidth was 453 Hz, from a filter derived by Glasberg and Moore (2000) from data of Baker *et al.* (1998). Thus the masker levels were estimated as the threshold spectrum levels plus 26.6 dB. A function proposed by Yasin and Plack (2003) was used to fit the data. This function is composed of three linear sections described by the following:

$$L_{\text{out}}=\begin{cases} L_{\text{in}}+G & (L_{\text{in}}<=\text{BP}_1), \quad (1)\\ cL_{\text{in}}+k_1+G & (\text{BP}_1<L_{\text{in}}<=\text{BP}_2), \quad (2)\\ L_{\text{in}} & (L_{\text{in}}>\text{BP}_2). \quad (3)\end{cases}$$

L_{in}, the input level, L_{out}, the output level, and G are all in units of decibels. These three equations produce a three-line fit to the data, with slopes of 1 below the lower breakpoint (BP_1, in hertz) and above the higher breakpoint (BP_2, in hertz), and a section with a slope of less than 1 (c) between the two breakpoints. The factor k_1, where $k_1=\text{BP}_1(1-c)$, is a correction so that the sections are linked appropriately. The maximum gain, G, was calculated as $-(\text{BP}_2(c-1)+k_1)$, so that above BP_2 the output was constrained to equal the input, which is a modification from Yasin and Plack (2003).

In general, the data for the no-precursor condition were fit first, with the constraints that BP_1 and BP_2 had to be within the data range, and $0<c<1$. For listeners with no clear upper breakpoint in the data, BP_2 was fixed at the highest signal level. The fitting program converged on the criterion ratio, c, BP_1, and BP_2. The criterion ratio and BP_2 were then fixed at these values, and data for each precursor level were fit. For some listeners (e.g., L2 with a 20-dB precursor), the slope of the increase in masker level with signal level was greater than 1 at the lowest signal levels when there was a precursor present. This may be due to an effect near threshold that does not fit the model presented above. These points were excluded from the fits. The parameters of the input–output functions for the different subjects and conditions are shown in Table I. The derived input–output functions are shown in Fig. 3. Masker thresholds predicted by the model

TABLE I. Parameters for three-line functions fit to the data in Fig. 2. Slope is the change in maximum gain with precursor level.

Listener	Precursor level	Criterion	c	BP_1	BP_2	G	rms error	Slope
L1	No	15.49	0.43	44.5	82.4	21.45	1.04	−0.59
	0		0.47	46.9		18.77	2.31	
	10		0.29	61.7		14.68	1.98	
	20		0.41	69.7		7.51	2.04	
	30		0.42	79.5		1.66	0.96	
L2	No	8.51	0.36	35.6	73.6	24.33	1.02	−0.58
	0		0.40	41.2		19.47	0.98	
	10		0.42	48.7		14.33	1.49	
	20		0.55	55.8		7.95	0.36	
	30		0.65	48.5		8.76	1.26	
L3	No	11.07	0.34	38.7	75.7	24.32	0.45	−0.51
	−10		0.35	39.4		23.47	1.25	
	0		0.26	47.6		20.70	1.01	
	10		0.27	54.8		15.20	0.80	
	20		0.27	60.8		10.87	1.00	
	30		0.82	59.7		2.68	0	
L4	No	14.39	0.47	41.1	81.7	21.64	0.64	−0.46
	0		0.57	45.5		15.55	2.12	
	10		0.60	53.7		11.13	0.87	
	20		0.69	61.2		6.28	0.53	
L5	No	6.08	0.18	28.2	87.4	48.59	0.56	−0.41
	−10		0.21	30.3		44.90	0.72	
	0		0.26	32.3		40.77	0.67	

are shown by the dashed lines in Fig. 2. The model predicts the data well, so that in some cases the dashed line is obscured by the data.

In examining Table I and Fig. 3, it can be seen that BP_1 systematically increases with precursor level, and the slope c increases slightly with precursor level for some listeners. The net effect is that maximum gain systematically decreases as precursor level increases. This decrease in gain is similar to the decrease in gain between normal-hearing and hearing-impaired listeners shown by Plack *et al.* (2004). The decrease in maximum gain is 4–6 dB for every 10 dB increase in precursor level, as shown by the values in the final column of Table I. For L2, the gain for the 30-dB precursor condition was not included in the fit in the final column, because the gain did not change between the 20- and the 30-dB precursor condition. Although a decrease in gain sounds as if it would be detrimental, it decreases the response to the masker more

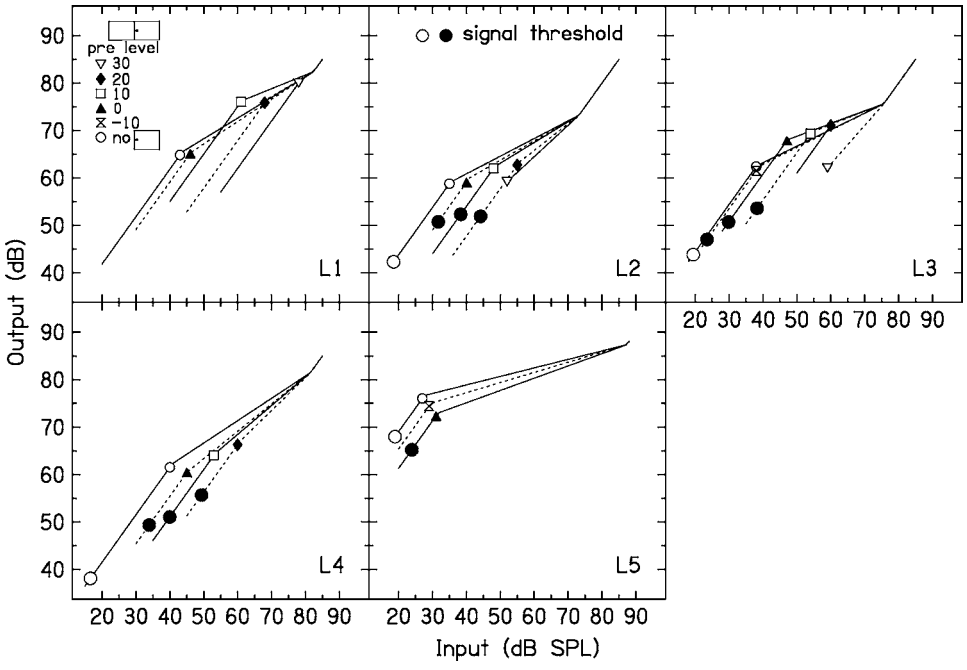


FIG. 3. Input–output functions estimated from the data in Fig. 2. The small symbols correspond to the different precursor levels in Fig. 2. The large circles are thresholds measured either in quiet (open) or in the presence of a precursor (closed). These are plotted on the input–output function estimated with the same precursor.

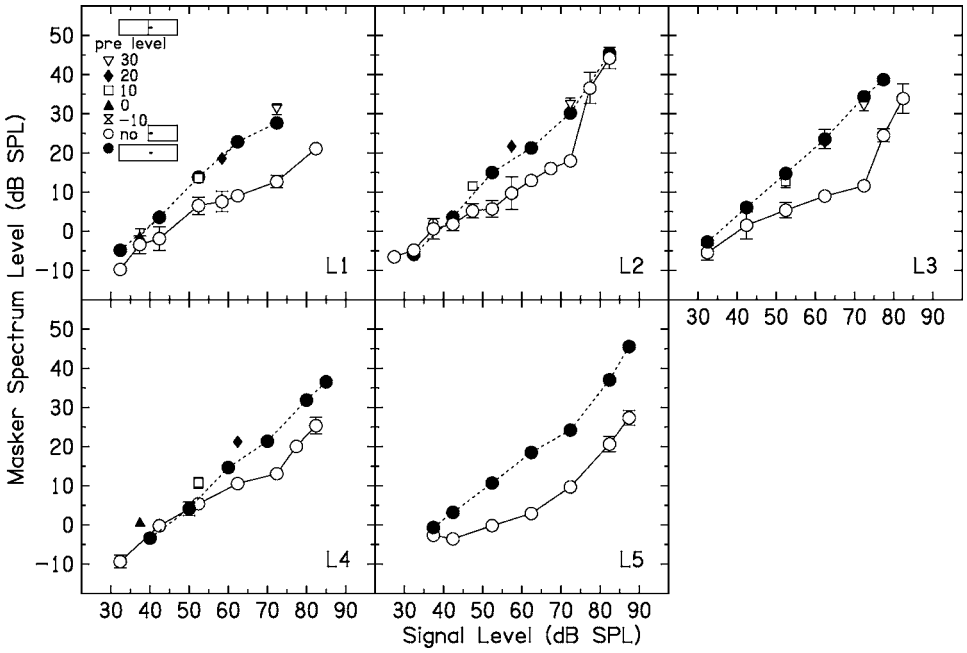


FIG. 4. Long-delay (closed circles) thresholds when the masker and precursor $\Delta f=0.0$, along with no-precursor (open circles) thresholds from Fig. 2. Also replotted from Fig. 2 are thresholds when the precursor and masker were at approximately equal levels. The long-delay data trace along these points from the functions from Fig. 2.

than the response to the signal (see, e.g., Strickland, 2004, Fig. 5), and this may be why the masker level has to be increased following a precursor.

In general, the maximum gain values for the no-precursor condition are lower than values reported in Strickland (2004) and Strickland and Krishnan (2005). In those studies, a polynomial function from Glasberg and Moore (2000) was used to fit the data, rather than the three-line function used in the present paper. Informal tests from the author's lab have found that the polynomial fit can give a maximum gain estimate that is over 20 dB higher than the estimate from the three-line fit. In the present paper, the main concern is with relative changes in gain, which would likely be the same with either fitting procedure. It is also important to note that the estimates of gain in all three of these studies are from simultaneous masking, and thus may include the effects of suppression, which would tend to decrease the gain.

On the input–output functions shown in Fig. 3, small symbols are plotted at BP1 of each function so that it may be paired with the legend indicating which precursor condition the fitted function represents. The large closed circles are thresholds for the signal in quiet and for the signal presented after the fixed-level precursors, with no simultaneous masker. The quiet threshold is plotted on the input–output function derived from the no-precursor data in Fig. 2; the threshold for the signal following a 0-dB spectrum level precursor is plotted on the input–output function for the 0-dB spectrum level precursor, etc. Missing symbols mean that the threshold was not measured. For a given listener, if the signal levels at threshold were all at the same output level, this would be consistent with forward masking by the precursor being due to a decrease in gain at the signal frequency place. For L5, the quiet threshold and the threshold in the presence of the precursor at 0-dB spectrum level are nearly equal. For L2, L3, and L4, the thresholds in the presence of a precursor are at an output above that for the signal in quiet, but are similar to each other. The precursor actually overlapped the signal by 3 ms, so some of the threshold shift following a precursor could be due to simultaneous masking. It is interesting to note, however, that Plack *et al.* (2004) found that when output levels for signal threshold were calculated in the same manner as noted earlier, the output levels for hearing-impaired listeners were above those for normal hearing listeners. Plack *et al.* suggested that this might indicate inner hair cell damage in the hearing-impaired listeners. If the precursor in the present study activates the MOCR, this would not be expected to affect the inner hair cells. Although the detection processes used near threshold may not be well understood, the thresholds shown in Fig. 3 show some support for the idea that forward masking is at least partially due to a decrease in gain at the signal frequency.

In Fig. 4, thresholds are shown for the long-delay condition (closed circles) and the no-precursor condition (open circles). Replotted from Fig. 2 are thresholds from conditions in which precursor and masker levels were nearly equal (individual symbols). Note that the long-delay function connects these points, as would be expected. The slope of the long-delay function is close to one, as has been noted in previous studies (Zwicker, 1965; Bacon, 1990). However, the long-delay function does not represent a condition where the gain is fixed at the signal frequency, but rather combines data across conditions where the gain is changing.

B. Off-frequency

Results for the five listeners when $\Delta f=0.3$ are shown in Fig. 5. Some listeners showed a clear improvement in thresholds in the no-precursor condition across sessions, which was not the case when $\Delta f=0.0$. For these listeners, the initial runs were discarded and only the final values used. The presence of a precursor causes an increase in masker level, and this is

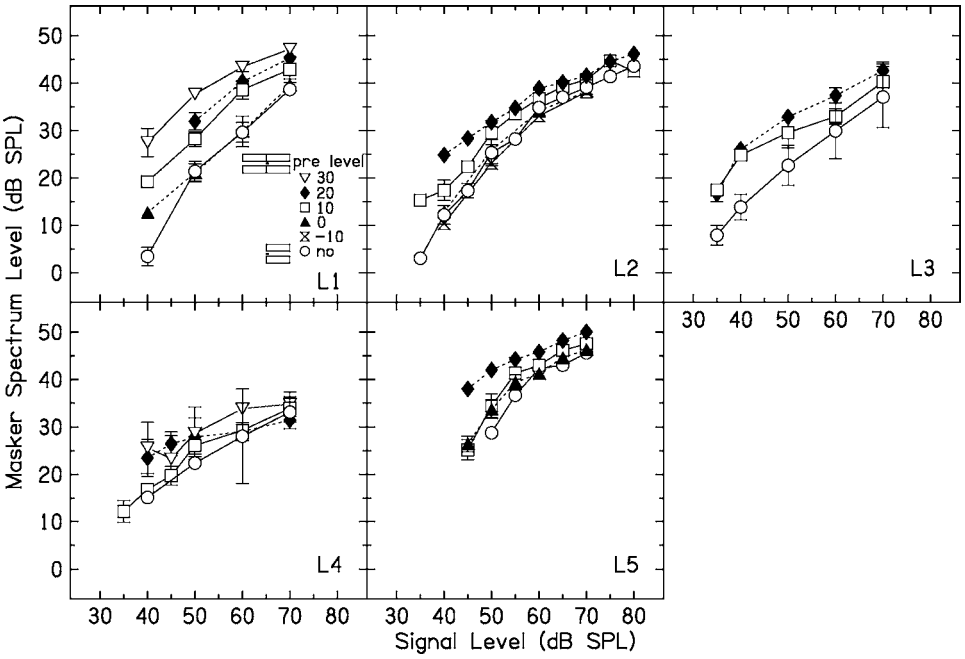


FIG. 5. No-precursor (open circles) and precursor (other symbols, see the legend) data for five listeners when the masker and precursor Δf=0.3.

graded with precursor level. The increase tends to be largest at low signal levels, and decreases to a fairly constant amount at higher signal levels.

How can these data be interpreted in terms of input–output functions? For this fairly wide notch width, the excitatory response to the masker at the signal place should be close to linear, based on physiological estimates from the chinchilla (Ruggero *et al.*, 1997). If masking were purely excitatory, the masking function would give an estimate of the input–output function at the signal place (such as the input–output functions in Fig. 3). If the data are interpreted in this way, it can be seen that the change in the input–output functions is the opposite of what it was when Δf=0.0. That is, the lower breakpoint of the input–output function moves to the left, and the slope of the function at higher signal levels may decrease with precursor level. This would appear to be consistent with an *increase* in the gain of the function with precursor level.

This curious result may be explained if suppressive masking is also considered. Earlier studies have proposed that results in this type of condition, where the masker energy is removed from the signal frequency, could be consistent with a decrease in suppression with masker duration (Viemeister and Bacon, 1982; Carlyon, 1987; Strickland, 2004). Studies that have directly measured suppression using forward masking have shown that suppression decreases as suppressor onset precedes masker onset (Viemeister and Bacon, 1982; Thibodeau *et al.*, 1991; Champlin and Wright, 1993). Some physiological data (Kiang *et al.*, 1965; Arthur *et al.*, 1971) also show that suppression decreases over the first tens of milliseconds of the suppressor tone.

If suppression decreases with increasing suppressor duration, the question is where this change would be taking place. There is some evidence that suppression may depend on gain at the suppressor frequency as well as the signal frequency, at least for suppressor frequencies above the signal frequency. Physiological data show that exposing the frequency region above a signal frequency to a high-intensity tone (Robertson and Johnstone, 1981) or to kanamycin (Dallos *et al.*, 1980) decreases suppression without affecting the tuning curve at the signal frequency. Although it is not clear whether this is a within-channel or an across-channel effect, it seems worth exploring the hypothesis that a decrease in suppression may be caused by a decrease in gain at the suppressor frequency.

This may be illustrated using the input–output functions that were estimated in the Δf=0.0 condition for the signal place as an estimate of input–output functions at the masker/suppressor place. As in Strickland (2004), it was assumed that at the lowest signal levels in Fig. 5, suppression dominates because the masker does not yet produce enough activity at the signal place for excitatory masking. The suppression was assumed to come from frequencies above the signal frequency, because this region produces suppression at the lowest masker level (Cooper, 1996). It was also assumed that the suppressor level is determined by the gain in the masker region above the signal. This is an unorthodox assumption, but is supported by some psychophysical data (Bacon *et al.*, 1988) and the physiological evidence discussed earlier. It will be assumed that the precursor affects the input–output function in the masker region in the same way that it did in the signal region when Δf=0.0. If the gain in the masker region decreases, then the masker level must be increased to cause the same amount of suppression. In Fig. 6, plots are made predicting what may be happening at the *suppressor* place. The input–output functions are the same as those in Fig. 3, which were derived for various precursor levels from the data in Fig. 2. The small symbols at the bottom of each function are included so that the input–output functions may be paired with the legend indicating which precursor condition the fitted function represents. The large closed circles are the masker levels for the lowest signal level for each listener in Fig. 5 at which there were multiple data points and also corresponding input–output functions [35 (L3, L4), 40

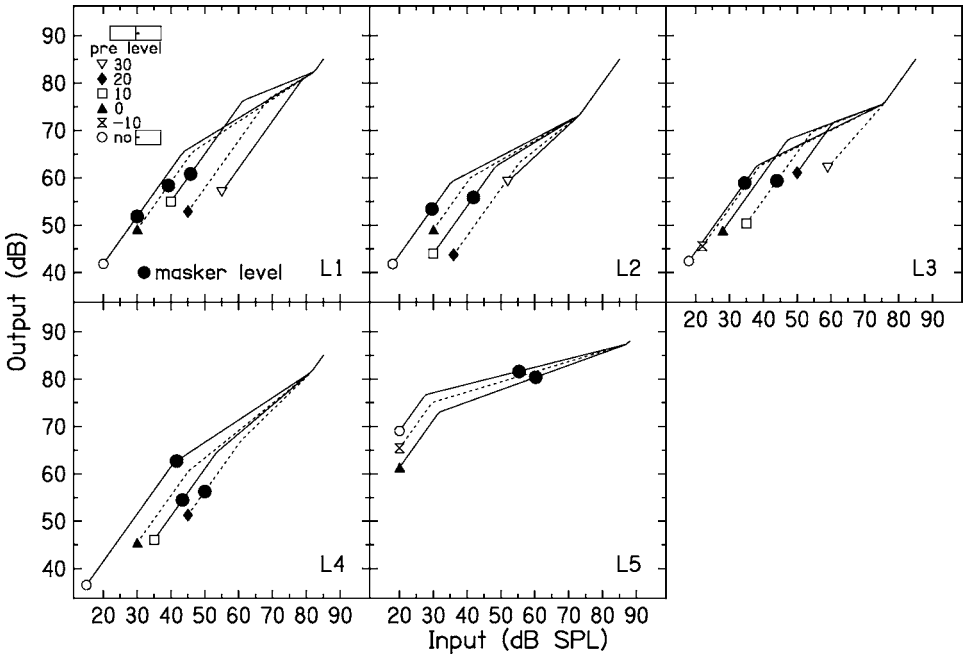


FIG. 6. The input–output functions from Fig. 3 are used as estimates of the input–output functions at the *suppressor* place. The symbols are masker values from Fig. 5 for signal levels of 35 (L3, L4), 40 (L1, L2), or 50 dB SPL (L5), the lowest signal levels for which multiple points could be plotted. In general, the output levels for the maskers are approximately equal, which would be consistent with the hypothesis that they are producing equal suppression of the signal.

(L1, L2), or 50 dB SPL (L5)]. The masker level from the no-precursor condition is plotted on the input–output function estimated from the no-precursor condition, etc. For three listeners (L2, L3, and L5) the suppressor levels at output are fairly constant (within listener), while for L1 this is true for two of the three suppressor levels. This would support the idea that in this condition, the temporal effect could be consistent with a decrease in gain at the masker (suppressor) frequency. For listener L4, the masker output levels are lower with a precursor than with no precursor. In this case, the masking may be excitatory. This listener showed forward masking from the precursor at a level of 10 dB SPL, which was not observed for the other listeners.

At higher signal levels, it will be assumed that the masker also causes excitatory masking of the signal, in addition to suppression. It will also be assumed that the input–output function for the masker at the signal frequency is linear. As the masker level is increased, the excitatory masking should increase at a faster rate than the suppressive masking (Yasin and Plack, 2007). The presence of the precursor would still decrease suppression, but the masker level would only need to be increased enough to increase the excitatory masking by that amount. Thus the slopes of the functions will reflect mainly excitatory masking, but the decrease in suppression is reflected in the vertical shift in the functions. In Fig. 7, thresholds for a signal delayed 202 ms from the onset of the masker are added to the data from Fig. 5, for comparison with other studies. The long-delay function is nearly identical to (L1, L2, and L5) or is parallel to (L3 and L4) the function measured with the highest precursor level.

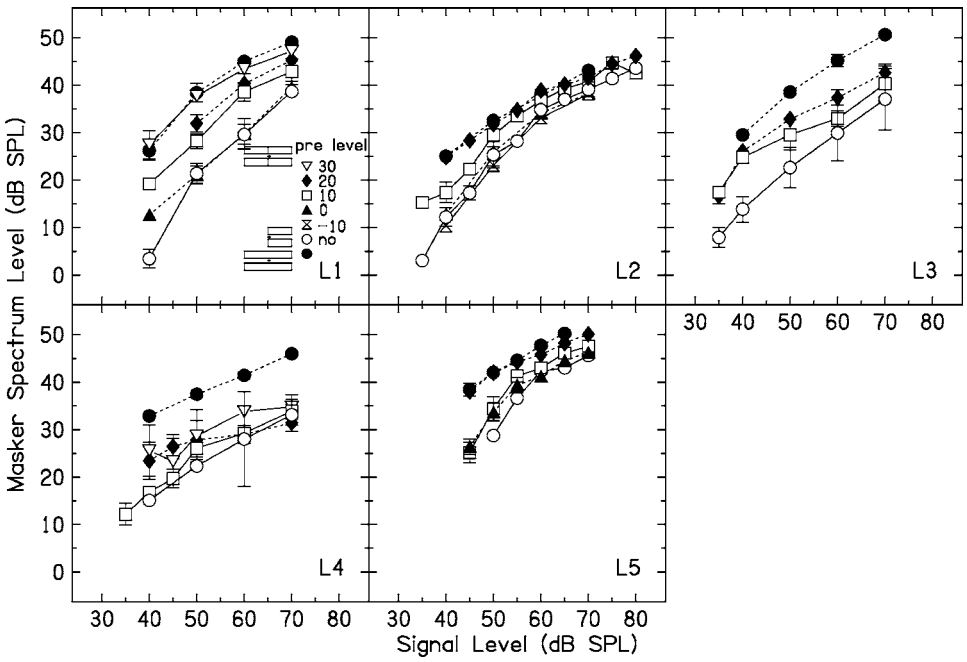


FIG. 7. Long-delay (closed circles) thresholds when the masker and precursor $\Delta f = 0.3$, along with data replotted from Fig. 5. The long-delay thresholds are nearly equal to (L1, L2, and L5) or parallel to (L3 and L4) the functions for the highest precursor level.

This supports the idea that above a certain masker level, excitatory masking dominates, so the shape of the function remains the same.

IV. DISCUSSION

The data shown in this study are consistent with previous results, showing a graded temporal effect with precursor level. In previous studies (Carlyon, 1987; Bacon and Smith, 1991), the effect of precursor level was measured for one fixed masker level. Bacon and Smith used a noise at a spectrum level of 20 dB for most of their listeners, and found that signal threshold for a 4-kHz signal decreased roughly 10 dB for each 10 dB increase in precursor level, up to a precursor spectrum level of 10 dB SPL. This is consistent with the results in the present study. In Fig. 2, at a masker spectrum level of 20 dB on the *y* axis, signal thresholds decrease as the precursor spectrum level increases up to roughly 10 dB. Above this level, for this masker level, the precursor itself may be producing masking. One listener in the Bacon and Smith (1991) study was tested with a masker spectrum level of 40 dB, and showed smaller level effects. In the present study L2 also shows smaller effects of precursor level at this higher masker level.

This study adds to previous results by quantifying the temporal effect in terms of changes in gain. The presence of a precursor decreased gain fairly linearly over a certain range. This suggests that the temporal effect acts like an automatic gain control, somewhat akin to the middle ear acoustic reflex. As the input increases, the maximum gain decreases. Previous studies have shown that thresholds were at the lowest signal-to-masker ratio when the precursor level was at or slightly below the masker level (Carlyon, 1987; Bacon and Smith, 1991). This is also true in the present study. In Fig. 2, the ratio of the signal level to the masker level at threshold is lowest when the precursor and the masker are at approximately equal levels. In natural environments, there would usually be continuous noise, not separate precursors and maskers. The results suggest that the auditory system may optimize the signal-to-noise ratio in this situation by adjusting the gain appropriately. The decrease in gain turns the output masker level down more than the signal, when the masker level is lower than the signal level as it was in the present conditions, improving the signal-to-noise ratio at the output of the cochlea. The long-delay results in Fig. 4 show that this adjustment results in a linear growth of masking.

The size of the temporal effect also depends on whether the masker and precursor are on-frequency or off-frequency relative to the signal. The results support the interpretation by Strickland (2004) that gain is decreased in the frequency region of the precursor. If this includes the signal frequency, the result is a decrease in gain at the signal frequency. If the precursor only contains energy at the suppressor frequency, the results are consistent with a decrease in gain at the suppressor frequency, leading to a decrease in suppression at the signal frequency.

Both of these results would be consistent with activation of the MOCR. Although there is a large body of research on the MOCR, its function is still not at all well understood. There is a small amount of behavioral data from animal studies showing efferent-mediated changes in response to sound. Smith *et al.* (2000) found that contralateral noise raised behavioral thresholds for tones in Japanese macaques, and that this effect disappeared with sectioning of the medial olivocochlear bundle. May and McQuone (1995) reported that cats showed a decrease in the ability to detect intensity changes in 8-kHz tones presented in ipsilateral noise after sectioning of the medial olivocochlear bundle. A recent comprehensive review article by Guinan (2006) mentioned almost no behavioral data on the action of the MOCR in humans.

The present study provides behavioral data in humans that would be consistent with the action of the MOCR. An examination of Fig. 2 provides evidence for what the role of the temporal effect might be when the masker and precursor are on-frequency. The precursor may be thought of as ongoing background noise. As the precursor level is increased, for a range of signal levels, the signal is audible in higher and higher levels of masking noise. This is consistent with a decrease in gain within the signal channel. Thus, decreasing the gain in response to background noise may optimize audibility. Likewise, Fig. 5 shows what happens when the interfering sound is probably partly suppressing the signal. As the level of the background sound is increased, the signal is audible in increasing levels of noise. Thus the temporal effect may be evidence that the auditory system decreases its response to ongoing stimuli, which optimizes the response to changing stimuli.

ACKNOWLEDGMENT

This research was partially supported by a grant from NIH (No. R03 DC03510).

[1]The onset and offset times were intended to be 5 ms for all the stimuli. The TDT function sets onset and offset times from 10% to 90% of maximum. For a nominal rise/fall time of 5 ms, the rise/fall time from 0 to 100% of maximum is approximately 8.5 ms. The signal had no steady state, so the rise/fall times were 5 ms, but the level of the signal was affected because the middle was affected by the onset and the offset ramp. The reported signal levels are approximately 7.6 dB below the nominal signal levels. The gating was still close to $\cos^2$.

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). "Properties of 'two-tone inhibition' in primary auditory neurons," J. Physiol. (London) **212**, 593–609.

Backus, B. C., and Guinan, J. J., Jr. (**2006**). "Time-course of the human medial olivocochlear reflex," J. Acoust. Soc. Am. **119**, 2889–2904.

Bacon, S. P. (**1990**). "Effect of masker level on overshoot," J. Acoust. Soc. Am. **88**, 698–702.

Bacon, S. P., Hedrick, M. S., and Grantham, D. W. (**1988**). "Temporal effects in simultaneous pure-tone masking in subjects with high-frequency sensorineural hearing loss," Audiology **27**, 313–323.

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

Bacon, S. P., and Takahashi, G. A. (**1992**). "Overshoot in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **91**, 2865–2871.

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

Baker, R. J., Rosen, S., and Darling, A. M. (**1998**). "An efficient characterization of human auditory filtering across level and frequency that is also physiologically reasonable," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield,

and R. Meddis (Whurr, London).
Carlyon, R. P. (**1987**). "A release from masking by continuous, random, notched noise," J. Acoust. Soc. Am. **81**, 418–426.
Carlyon, R. P. (**1989**). "Changes in the masked thresholds of brief tones produced by prior bursts of noise," Hear. Res. **41**, 223–235.
Champlin, C. A., and McFadden, D. (**1989**). "Reductions in overshoot following intense sound exposures," J. Acoust. Soc. Am. **85**, 2005–2011.
Champlin, C. A., and Wright, B. A. (**1993**). "Manipulations of the duration and relative onsets of two-tone forward maskers," J. Acoust. Soc. Am. **94**, 1269–1274.
Cooper, N. P. (**1996**). "Two-tone suppression in cochlear mechanics," J. Acoust. Soc. Am. **99**, 3087–3098.
Dallos, P., Harris, D. M., Relkin, E., and Cheatham, M. A. (**1980**). "Two-tone suppression and intermodulation distortions in the cochlea: Effect of outer hair cell lesion," in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. Van den Brink and F. A. Bilsen (Delft University Press, Delft, The Netherlands), pp. 242–249.
Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.
Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.
Hicks, M. L., and Bacon, S. P. (**1992**). "Factors influencing temporal effects with notched-noise maskers," Hear. Res. **64**, 123–132.
Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). "Discharge patterns of single fibers in the cat's auditory nerve," *Monograph No. 35* (MIT, Cambridge, MA).
Kimberley, B. P., Nelson, D. A., and Bacon, S. P. (**1989**). "Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1660–1665.
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.
May, B. J., and McQuone, S. J. (**1995**). "Effects of bilateral olivocochlear lesions in pure-tone intensity discrimination in cats," Aud. Neurosci. **1**, 385–4000.
McFadden, D., and Champlin, C. A. (**1990**). "Reductions in overshoot during aspirin use," J. Acoust. Soc. Am. **87**, 2634–2642.
Oxenham, A. J., and Bacon, S. P. (**2004**). "Psychophysical manifestations ofcompression: Normal-hearing listeners," in *Compression*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 62–106.
Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.
Robertson, D., and Johnstone, B. M. (**1981**). "Primary auditory neurons: Nonlinear responses altered without changes in sharp tuning," J. Acoust. Soc. Am. **69**, 1096–1098.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
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.
Schmidt, S., and Zwicker, E. (**1991**). "The effect of masker spectral asymmetry on overshoot in simultaneous masking," J. Acoust. Soc. Am. **89**, 1324–1330.
Scholl, H. (**1962**). "Das dynamische verhalten des gehörs bei der unterteilung des schallspektrums in frequenzgruppen" ("The dynamic performance of the hearing system when separating the sound spectrum into critical bands"), Acustica **12**, 101–107.
Smith, D. W., Turner, D. A., and Hensen, M. M. (**2000**). "Psychophysical correlates of contralateral efferent suppression. I. The role of the medial olivocochlear system in 'central masking' in nonhuman primates," J. Acoust. Soc. Am. **107**, 933–941.
Strickland, E. A. (**2001**). "The relationship between frequency selectivity and overshoot," J. Acoust. Soc. Am. **109**, 2063–2073.
Strickland, E. A. (**2004**). "The temporal effect with notched-noise maskers: Analysis interms of input-output functions," J. Acoust. Soc. Am. **115**, 2234–2245.
Strickland, E. A., and Krishnan, L. A. (**2005**). "The temporal effect in listeners with mild to moderate cochlear hearing impairment," J. Acoust. Soc. Am. **118**, 3211–3217.
Thibodeau, L. M., Champlin, C. A., and Stiritz, L. (**1991**). "Suppressor duration effects in forward masking," J. Acoust. Soc. Am. **90**, 2268(A).
Viemeister, N. F., and Bacon, S. P. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**, 1502–1507.
von Klitzing, R., and Kohlrausch, A. (**1994**). "Effect of masker level on overshoot in running- and frozen-noise maskers," J. Acoust. Soc. Am. **95**, 2192–2201.
Warr, W. B. (**1980**). "Efferent components of the auditory system," Ann. Otol. Rhinol. Laryngol. **90**, 114–190.
Warr, W. B., and Guinan, J. J., Jr. (**1979**). "Efferent innervation of the organ of Corti: Two separate systems," Brain Res. **173**, 152–155.
Yasin, I., and Plack, C. J. (**2003**). "The effects of a high-frequency suppressor on tuning curves and derived basilar-membrane response functions," J. Acoust. Soc. Am. **114**, 322–332.
Yasin, I., and Plack, C. J., (**2007**). "The effects of low- and high-frequency suppressors on psychophysical estimates of basilar-membrane compression and gain," J. Acoust. Soc. Am. **121**, 2832–2841.
Zwicker, E. (**1965**). "Temporal effects in simultaneous masking by white-noise bursts," J. Acoust. Soc. Am. **37**, 653–656.

The effect of hearing impairment on the identification of speech that is modulated synchronously or asynchronously across frequency

Joseph W. Hall III,[a] Emily Buss, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 22 June 2007; accepted 11 November 2007)

This study investigated the effect of mild-to-moderate sensorineural hearing loss on the ability to identify speech in noise for vowel-consonant-vowel tokens that were either unprocessed, amplitude modulated synchronously across frequency, or amplitude modulated asynchronously across frequency. One goal of the study was to determine whether hearing-impaired listeners have a particular deficit in the ability to integrate asynchronous spectral information in the perception of speech. Speech tokens were presented at a high, fixed sound level and the level of a speech-shaped noise was changed adaptively to estimate the masked speech identification threshold. The performance of the hearing-impaired listeners was generally worse than that of the normal-hearing listeners, but the impaired listeners showed particularly poor performance in the synchronous modulation condition. This finding suggests that integration of asynchronous spectral information does not pose a particular difficulty for hearing-impaired listeners with mild/moderate hearing losses. Results are discussed in terms of common mechanisms that might account for poor speech identification performance of hearing-impaired listeners when either the masking noise or the speech is synchronously modulated.
 [DOI: 10.1121/1.2821967]

PACS number(s): 43.66.Dc, 43.66.Mk, 43.66.Sr, 43.71.Ky [RYL] Pages: 955–962

I. INTRODUCTION

The purpose of this study was to gain insight into the factors affecting the ability of listeners with sensorineural hearing impairment to understand speech in the presence of masking noise. It is widely recognized that hearing-impaired listeners often show particular difficulty in understanding speech at poor signal-to-noise ratios (e.g., Plomp and Mimpen, 1979; Festen and Plomp, 1990; Peters *et al.*, 1998). The present study focused on the ability of hearing-impaired listeners to integrate asynchronous, spectrally distributed speech information. We have noted that such an ability may be of benefit in processing signals in fluctuating noise backgrounds, where the signal-to-noise ratio varies dynamically with respect to both frequency and time (Buss *et al.*, 2003). For example, in a given temporal epoch the signal-to-noise ratio may be favorable at Frequency A but not at Frequency B, but in a successive epoch the signal-to-noise ratio may be favorable at Frequency B but not Frequency A. In such cases, a listener may construct an auditory target from asynchronous information arising from different spectral regions.

Previous studies of speech perception in normal-hearing listeners have demonstrated evidence for the ability to combine spectrally distributed, asynchronous speech information. An innovative "checkerboard speech" study by Howard-Jones and Rosen (1993) was perhaps the first study to test this ability rigorously. That study assessed the perception of consonants in a masker composed of multiple, contiguous noise bands that were amplitude modulated. In one condition, modulation of neighboring bands was out of phase, such that when odd-numbered bands were gated on, even numbered bands were gated off, and vice versa; in this condition, masking noise formed a "checkerboard" pattern when displayed as a spectrogram. Speech identification thresholds in the checkerboard conditions were sometimes better than those obtained in conditions where only the odd-numbered masker bands or only the even numbered masker bands were present, and this was interpreted as reflecting the combination of asynchronous cues for speech identification. Evidence for the combination of asynchronous, spectrally distributed speech information has also been obtained in studies where speech is amplitude modulated in such a way that the availability of simultaneous across-frequency consonant information is limited (Buss *et al.*, 2003; 2004a). Carlyon *et al.* (2002) also showed that vowels can be identified when two formants are presented such that they do not overlap in time. Overall, these results suggest a robust ability to integrate spectrally distributed, asynchronous speech cues and are consistent with the notion that successful hearing in noise may involve the processing of successive "glimpses" of spectrally and temporally fragmented target sounds (e.g., Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Assmann and Summerfield, 2004; Buss *et al.*, 2004a; Cooke, 2006). The particular approach used here follows that of Buss *et al.* (2004a) where consonant identification in a vowel-consonant-vowel (VCV) context is determined under two different conditions of modulation. In this approach, speech

[a]author to whom correspondences should be addressed. Electronic mail: jwh@med.unc.edu

TABLE I. Audiometric data for hearing-impaired listeners (HI1-HI7). The ear tested is associated with the entries in bold. The %Sp column refers to speech recognition for monosyllabic words in quiet with presentation level approximately 30 dB above the speech recognition threshold for spondee words. "NR" indicates that the threshold was above the limits of audiometric testing and "DNT" indicates that the test was not performed.

	Left							Right						
	0.25	0.5	1.0	2.0	4.0	8.0	%Sp	0.25	0.5	1.0	2.0	4.0	8.0	%Sp
HI1	**40**	**45**	**45**	**35**	**45**	**85**	**84**	NR	NR	NR	NR	NR	NR	DNT
HI2	50	55	60	60	60	65	92	**40**	**35**	**45**	**45**	**45**	**45**	**88**
HI3	**25**	**30**	**50**	**50**	**50**	**35**	**92**	25	30	40	50	50	40	92
HI4	20	20	20	20	20	15	100	**50**	**55**	**55**	**40**	**40**	**40**	**92**
HI5	35	35	25	30	40	60	100	**25**	**30**	**30**	**30**	**45**	**65**	**100**
HI6	**25**	**40**	**45**	**30**	**55**	**85**	**88**	15	15	10	5	5	10	100
HI7	**25**	**50**	**60**	**45**	**35**	**50**	**88**	10	40	45	45	45	35	96

tokens are filtered into a number of contiguous, log-spaced frequency bands, and the bands are then amplitude modulated such that the pattern of modulation is either in phase across bands (synchronous modulation) or 180° out of phase for adjacent bands (asynchronous modulation).

Although the effect of hearing loss on the ability to integrate asynchronous speech information is largely unknown, a study by Healy and Bacon (2002) provides at least a suggestion that this capacity might be reduced in listeners with sensorineural hearing impairment. These investigators employed a speech perception method similar to that used in several recent studies where speech is divided into a number of frequency bands and the envelope of each band is used to modulate a corresponding frequency region of either a tonal or noise carrier (e.g., Shannon *et al.*, 1995; Turner *et al.*, 1995; Dorman *et al.*, 1997; Turner *et al.*, 1999). Such methods have demonstrated that good speech perception can occur on the basis of envelope fluctuations imposed upon the carrier(s). The study by Healy and Bacon (2002) employed two widely separated tonal carriers. The individual, modulated carriers did not result in recognizable speech, but the pair of modulated carriers did. Of greater potential relevance here was the finding that small temporal delays imposed between the modulated carriers adversely affected performance to different degrees in listeners with normal hearing and hearing impairment. Listeners with normal hearing were able to maintain relatively good performance for temporal delays of 12.5–25 ms between stimuli in the two spectral regions, a finding that is in agreement with previous reports (e.g., Greenberg and Arai, 1998). In contrast, most of the hearing-impaired listeners of Healy and Bacon showed steep declines in performance for such delays. One possible interpretation of this result is that hearing-impaired listeners have a reduced ability to recognize speech on the basis of information that is asynchronous across frequency. A potential practical consequence of this finding is related to signal processing associated with some digital hearing aid strategies. For example, Stone and Moore (2003) noted that digital hearing aid processing that attempts to mimic the frequency selectivity of the normal ear can result in speech delays that vary across frequency. It is possible that such effects could interact with hearing loss to produce undesirable consequences on speech intelligibility.

As noted above, the results of Healy and Bacon (2002) suggest that hearing-impaired listeners may have a reduced ability to integrate asynchronous speech information when compared with normal-hearing listeners under conditions where different speech frequency regions are delayed with respect to each other. Under such circumstances, the temporal relationship of information across frequency is disrupted. In contrast, when speech is presented in a masker with spectro-temporal modulations, the available speech cues may be sparse but the temporal relationship between those cues is unchanged. The cues presented in such a masker may be simulated using the *band-AM* paradigm, where the speech stimulus is filtered into contiguous narrow bands which are then amplitude modulated independently (Buss *et al.*, 2004a). When amplitude modulation is out of phase across neighboring bands, there are asynchronous cues, and the temporal relationship between those cues that remain is natural and unaffected by stimulus processing. At the outset of this study it was hypothesized that the poorer ability to integrate asynchronous speech information demonstrated by Healy and Bacon (2002) using the *delayed-band* paradigm may reflect a more general reduction in the capacity of hearing-impaired listeners to combine asynchronous, across-frequency information, like that associated with the *band-AM* paradigm.

II. METHODS

A. Listeners

Listeners with normal hearing and listeners with sensorineural hearing loss participated. There were eight normal-hearing listeners, two male and six female, ranging in age from 24 to 55 years (with a mean age of 36.5 years and standard deviation of 11.7 years). These listeners had no history of hearing problems, and had pure-tone thresholds of 15 dB hearing leavel or better at octave frequencies from 250 Hz to 8000 Hz (ANSI, 2004). The hearing-impaired listeners had mild to moderate sensorineural hearing losses that were relatively flat in configuration (see Table I). There were seven listeners in this group, two male and five female, ranging in age from 24 to 55 years (with a mean age of 46.2 years and standard deviation of 5.4 years). All listeners were paid for their participation.

B. Stimuli

The stimulus processing was very similar to that used by Buss *et al.* (2004a) in a study of normal-hearing listeners. The stimuli were 12 vowel-consonant-vowel (VCV) tokens spoken by an American English speaking female. There were five separate samples of each of the 12 tokens, and each token was in the form /a/ C /a/ (e.g., /aka/). The consonants were b, d, f, g, k, m, n, p, s, t, v, and z. The VCV samples were 528–664 ms in duration with a mean duration of 608 ms. The speech tokens were scaled to have equal total root mean square level, and the presentation level of the tokens prior to modulation was approximately 75 dB sound pressure level (SPL). The tokens were digitally filtered into 4 or 16 contiguous bands, with edge frequencies logarithmically spaced from 0.1 to 10 kHz. Our previous research (Buss *et al.*, 2004a) indicated that listeners with normal hearing showed evidence of an ability to integrate asynchronous spectral information in the perception of speech for 2, 4, 8, or 16 contiguous bands. The present choice of 4 and 16 bands was motivated by a desire to determine whether listeners with hearing impairment also showed an ability to integrate asynchronous spectral information in the perception of speech over a similarly wide range of spectral bands. The filtered speech tokens were saved to disk in separate files containing either the odd-numbered bands or the even-numbered bands. Modulation was accomplished by multiplying the speech tokens by a 20 Hz raised square wave having a 50% duty cycle. The abrupt duty cycle transitions were replaced by 5 ms $\cos^2$ ramps.

In one condition, unprocessed speech tokens were presented. This condition was intended to allow an assessment of group differences related to the identification of unprocessed speech in noise. All other conditions were associated with modulated speech. In *just-odd* conditions, only the odd-numbered bands were included and modulation began with the "on" half of the duty cycle. In the *just-even* conditions, only the even-numbered bands were included and modulation began with the "off" half of the duty cycle. The *asynchronous* conditions were formed by the combination of the *just-odd* and *just-even* stimuli as defined above. For the *synchronous* condition, both the odd and even numbered bands were included; modulation was in phase across all bands, and the modulation began with the on half of the duty cycle. Previous findings (Buss *et al.*, 2004a) showed that the starting phase of the modulation cycle (on versus off) did not affect the outcome for these stimuli when modulation was synchronous across bands. Note that the number of bands (4 or 16) did not matter for the synchronous conditions, as the contiguous bands formed identical stimuli when bands were modulated synchronously.

The masker was a noise with a spectral shape matching the long-term spectrum of the speech stimuli. The level of this masker was adjusted adaptively, and played continuously over the duration of a threshold track. Stimuli were presented monaurally via a Sennheiser HD265 headphone. For normal-hearing listeners, the stimuli were presented to the left ear. For hearing-impaired listeners, the ear tested was the poor ear in the one case of a unilateral hearing loss, the better ear in the one case where hearing thresholds were beyond audiometric limits in one ear, and was assigned randomly in cases of bilateral hearing loss. A continuous, 50 dB SPL speech-shaped noise was presented to the contralateral ear of all listeners in order to mask possible crossover to the contralateral ear.

1. Procedure

In stage 1 of the procedure, the threshold level for detecting the speech-shaped masking noise was obtained. This threshold was ascertained because the main procedure determined the level of the speech-shaped noise that just masked the identification of the speech token. It was therefore important to know whether the masker level that just masked the speech was above the detection threshold of the masker, in order to identify possible floor effects. An observation interval was marked visually, and the listener pressed a button to report whether or not the sound had been detected. Based on the listener's response, the experimenter raised or lowered the noise level in steps of 2 dB and bracketed the threshold level based upon the yes/no responses of the listener. In stage 2 of the procedure, listeners were presented with the samples of the unprocessed and modulated VCVs presented in quiet. The purpose of this stage was to give the listeners a general familiarity with the speech material on which they would be tested.

In stage 3 of the procedure, computer-controlled threshold runs were obtained to determine masked VCV identification thresholds. During trials on these runs, listeners were presented with a randomly selected VCV token. The listeners were then visually presented with the 12 possible consonants and asked to enter a response via the keyboard. No feedback regarding the correct response was provided. The level of the masker was adjusted using a 1-up, 1-down adaptive rule, estimating the masker level necessary to obtain 50% correct identification. There were 26 total reversals per threshold run. The first two reversals were made with 3 dB steps, and the final 24 reversals were made in steps of 2 dB. The threshold estimate was taken as the mean masker level at the last 24 track reversals. Three to five threshold runs were obtained and averaged for each condition. If, during the course of a threshold run, a total of six reversal values occurred at or below the masker threshold (as determined in stage 1), it was assumed that the threshold was influenced by a floor effect, and the masked VCV identification threshold was considered to be unmeasurable.

III. RESULTS

Figure 1 provides a summary of the results of the conditions where the speech was unprocessed, modulated synchronously, and modulated asynchronously. All of the normal-hearing and hearing-impaired listeners obtained measurable thresholds in these conditions. Table II shows individual data for these conditions and also for the conditions where only odd or even bands were present. The latter data will be considered last because there were many cases where the thresholds of the hearing-impaired listeners were unmeasurable in these conditions. In interpreting the masked

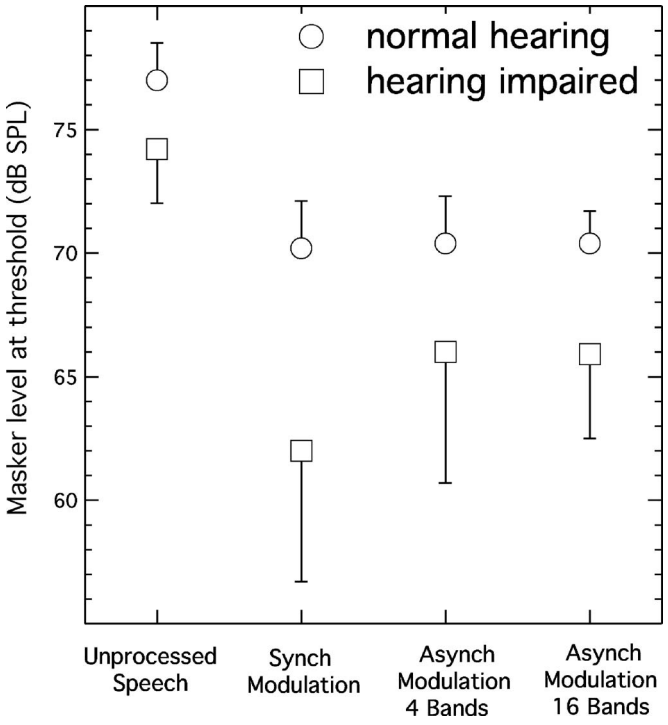


FIG. 1. Masker level at threshold for the unprocessed speech condition, the synchronous modulation condition, the asynchronous modulation/4-band condition, and the asynchronous modulation/16-band condition. The error bars show +1 standard deviation for the normal-hearing listeners and −1 standard deviation for the hearing-impaired listeners.

threshold values from the experimental conditions, it should be kept in mind that higher values represent better performance, because threshold represents the masker level associated with 50% correct identification of a fixed-level speech signal.

A. Group differences for unprocessed speech

In agreement with the findings of several previous studies (Plomp and Mimpen, 1979; Dubno *et al.*, 1984; Glasberg and Moore, 1989; Turner *et al.*, 1992; Peters *et al.*, 1998), the present results for the unprocessed speech indicated that the hearing-impaired listeners required a higher signal-to-noise ratio than the normal-hearing listeners at the masked speech identification threshold. As shown in the points to the left of Fig. 1, the mean masker level at threshold was 77.0 dB SPL for the normal-hearing listeners and 74.3 dB SPL for the hearing-impaired listeners. An independent samples t test indicated that this difference was statistically significant ($t_{13}=2.8$; $p=0.01$).

B. Modulated speech conditions

The results of central interest in the present study concerned differences in performance between the synchronously modulated and asynchronously modulated speech conditions. Our previous data for normal-hearing listeners and a 20 Hz rate of modulation indicated a relatively good ability to integrate asynchronous, across-frequency speech information, with masked thresholds being approximately the same for conditions of synchronous and asynchronous modulation (Buss *et al.*, 2004a). The present results are in agreement with this: the normal-hearing listeners had an average threshold of 70.4 dB SPL for both the 4-band and 16-band asynchronous conditions, which compares closely to the average threshold of 70.2 dB SPL for the synchronous modulation condition.

Inspection of Fig. 1 suggests generally poorer performance by the hearing-impaired listeners in the modulated speech conditions, with particularly poor performance for the

TABLE II. Masker level at threshold (dB SPL) for the various experimental conditions is shown for the normal-hearing (N1–N8) and hearing-impaired (HI1–HI7) listeners. The numbers in parentheses are the inter-listener standard deviations. Dashes indicate thresholds that were unmeasurable. Synchronous and Asynchronous modulation conditions are labeled "Synch" and "Asynch," respectively.

			4 Bands			16 Bands		
	Unprocessed	Synch	Asynch	Just-Odd	Just-Even	Asynch	Just-Odd	Just-Even
N1	78.9	72.6	73.1	68.3	65.2	72.5	65.3	67.5
N2	79.6	73.4	71.6	59.0	37.9	71.6	61.2	68.5
N3	77.0	70.1	72.1	47.2	55.2	70.5	58.0	63.7
N4	75.8	68.5	67.8	30.7	56.5	69.7	56.0	60.0
N5	76.0	69.4	68.1	35.0	46.9	68.8	55.0	58.3
N6	76.6	69.2	69.9	39.5	45.7	70.1	56.8	63.8
N7	76.4	70.2	69.3	38.9	55.5	68.8	62.1	60.7
N8	75.5	67.8	71.5	47.9	59.6	70.8	53.8	65.9
Mean	77.0	70.2	70.4	45.8	52.8	70.4	58.5	63.6
	(1.5)	(1.9)	(1.9)	(12.6)	(8.7)	(1.3)	(4.0)	(3.7)
HI1	72.2	62.8	58.2	—	—	63.0	—	47.1
HI2	76.6	68.4	69.7	—	—	70.1	63.1	—
HI3	75.5	57.0	60.9	—	56.5	64.3	50.0	—
HI4	76.7	70.3	73.3	57.1	—	71.0	62.2	66.1
HI5	74.9	59.0	68.8	—	—	66.8	—	—
HI6	72.3	58.6	67.7	48.6	—	63.1	48.8	46.0
HI7	71.6	58.1	63.5	—	56.0	63.3	—	—
Mean	74.3	62.0	66.0			65.9		
	(2.2)	(5.3)	(5.3)			(3.4)		

synchronous modulation condition. In order to evaluate this impression, a repeated measures analysis of variance was performed on the modulated speech data with a within-subjects factor of condition (synchronous modulation, 4-band asynchronous modulation, and 16-band asynchronous modulation) and a between-subjects grouping factor of hearing impairment. Results of this analysis indicated a significant effect of condition ($F_{2,26}=5.4$; $p=0.01$), a significant effect of hearing impairment ($F_{1,13}=7100$; $p<0.001$), and a significant interaction between condition and hearing impairment ($F_{2,26}=4.3$; $p=0.02$). The interaction reflects the finding that whereas the normal-hearing listeners showed comparable performance across the asynchronous and synchronous modulated speech conditions, the hearing-impaired listeners performed relatively poorly in the synchronous modulation condition. A further indication of the poor performance of the hearing-impaired listeners in the synchronously modulated speech condition is the fact that the hearing-impaired were, on average, 8.2 dB worse than normal in this condition, compared to only 2.7 dB worse than normal in the unprocessed speech condition.

C. Performance in just-odd and just-even conditions

As can be seen in Table II, many of the thresholds of the hearing-impaired listeners were at floor (unmeasurable) in the conditions where just even numbered or just odd numbered bands were present. Statistical analyses were therefore not performed for these thresholds. The available data are nevertheless informative. For the normal-hearing listeners, where no thresholds were unmeasurable, it was always the case that the thresholds for the asynchronous conditions were better than for either the *just-odd* or *just-even* conditions. This finding is consistent with the interpretation that the threshold in the asynchronous condition did not simply reflect the better of the *just-odd* or *just-even* conditions, but instead reflected an integration of the asynchronously presented information; a previous study, where data were collected on individual speech tokens, also supported this interpretation (Buss *et al.*, 2004a). A similar pattern of results was found for the hearing-impaired listeners, where the thresholds for the asynchronous conditions were better than the thresholds associated with the *just-odd* or *just-even* conditions (with a number of the *just-odd* or *just-even* thresholds being unmeasurable).

IV. GENERAL DISCUSSION

The discussion begins with the central question of the study, whether hearing-impaired listeners appear to have a deficit in the ability to utilize spectrally distributed, asynchronous cues for speech identification. Possible accounts are then considered for the effects of hearing loss on the processing of synchronously and asynchronously modulated speech signals.

A. Performance for synchronous versus asynchronous modulation

The primary motivation for the present study was to determine whether sensorineural hearing impairment might be associated with a reduced ability to integrate asynchronous, spectrally distributed speech information. Because this ability may aid the processing of speech at poor signal-to-noise ratios, such a reduced capacity might help to account for the commonly reported difficulties of hearing-impaired listeners understanding speech in noisy backgrounds. Rather than indicating that hearing-impaired listeners have particular difficulty processing asynchronous speech information, the present results indicated that many hearing-impaired listeners actually performed better for asynchronous modulation of spectrally distributed speech information than for synchronously modulated speech. As noted in the introduction, a possible interpretation of the results of Healy and Bacon (2002) is that hearing-impaired listeners have a deficit in the ability to integrate asynchronous speech information. In that study listeners with sensorineural hearing loss were more adversely affected than normal-hearing listeners by small temporal asynchronies between spectrally separated bands of speech information. The fact that the present study found that hearing-impaired listeners performed better for sparse cues that are distributed over time than for cues that are clustered in time (asynchronous vs synchronous AM) does not constitute a conflict with the findings or conclusions of Healy and Bacon, however, these two studies used very different means of presenting speech information and of manipulating asynchrony. The present results suggest that mild/moderate sensorineural hearing loss is not associated with a general deficit in the ability to integrate spectrally distributed, asynchronous speech information. Before considering the results of the asynchronous modulation conditions further, we will discuss the possible significance of the poor performance of the hearing-impaired listeners in the synchronous modulation condition.

B. Performance of hearing-impaired listeners for modulated speech and its possible relation to performance in modulated noise

As noted above, the average difference in performance between the normal-hearing listeners and the hearing-impaired listeners was relatively large in the synchronous modulation condition. This finding is noteworthy because of its possible relation to previous results from studies where the masking noise was modulated instead of the speech signal (e.g., Wilson and Carhart, 1969; Festen and Plomp, 1990; Takahashi and Bacon, 1992; Eisenberg *et al.*, 1995; Peters *et al.*, 1998; George *et al.*, 2006). A common outcome in such studies is that performance deficits for speech identification by hearing-impaired listeners are greater in modulated noise than in unmodulated noise. Although reduced audibility may sometimes contribute to this effect, the effect persists when the audibility factor is controlled (e.g., Eisenberg *et al.*, 1995; George *et al.*, 2006). Perhaps the most obvious factor that might underlie poor speech identification performance in modulated noise by hearing-impaired listeners is reduced temporal resolution (e.g., greater than normal forward masking), with a resulting reduction in the ability to take advantage of the good signal-to-noise ratios associated with masker envelope minima (e.g., Zwicker and Schorn, 1982; Festen and Plomp, 1990). The present finding that hearing-

impaired listeners also show relatively poor performance when the speech is synchronously modulated instead of the masker suggests that factors in addition to temporal resolution may play an important role in the relatively poor speech identification performance obtained by hearing-impaired listeners in modulated noise. This follows because, in the present modulated speech masking conditions, it would seem reasonable to attribute most of the masking to the speech-shaped masker energy that occurs *simultaneously* with the speech signal rather than to forward masking.

As noted above, poor speech identification performance of hearing-impaired listeners in modulated noise persists when the audibility factor is controlled (e.g., Eisenberg *et al.*, 1995; George *et al.*, 2006). In the present paradigm, a single speech level was employed and it is therefore difficult to draw firm conclusions about possible effects related to audibility. However, there are bases for speculating that audibility was not a major factor contributing to the difference in results between the synchronous and asynchronous modulation conditions obtained here for the hearing-impaired listeners. First, because the speech level was the same in the synchronous and asynchronous modulation conditions, it seems reasonable to infer that the difference in results between these conditions depended upon factors other than audibility. Second, if audibility contributed strongly to the pattern of results in the hearing-impaired listeners, significant correlations might be expected between audiometric thresholds and the speech identification threshold in synchronously modulated speech (which was relatively poor in the hearing-impaired group), and/or between audiometric thresholds and the difference in speech identification thresholds for synchronously versus asynchronously modulated speech (which was relatively large in the hearing-impaired group). To examine this question, the correlation was determined between these speech measures and the audiometric thresholds averaged over 500, 1000, 2000, and 4000 Hz. The correlations with audiometric threshold did not approach significance for either the speech identification threshold for synchronously modulated speech ($r=0.24$; $p=0.61$), or for the difference in speech identification thresholds for synchronously versus asynchronously modulated speech ($r=0.34$; $p=0.46$).

The fact that hearing-impaired listeners show a marked deficit in speech identification performance if either the masking noise *or* the speech is modulated invites speculation about a processing mechanism that may be common to both types of modulation. Two related (and not mutually exclusive) possibilities are considered below.[1]

1. Interaction between hearing impairment and speech redundancy

One possibility that could account for poor speech perception in hearing-impaired listeners when either the masking noise or the speech is synchronously modulated involves an interaction between hearing impairment and a reduction of speech redundancy. One feature that is shared whether the noise or the speech is modulated is that some of the listening epochs are systematically corrupted (by periodically increasing the masker level or decreasing the speech level). It is possible that the reduction of speech cue redundancy resulting from noise or speech modulation interacts with the degradation of speech cues associated with sensorineural hearing loss (Plomp, 1978) to produce a more substantial deterioration than occurs for unprocessed speech in unmodulated masking noise. This account is similar to that developed by Baer and Moore (1994) who spectrally "smeared" speech material presented to normal-hearing listeners in order to simulate the effects of reduced frequency selectivity that is common in listeners with sensorineural hearing impairment. Using this technique, both Baer and Moore (1994) and ter Keurs *et al.* (1993) have reported that spectral smearing has a greater deleterious effect for a fluctuating masker than for a steady masker. Baer and Moore reasoned that although the fluctuating masker was associated with sporadic good listening epochs (masker envelope minima), the masker energy present during envelope maxima tended to reduce the redundancy of the speech signal and interacted with the spectral smearing manipulation to cause a relatively large performance deficit. The results of the present study, where speech was modulated, and past studies, where noise was modulated, are consistent with an interpretation that corruption of speech information associated with modulation of either the speech or the noise can interact with the further corruption of speech information that results from hearing impairment to cause such a performance deficit. Although psychoacoustic abilities were not evaluated in the present hearing-impaired listeners, one likely way in which the coding of the auditory stimulus could have been corrupted is reduced frequency selectivity (e.g., Tyler *et al.*, 1984; Stelmachowicz *et al.*, 1985; Leek and Summers, 1996), analogous to the spectral smearing manipulation that was used in the Baer and Moore study.

The above reasoning is compatible with the present finding that the results of the hearing-impaired listeners were generally less abnormal for the asynchronously modulated speech. For example, it is possible that the spectral discontinuities associated with asynchronous speech modulation may have had the beneficial consequence of reducing the potential for interactions among widely separated spectral components of speech. Such interactions would be minimal in the normal ear, which is highly frequency selective, but could occur in listeners with sensorineural hearing loss. This account has some similarity to previous research on the question of whether signal preprocessing can aid speech identification in hearing-impaired listeners. Several investigators have noted that it is theoretically possible to ameliorate effects related to poor frequency selectivity via forms of speech signal processing intended to sharpen the spectrum of the speech signal (e.g., Summerfield *et al.*, 1985; Simpson *et al.*, 1990; Baer *et al.*, 1993; Miller *et al.*, 1999). Although such approaches have not resulted in large improvements in the speech identification abilities of hearing-impaired listeners, they have met with modest success (e.g., Baer *et al.*, 1993). The asynchronous modulation of speech investigated here, with its reduced potential for interaction among widely spaced spectral components, might be regarded as a special case of "spectral sharpening" when compared with the synchronous condition.

2. Temporal envelope versus temporal fine structure cues

A second, related, possibility that could account for poor performance of hearing-impaired listeners when either the masker or the speech stimulus is synchronously modulated concerns the relative utility of temporal envelope versus temporal fine structure cues for speech perception. Although both temporal envelope and temporal fine structure cues contribute importantly to normal speech perception (Rosen, 1992), it has been hypothesized that the ability to process speech envelope cues in modulated noise may be limited, due to the modulations of the masker interfering with the ability of the listener to process the envelope modulations of the speech stimulus (e.g., Nelson *et al.*, 2003; Lorenzi *et al.*, 2006). By this account, envelope cues have a reduced role for speech perception in modulated noise, elevating the relative importance of temporal fine structure speech cues that are available in the envelope minima of the masker. Although the results of psychoacoustical (e.g., Bacon and Viemeister, 1985) and speech studies (e.g., Turner *et al.*, 1995) suggest that hearing-impaired listeners are often capable of using temporal envelope cues well (at least for stimuli presented in quiet), there is growing behavioral evidence that such listeners often have a reduced ability to benefit from temporal fine structure cues (e.g., Lacher-Fougere and Demany, 1998; Moore and Moore, 2003; Buss *et al.*, 2004b). Lorenzi *et al.* (2006) hypothesized that poor speech perception of hearing-impaired listeners in modulated noise may result from (1) the increased importance of temporal fine structure cues for speech in such noise, and (2) the reduced ability of listeners with sensorineural hearing loss to utilize temporal fine structure cues. This reasoning is also consistent with the present finding that listeners with sensorineural hearing loss showed relatively poor performance for synchronously modulated speech: this external modulation may reduce the utility of speech envelope cues, thereby increasing the importance of temporal fine structure cues for which the hearing-impaired listeners have a diminished processing ability.

The above account raises the question of why asynchronous modulation of speech did not appear to be as problematic as synchronous modulation for the hearing-impaired listeners. One possibility is that it is easier to follow speech envelope cues over time in the asynchronous modulation conditions than in the synchronous modulation condition. It is well known that the envelopes associated with different frequency regions of speech are often correlated (e.g., Remez *et al.*, 1994). This raises the possibility that speech envelope cues can be followed across the odd/even-band phases of asynchronous modulation, allowing better utilization of speech envelope cues than is possible in the case of synchronous modulation. This would be particularly likely in hearing-impaired listeners, as reduced frequency selectivity would increase the likelihood that adjacent odd/even speech bands would stimulate common frequency channels in which speech envelope cues could be represented.

V. CONCLUSIONS

When compared to normal-hearing results, the thresholds of the hearing-impaired listeners were more elevated in the synchronous modulation condition than for unprocessed speech in noise. This finding with modulated speech mirrors that demonstrated in the literature for speech in modulated noise. Given that the present results were not likely to have been influenced significantly by forward masking, these results suggest that factors other than temporal resolution play an important role in the speech perception deficits demonstrated here and those shown previously with modulated noise. One hypothesis for this poorer performance is that the corruption of speech information associated with modulation interacts with the additional corruption of speech information associated with hearing loss. A second, related hypothesis is that synchronous modulation of the speech signal reduces the ability of the listener to benefit from cues related to the speech envelope. With the reduction in the availability of speech envelope cues, the performance of the hearing impaired suffers because the listeners are forced to depend upon a poorly encoded cue, temporal fine structure.

The results of the present study did not suggest a general deficit by hearing-impaired listeners to integrate asynchronous spectral information in the perception of speech. In fact, many of the hearing-impaired listeners in this study showed better performance in the asynchronous conditions, where adjacent bands of speech were modulated out of phase, than in the synchronous conditions where the modulation of all bands of speech was in phase. Two interpretations of this result were considered: (1) the better performance in the asynchronous modulation condition occurred because deleterious effects related to poor frequency selectivity were reduced under conditions of asynchronous modulation; (2) speech envelope cues were more accessible in conditions where the modulation of the speech was asynchronous rather than synchronous.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD Grant No. R01 DC00418. The authors thank Heidi Reklis and Madhu B. Dev for assistance in running subjects.

[1]In these accounts, it is assumed that the same factors may underlie the relatively poor performance for hearing-impaired listeners when either the speech signal or a masking noise is modulated. Although these accounts have the virtue of parsimony, a caveat is that there are potentially important perceptual differences associated with the noise modulation and speech modulation paradigms. For modulated noise, informal listening suggests that speech has a relatively natural quality at both high and low signal-to-noise ratios (e.g., 10–15 dB above the 50% identification threshold versus a few dB above the 50% identification threshold). For modulated speech, the speech signal has a somewhat unnatural quality at high signal-to-noise ratios, but a more natural quality at low signal-to-noise ratios, perhaps related to the induction effect, where background noise may promote "filling in" of missing parts of an auditory image (e.g., Bashford and Warren, 1987). Thus, although the modulated noise and modulated speech paradigms have conceptual similarities, they are associated with perceptual differences, and it is not clear how such perceptual differences might interact with the factor of hearing impairment.

ANSI (**2004**). "ANSI S3.6-2004, Specification for audiometers," (American National Standards Institute, New York).

Assmann, P. F., and Summerfield, A. Q. (**2004**). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer Verlag, New York).

Bacon, S. P., and Viemeister, N. F. (**1985**). "Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects," Audiology **24**, 117–134.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Baer, T., Moore, B. C. J., and Gatehouse, S. (**1993**). "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment - effects on intelligibility, quality, and response-times," J. Rehabil. Res. Dev. **30**, 49–72.

Bashford, J. A., Jr., and Warren, R. M. (**1987**). "Effects of spectral alternation on the intelligibility of words and sentences," Percept. Psychophys. **42**, 431–438.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2004a**). "Spectral integration of synchronous and asynchronous cues to consonant identification," J. Acoust. Soc. Am. **115**, 2278–2285.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2004b**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Buss, E., Hall, J. W., and Grose, J. H. (**2003**). "Effect of amplitude modulation coherence for masked speech signals filtered into narrow bands," J. Acoust. Soc. Am. **113**, 462–467.

Carlyon, R. P., Deeks, J. M., Norris, D., and Butterfield, S. (**2002**). "The continuity illusion and vowel identification," Acta. Acust. Acust. **88**, 408–415.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

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.

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). "Effects of age and mild hearing loss on speech recognition in noise," J. Acoust. Soc. Am. **76**, 87–96.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

George, E. L., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

Glasberg, B., and Moore, B. (**1989**). "Psychoacoustics abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Greenberg, S., and Arai, T. (**1998**). "Speech intelligibility is highly tolerant of cross-channel spectral asynchrony," in *Joint Proceedings of the Acoustical Society of America and the International Congress on Acoustics* (Seattle).

Healy, E. W., and Bacon, S. P. (**2002**). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," J. Speech Lang. Hear. Res. **45**, 1262–1275.

Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in checkerboard noise," J. Acoust. Soc. Am. **93**, 2915–2922.

Lacher-Fougere, S., and Demany, L. (**1998**). "Modulation detection by normal and hearing-impaired listeners," Audiology **37**, 109–121.

Leek, M. R., and Summers, V. (**1996**). "Reduced frequency selectivity and the preservation of spectral contrast in noise," J. Acoust. Soc. Am. **100**, 1796–1806.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999**). "Contrast enhancement improves the representation of /epsilon/-like vowels in the hearing-impaired auditory nerve," J. Acoust. Soc. Am. **106**, 2693–2708.

Moore, B. C., and Moore, G. A. (**2003**). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," Hear. Res. **182**, 153–163.

Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Plomp, R., and Mimpen, A. M. (**1979**). "Speech-reception threshold for sentences as a function of age and noise level," J. Acoust. Soc. Am. **66**, 1333–1342.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). "On the perceptual organization of speech," Psychol. Rev. **101**, 129–156.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Simpson, A. M., Moore, B. C. J., and Glasberg, B. R. (**1990**). "Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners," Acta Oto-Laryngol., Suppl. **469**, 101–107.

Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., and Mott, J. (**1985**). "Speech perception ability and psychophysical tuning curves in hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 620–627.

Stone, M. A., and Moore, B. C. J. (**2003**). "Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay," Ear Hear. **24**, 175–183.

Summerfield, A. Q., Foster, J., Tyler, R., and Bailey, P. J. (**1985**). "Influences of formant narrowing and auditory frequency selectivity on identification of place of articulation in stop consonants," Speech Commun. **4**, 213–229.

Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410–1421.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Effect of spectral envelope smearing on speech reception. II.," J. Acoust. Soc. Am. **93**, 1547–1552.

Turner, C. W., Chi, S. L., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **42**, 773–784.

Turner, C. W., Fabry, D. A., Barrett, S., and Horwitz, A. R. (**1992**). "Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **35**, 942–949.

Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.

Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C. J., and Patterson, R. D. (**1984**). "Auditory filter asymmetry in the hearing impaired," J. Acoust. Soc. Am. **76**, 1363–1376.

Wilson, R. H., and Carhart, R. (**1969**). "Influence of pulsed masking on the threshold for spondees," J. Acoust. Soc. Am. **46**, 998–1010.

Zwicker, E., and Schorn, K. (**1982**). "Temporal resolution in hard of hearing patients," Audiology **21**, 474–492.

Temporal weights in the level discrimination of time-varying sounds[a)]

Benjamin Pedersen[b)] and Wolfgang Ellermeier[c)]
Sound Quality Research Unit (SQRU), Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7-B5, 9220 Aalborg Øst, Denmark

(Received 12 July 2006; revised 17 November 2007; accepted 19 November 2007)

To determine how listeners weight different portions of the signal when integrating level information, they were presented with 1-s noise samples the levels of which randomly changed every 100 ms by repeatedly, and independently, drawing from a normal distribution. A given stimulus could be derived from one of two such distributions, a decibel apart, and listeners had to classify each sound as belonging to the "soft" or "loud" group. Subsequently, logistic regression analyses were used to determine to what extent each of the ten temporal segments contributed to the overall judgment. In Experiment 1, a nonoptimal weighting strategy was found that emphasized the beginning, and, to a lesser extent, the ending of the sounds. When listeners received trial-by-trial feedback, however, they approached equal weighting of all stimulus components. In Experiment 2, a spectral change was introduced in the middle of the stimulus sequence, changing from low-pass to high-pass noise, and vice versa. The temporal location of the stimulus change was strongly weighted, much as a new onset. These findings are not accounted for by current models of loudness or intensity discrimination, but are consistent with the idea that temporal weighting in loudness judgments is driven by salient events. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2822883]

PACS number(s): 43.66.Fe, 43.66.Cb, 43.66.Ba, 43.66.Mk [RAL] Pages: 963–972

I. INTRODUCTION

A. Weighting level information in auditory discrimination tasks

When discriminating or evaluating complex sounds, the auditory system may be assumed to integrate information both across spectral regions and over time. A powerful tool to study such integration processes has been the *analysis of weights* given to the stimulus components defined in the experiment. Pioneered by COSS analysis (i.e., analyzing responses "Conditional On a Single Stimulus" or stimulus component; Berg, 1989), a number of related methodologies have evolved (e.g., Lutfi, 1995), all of which have in common that the listener does not have to be explicitly queried as to his or her weighting of the informational elements. Rather, all but a *global* judgment of pitch (Berg, 1989), loudness (Willihnganz *et al.*, 1997), or lateralization (Saberi, 1996; Stecker and Hafter, 2002) is required, from which, via statistical analysis or the construction of psychometric functions, its relation to the particular informational components is derived.

1. Spectral weights

Most of the few studies applying the analysis-of-weights methodology to the auditory system's use of level information have been concerned with the determination of *spectral* weights in level-discrimination tasks (Doherty and Lutfi, 1996, 1999; Kortekaas *et al.*, 2003; Willihnganz *et al.*, 1997). To that end, in a two-interval, forced-choice paradigm, random, independent level perturbations were added to each of a number of tonal components of different frequency, and the effect of these frequency-specific perturbations on the listener's overall decision yielded the spectral weights in question. Typically, the average weighting functions were found to be relatively flat, though sometimes with greater emphasis given to the highest or lowest frequency components (see Kortekaas *et al.*, 2003).

2. Temporal weights

There have been hardly any studies on the weighting of level information as a function of time (for a review see Stellmack and Viemeister, 2000). Buus (1999) investigated the detectability of a series of six adjacent 25-ms, 1-kHz tone pulses in masking noise. By adding independent level perturbations to the pulses, he was able to construct conditional psychometric functions relating detectability to the random level variations, separately for each of the six temporal pulse locations. From the slopes of these psychometric functions, much like in COSS analysis, relative weights were derived specifying the contribution of each temporal position in the pulse sequence to overall detectability. Analyzing three listeners in a number of experimental conditions, Buus found their weighting functions to be nearly optimal, i.e., giving

[a)]Parts of this work were presented at the 149th meeting of the Acoustical Society of America, Vancouver, Canada, May 2005 and at the joint meeting of the German and the French acoustical societies (CFA/DAGA), Strasbourg, France, March 2004.
[b)]Now at Atlantic Technologies, Jónas Broncks gøta 35, FO-110 Torshavn, Faroe Islands. Electronic mail: ja_min_pedersen@hotmail.com.
[c)]Now at Technische Universität Darmstadt, Germany. Electronic mail: ellermeier@psychologie.tu-darmstadt.de.

equal weight to each of the (equally informative) six pulses, with small, but statistically significant departures favoring the middle portion of the pulse sequence (see his Fig. 3).

Lufi's (1990) studies of sample discrimination contained one condition in which sequences comprised of up to 12 tones had to be discriminated on the basis of an overall level difference between target and standard sequence. COSS analysis (performed on the data of a single listener, see Lufi's Fig. 9) showed the weights assigned to the elements in the sequence to be approximately equal.

In a study involving one of the present authors (Ellermeier and Schrödl, 2000), using a 2IFC paradigm, on each trial listeners compared two 1-s samples of broadband noise (one of which was incremented relative to the other by 1 dB) with respect to their overall loudness. The noise samples were divided into ten segments of 100 ms each onto which small, random level perturbations were imposed. Using COSS analysis (Berg, 1989), weights were derived for the ten temporal segments. They exhibited a bowl-shaped pattern with the beginning of the noise sequence, and (to a lesser extent) the end being emphasized.

B. Memory effects

Further evidence for an unequal weighting as a function of time comes from studies investigating performance effects supposedly related to the functioning of auditory memory. These studies, however, looked at the discriminability of tone patterns in which *frequency* (or pitch) changes rather than level changes had to be tracked. McFarland and Cacace (1992) found strong primacy and recency effects in tone patterns being between seven and thirteen elements long, i.e., significantly better discrimination at the beginning or end of the sequence.

Surprenant (2001) varied the interstimulus interval (ISI) between the sequences to be discriminated, and found strong recency effects, with additional primacy effects emerging as the ISI was increased. Whether such memory effects are obtained for the discrimination of level changes as well remains an open question.

C. Rationale

Given the scarce and equivocal evidence regarding temporal weighting in level discrimination, it appears worthwhile to reinvestigate the issue. In contrast to earlier investigations that shall be done using a *one-interval task* much like in the original study illustrating the weights technique (Berg, 1989). In the present implementation, subjects will be presented with a single stimulus on each trial, and will simply have to classify it as belonging to the "loud" or "soft" set defined by the experiment. This task is conceptually much simpler than a 2IFC task (see Kortekaas *et al.*, 2003), and it does not require assumptions about within-trial memory processes involved, such as making different predictions depending on the length of the ISI (Surprenant, 2001).

Furthermore, since it is conceivable that the contradictory outcomes of some of the studies of temporal weighting may be due to different degrees of practice with the task, or to different strategies used, in Experiment 1, the opportunity to acquire an optimal weighting (with respect to using the physical level information) shall be experimentally manipulated by giving one group of listeners explicit trial-by-trial feedback as to the "correct" response alternative, while another group receives no such feedback, and thus no chance to optimize their strategy.

Finally, since those authors motivated by theories of memory have speculated on the "distinctiveness" of certain events in the temporal sequence, such as the beginning and end of a sound (Neath *et al.*, 2006; Surprenant, 2001), in Experiment 2 additional distinct events shall be experimentally induced by abruptly changing the spectral content of the sound to be judged. In particular, noise sequences will be designed that instantaneously shift from a low-pass to a high-pass characteristic (and vice versa) in the middle of the temporal sequence. Potentially, the spectral shift might constitute a new "distinct" event, e.g., signaling a new "onset," and thereby altering the weight pattern when compared to a control sequence of nonchanging broadband noise.

By virtue of the use of trial-by-trial feedback based on the physical sound generation (in Experiment 1), the task becomes one of *intensity discrimination* (albeit based on multiple channels). It is reasonable to assume, however, that—no matter whether they receive feedback, or not—the subjective quality listeners base their decisions on is related to some "internal" computation of instantaneous or overall *loudness*. The advantage of this view is that it brings models of time-varying loudness and of loudness integration to bear on the behavior observed.

II. EXPERIMENT 1: LEVEL-FLUCTUATING SOUNDS

A. Method

1. Listeners

Ten listeners (one female, nine male) including the authors ("WE" and "BP" in the figures) participated in the experiment. The mean age of the participants was 26 years (range: 18–46 years). All were audiometrically screened, and no one was found to have significant hearing loss (more than 20-dB hearing loss at more than one frequency of 0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz). Except for the authors, the participants were students with little or no experience in listening experiments.

2. Apparatus

Stimuli were generated digitally on the PC controlling the experiment. A Tucker Davis Technologies System 3 was used for digital-to-analog conversion (RP2.1 unit), setting appropriate levels (two PA5 attenuators), and for powering the headphones (HB7 unit). Signals were presented diotically via headphones (Beyerdynamic DT 990 PRO), at a sample rate of 50 kHz and with 24 bit resolution.

The listeners were seated in a double walled listening cabin during the experiment and made responses using two buttons marked "soft" and "loud" on a special button box connected to the Tucker Davis RP2.1 unit. The box was also used for providing feedback using red and green lights.

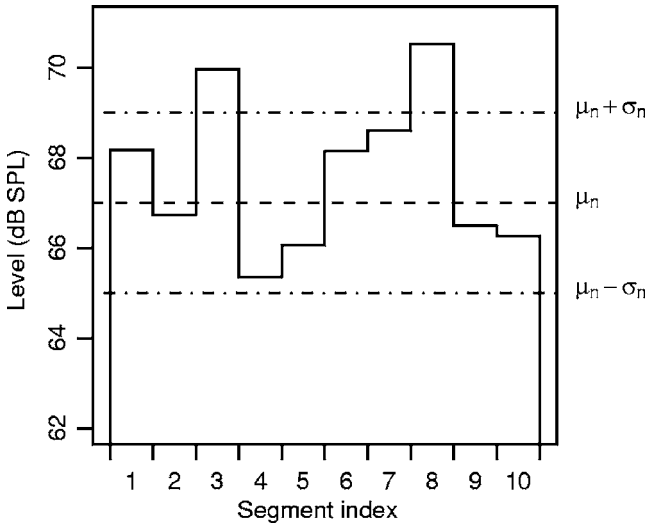


FIG. 1. Temporal envelope of a sound sample (here, "noise").

3. Stimuli

The sounds used in the experiment were samples of white noise having 1-s duration. Their overall level was randomly varied every 100 ms, thus producing a stepwise level-fluctuating sound consisting of ten segments (see Fig. 1). The overall level of each segment was picked randomly from one of two normal distributions denoted "signal" and "noise," with the signal distribution having a higher mean value. The signal distribution had mean value μ_s=68 dB SPL and a standard deviation of σ_n=2 dB. The noise distribution had a mean value μ_n=67 dB SPL and a standard deviation of σ_s =2 dB. Consequently, approximately 95% of the segment levels for each distribution fall in the range $\mu \pm 4$ dB.

Further, the noise and signal distributions overlapped considerably, such that the mean of the ten segments of a given noise sound was sometimes higher than the overall mean (67.5 dB) and vice versa for signal sounds. How often that is expected to happen can be estimated using the properties of the normal distribution. The standard deviation of the mean (given ten segments) is: $\sigma_{10}=\sigma_n/\sqrt{N}$, where N is the number of segments per stimulus (ten). Thus there is approximately a 21% chance for the mean of the ten noise segments to exceed the midpoint between the noise and signal distributions (67.5 dB).

The setup was calibrated using an artificial ear (Brüel & Kjær 4153) with a microphone (Brüel & Kjær 4134). When sound pressure levels are used throughout this article, they refer to the rms sound pressure level of a continuous broadband noise as would be measured in the artificial ear at the given presentation level.

4. Experimental procedure

Participants were instructed that the sounds "were randomly generated," and came from a "soft" or a "loud" set of levels with equal probability. A one-interval two-alternative forced-choice paradigm was used. On each trial, the listener heard a single sound and was asked to judge it as being either soft or loud. In the sequence of trials, noise and signal sounds were presented in random order.

Listeners were divided into two groups in one of which the listeners received trial-by-trial feedback. If the generated sound was from the noise distribution and the listener responded soft or if the sound was from the signal distribution and the response was loud the feedback was a green light, in the other cases it was a red light. No such feedback was given to the other group.

After the completion of each block of 130 trials, overall feedback was given by telling the participants the percentage of "correct" responses they had obtained, i.e., responses which agreed with the noise or signal property of the stimulus. This type of overall feedback was given to all listeners. It helped to motivate the listeners, however based on this type of feedback, it was impossible to change a decision strategy based on trial-by-trial learning.

The first two and a half sessions were used for training. During training the difference between the noise and signal means, μ_s and μ_n, was successively decreased from 3 dB over 2 dB to a final 1-dB difference.

5. Data collection

The experiment was arranged in blocks of 130 trials of which only trials 10–130 were analyzed, leaving the first 9 trials for building up a decision criterion. Five such blocks made up one session, which lasted approximately 40 min. Each listener proceeded through 10 sessions.

6. Determination of temporal weights

In making an overall judgment, listeners are assumed to base their responses on a decision variable, D, defined as

$$D(\mathbf{x}) = \left(\sum_{i=1}^{10} w_i x_i \right) - c, \tag{1}$$

where $\mathbf{x}$ is a vector of the ten segment levels constituting a given sound. x_i refers to the sound pressure level in decibels of each of the ten segments and w_i is a perceptual weight given to the ith segment. It is assumed that the weighted sum of the segment levels is compared to a fixed decision criterion c. So the strength of the decision variable is given by the difference between the magnitude of the weighted sound levels and the fixed decision criterion.

A logistic function was employed to statistically relate the binary dependent variable (judgments of loud and soft) to the strength of the decision variable:

$$\Psi(D) = p(\text{"loud"}) = \frac{e^D}{1+e^D} = \frac{1}{1+e^{-D}}, \tag{2}$$

where Ψ describes the probability, p, of a loud response. Note that sometimes other functions (e.g., normal ogives, Berg, 1989) are used to characterize Ψ, but it has been shown, and is true for the present data, that the estimated weights are to a great extent insensitive to the choice of function (Tang *et al.*, 2005).

Insertion of Eq. (1) in Eq. (2) gives

$$\Psi(\mathbf{x}) = p(\text{loud}|\mathbf{w}, c, \mathbf{x}) = \frac{1}{1+e^{c-\Sigma_i w_i x_i}}. \tag{3}$$

TABLE I. Performance of all listeners in Experiment 1 as the percentage of trials which were correctly identified according to the distribution of origin (DIST rows) and according to the mean sound pressure level of the ten segments of a given sound (SPL rows). Signal detection theory sensitivity (d') and bias (β) scores also indicated in separate rows. Listeners in the no-feedback condition (NF) are in the top half and listeners in the feedback condition (FB) are in the bottom half.

		BB	BJ	BP	CP	JJ	Mean
NF	DIST	63	66	69	65	66	**66**
	SPL	68	72	76	71	71	**72**
	d'	0.67	0.86	1.00	0.78	0.85	**0.83**
	β	0.93	0.86	0.91	0.89	1.06	**0.93**
		BL	EH	JV	LH	WE	Mean
FB	DIST	64	72	73	62	69	**68**
	SPL	68	83	81	66	76	**75**
	d'	0.74	1.17	1.21	0.61	0.97	**0.94**
	β	1.05	0.92	0.90	0.85	0.91	**0.93**

The outcome of the experiment is a sequence of loud and soft responses with associated values for **x**. The values of **w** and c which are most likely to yield the results, under the given model, can be estimated by maximum likelihood optimization. For the logistic function, as applied here, this is also known as logistic regression. Standard test statistics for the validity of the model can be applied and furthermore the logistic regression has the benefit of being directly applicable to binary (loud and soft) data (see, for example, Cohen, 2003). These are the main reasons for choosing logistic regression over alternative methods used in other studies estimating weights (for example, Berg, 1989; Ellermeier and Schrödl, 2000; Lutfi, 1995). Though conceptually different, the various methods at hand give very similar estimates for perceptual weights in practice.

It is seen from Eq. (3) that the regression coefficients (**w** and c) are not linearly related to the predicted probability of making loud response. The nonlinear relationship is generally true for logistic regression. In this work however, the logistic function is used as a psychometric function, and the regression coefficients are linearly related to the strength of the underlying decision variable as stated in Eq. (1).

In Eq. (1), a linear relationship between the decision variable, D, and the segment levels, **x**, is assumed. Generally, however, the loudness of steady-state sounds is not linearly related to the sound pressure level in decibels, but within the range of levels used in the present experiment (approximately 60 dB to 75 dB SPL) the relationship is close to linear (see Moore, 2003).

B. Results of Experiment 1

In Table I the performance of the listeners is evaluated via four different measures (DIST, SPL, d', and β in Table I). The DIST score indicates the percentage of trials on which the listeners correctly identified whether the sound originated from the signal or noise distribution. Thus, it evaluates performance in the same way as the feedback given during the experiment. Based on this statistic, performance is very similar in the no-feedback (66%) and feedback (68%) conditions. An alternative is to compute d' and β as defined in signal detection theory. The basis for the two measures is the percentage of trials where the signal distribution was correctly identified (hit-rate) and the percentage of trials where noise sounds were incorrectly identified as loud (false alarm rate). It appears that d' is slightly higher (0.94 vs 0.83) in the feedback condition, but due to the interindividual variance this difference does not reach statistical significance. The bias is nearly identical in the two conditions, marginally favoring loud judgments (β=0.93). It might be argued that the distribution-based performance measures (DIST and d') are unfair, because it is impossible to get all trials correct, since, by chance, high levels can originate from the noise distribution, and vice versa. Therefore, another performance measure (termed "SPL" in Table I) was computed based on the trial-by-trial mean sound pressure level of the ten sound segments. If this mean was higher than 67.5 dB (the midpoint between the two distributions), a loud response was considered correct and when lower than the overall mean, a soft response was considered correct. It appears that performance measured in this way is only slightly higher in the feedback group (75%) when compared to the no-feedback group (72%). The interindividual variance in performance is significantly larger than the mean difference between the two experimental conditions, ranging from 66% for listener "LH" to 83% for listener "EH." But note that this performance measure may not constitute a "fair" comparison, since it favors "flat" weighting curves and a decision criterion close to the overall mean.

In total, 4598 trials per listener (38 blocks $\times$ 121 trials) were used to derive weighting curves. The individual weighting curves are seen in Fig. 2. The weights are the scaled regression coefficients of the logistic regression [w_i in Eq. (1)], which provided the most likely fit to the listeners responses given the segment levels (x_i). The coefficients (w_i) are scaled by a factor so the sum of the ten weights is 1. This normalization makes the *relative* importance of each segment (the weighting curve) comparable across listeners. Different scaling values for different listeners reflect individual differences in sensitivity to level changes, which imply that the overall sensitivity is not reflected in the scaled weighting curves.

Figure 2 shows the derived weighting curves for listen-

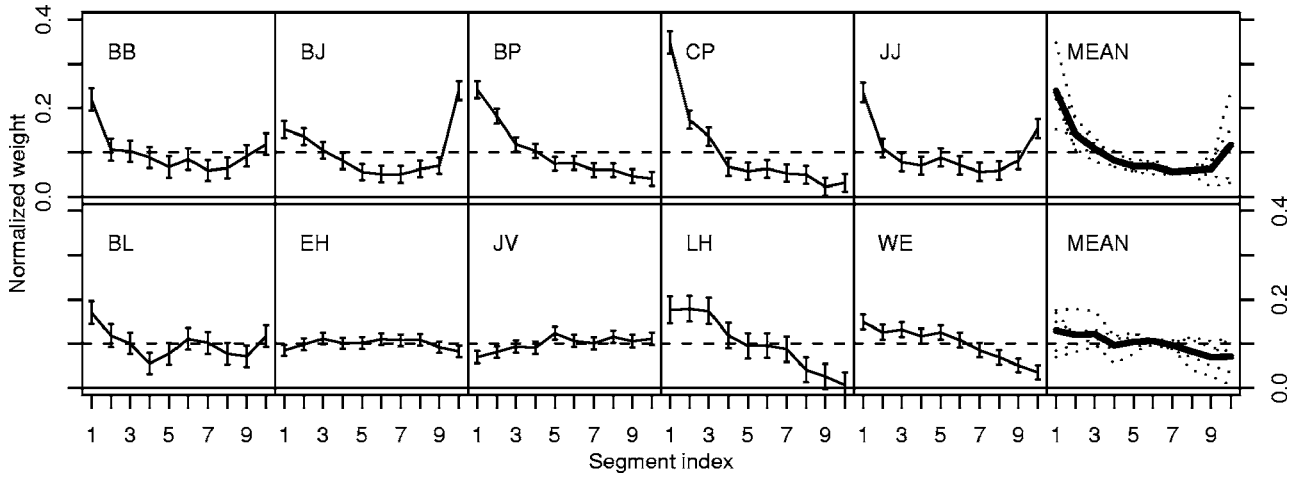


FIG. 2. Temporal weights for Experiment 1. Top row: Without trial-by-trial. Bottom row: Feedback. The error bars indicate the 95%-confidence intervals for the weights as calculated from the logistic regression. Average weights for the two conditions are depicted in the right column.

ers receiving feedback (bottom row) and those of listeners not receiving feedback (top row). For each segment weight, the error bar indicates the 95%-confidence interval. The end points of each interval were calculated prior to normalization and afterwards scaled by the same normalization factor as the weights. Comparing the size of the error bars to the weight differences between segments, it is clear that the shape of the weighting curve is meaningful for a given listener and not a product of random processes. It is also clear that the weighting curves are highly individual, consider "BJ" versus "CP" for example: CP heavily weights the beginning of the sound, while BJ put most weight on the end. For most listeners either the beginning or ending of the sound is weighted more heavily. Exceptions from this are "EH" and "JV," who do not show pronounced weighting of specific segments.

The effect of feedback can be inspected by comparing listeners in the upper row of Fig. 2 to those in the lower one. Comparing the two mean weighting curves it looks as if feedback did influence the overall shape of the weighting curves. The tendency to emphasize the beginning or the end of a sound seems to be more pronounced in the group of listeners who did *not* receive feedbacks.[1]

An estimate of the statistical significance of this apparent influence of feedback is not easily made, since (a) all weights are normalized to sum to 1, and (b) the weights for the ten segments are not statistically independent for a given listener. Therefore the following testing strategy is suggested: If a listener does not receive feedback, either the beginning or the ending of the sound is weighted more heavily. In any case (beginning, ending or both being weighted more heavily), the central part of the sound must receive less weight due to the normalization. A score for each listener's weighting of the central part of the sound can be obtained by calculating the sum of the "central" weights 4–8. One score is thus obtained for each of the listeners in each group, and the scores in the groups can be compared using a two-sample t-test. It turned out to be highly significant, $t(7.16)=5.30$;[2] $p=0.001$ indicating that the central weights in the no-feedback group were lower than in the feedback group. This in turn means that the curves in the feedback and no-feedback conditions do indeed have different shapes. If non-normalized weights are used, the certainty is even greater, however, this merely implies that listeners receiving feedback perform better than those not receiving feedback.

The same approach can be used to compare the mean weighting curve in each group to "flat" weights (all weights being equal to 0.1). When, in the no-feedback group, the central weights are compared to a value of 0.1, a one sample t-test results in $t(4)=10.01$; $p<0.001$, and in the feedback group: $t(4)=0.54$; $p=0.62$. That is, the central part of the mean curves is significantly different from optimal weighting for the no-feedback group only. However, from the 95%-confidence intervals in Fig. 2 it is clear that some weights are significantly different from the optimal 0.1 for individual listeners both in the feedback and no-feedback group.

C. Discussion

Global loudness judgments of level-fluctuating noise samples produced evidence for a nonoptimal temporal weighting in that onsets (and to a lesser extent offsets) were weighted more heavily in contributing to overall loudness. A similar, u-shaped weighting pattern as a function of time was observed by Sadralodabai and Sorkin (1999), though for an entirely different task (detecting temporal-pattern changes in a sequence of sinusoids). Furthermore, in the present experiment, trial-by-trial feedback—the presence of which turns the loudness classification into an intensity discrimination task—significantly reduced this emphasis, effectively resulting in an approximately equal (i.e., optimal) weighting of all segments of the sounds.

The present experiment thus provides support both for equal (as in Buus, 1999; Lutfi, 1990) and unequal (as in Ellermeier and Schrödl, 2000) temporal weights, and though all previous studies used some form of feedback it may be speculated that it may have been implemented more or less efficiently. The fact, however, that those participants receiving feedback in the present study were able to "optimize" their performance (with respect to correctly identifying noise

and signal sounds) to approximate ideal weights, suggests that there is considerable potential for "perceptual learning" in the temporal weighting patterns.

Earlier investigations have shown that feedback affects the weighting pattern in complex detection or discrimination tasks, e.g., by (a) shifting attention to different spectral regions (Doherty and Lutfi, 1999), (b) focusing on different temporal components (Plank and Ellermeier, 2003), or (c) using different physical cues altogether (Richards, 2002). Note that in all of these examples, however, the task per se was changed (e.g., feedback was made contingent on a different signal property; Richards, 2002), while in the present experiment the mere presence or absence of feedback altered the weighting pattern.

It thus appears, as has been shown for spectral weights (Lutfi, 1995; Southworth and Berg, 1995), there is considerable liberty in how listeners weight the components of perceptual information available, and that, depending on the task requirements, different weighting patterns may emerge. The considerable individual differences evident in the present data also argue for a certain flexibility in the assignment of weights.

One might speculate whether the overall pattern of weights decreasing with segment number (see Fig. 2) is due to earlier noise segments masking the later ones (forward masking). This is unlikely due to several reasons: First, the relative level differences between adjacent segments are too small (a few decibels), and the segment duration is too long (100 ms) to expect much forward masking. Second, the fact that essentially flat patterns (Subjects EH and JV), or a pattern that emphasizes the end (Subject "BJ") are observed, argues against a peripheral process such as masking being causal. Third, one would not expect feedback to produce a release from forward masking. Fourth, and finally, simulations using the loudness model by Glasberg and Moore (2002), which takes masking into account, failed to predict a decaying pattern of weights (Pedersen, 2007, Chap. 3).

The outcome of Experiment 1, however, does not specify the nature of the processes very well. It remains open, for example, whether the emphasis of beginning and ending observed in the unbiased listening condition is due to memory effects (primacy and recency), or simply to the perceptual salience of onsets and offsets.

III. EXPERIMENT 2: TWO-EVENT SOUNDS

To further clarify the issues raised by Experiment 1, a second experiment was performed, in which sounds of the same duration and temporal structure as those used in Experiment 1 were subjected to a sudden spectral change in the middle of the temporal sequence. The spectral change thus constitutes a salient event which is not tied to primacy or recency, and the effect of which on the temporal weighting pattern may be observed.

A. Method

1. Listeners

Six naive listeners took part in the experiment, none of whom had participated in Experiment 1. Their hearing was screened, and no one was found to have significant hearing loss (more than 20-dB hearing loss at more than one frequency of 0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz). The participants were five males and one female with an average age of 24 years (range: 22–28 years).

2. Apparatus

In Experiment 2, different hardware was used for signal generation: Signals were digitally generated using a sound card (RME HDSP9632) and subsequently converted to an analog signal via a digital-to-analog converter (Tracer Technologies Big DAADI), using 16-bit resolution and a sample rate of 44.1 kHz. The resulting signal was fed to a headphone amplifier (Behringer HA4400) and diotically played over headphones (Beyerdynamic DT 990 PRO).

3. Stimuli

As in Experiment 1, all sounds were of 1-s duration and the levels of the ten temporal segments were chosen from random distributions having the same parameters as in Experiment 1. The only difference was the spectral content of the sounds. In one condition of Experiment 2 the first half of the sound (i.e., the first five segments) was low-pass filtered and the last part (the last five segments) high-pass filtered. This type of sound is denoted "LH," indicating the change from low to high frequency content. In a different condition the segments were filtered in the opposite order, denoted "HL," i.e., changing from high-pass to low-pass filtered noise. The cut-off frequency was 1 kHz for both high- and low-pass filters. The filtering was done using digital finite impulse response filters (FIR of order 501), for which the attenuation was more that 50 dB in the nonpass section at a distance of more than 150 Hz from the cut-off frequency (1 kHz). The phase response of each filter was linear. The two filtered blocks were aligned so no silent interval occurred. A third condition, where no spectral change occurred, was included for comparison with Experiment 1. In this condition white noise was used as in Experiment 1 (denoted "WN").

4. Experimental procedure

The listeners' task was the same as in Experiment 1. After hearing a single sound, the listener responded whether it was loud or soft. No trial-by-trial feedback was given. After each block of 200 trials the percentage of "correct" responses based on the distribution from which the sounds were drawn was communicated to the participants. Because of the difference in quality of the filtered blocks, the listeners were specifically instructed to judge the composite sound as one whole.

Before data collection started all listeners learned the task in a similar way as in Experiment I. The difference in mean between the noise and signal distributions was slowly decreased (from 4 to 1 decibels). The training blocks contained fewer trials (50) and LH, HL, and WN blocks were included. Feedback on the percentage of correct responses helped listeners to realize whether they were on the right track.

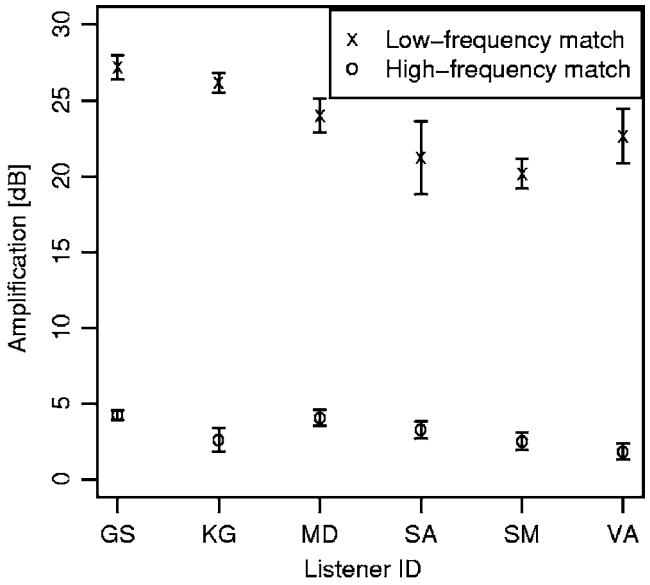


FIG. 3. Loudness matches for Experiment 2. Amplification required for the low-pass and high-pass noises to match a 67.5 dB SPL white-noise reference. Individual outcomes for each of the six listeners are depicted. The error bars indicate the 95%-confidence intervals.

5. Data collection

The experiment was arranged in blocks of 200 trials each. A given block contained either filtered noise (both LH and HL) or broadband noise (WN). In blocks containing filtered noise, LH and HL trials were presented in a random sequence. A total of 1200 trials per condition (LH, HL, or WN) was presented. Each session, lasting approximately 40 min, contained three blocks, one in which unchanging white-noise stimuli were presented (WN), and two containing spectral changes (LH and HL). The order of the blocks was counterbalanced within listeners, and across the six sessions used for data collection. 1140 trials were used per condition and listener in the regression analysis, since the first 9 trials in each block were discarded for practice.

6. Loudness calibration

In order to present the filtered noises at equal loudness, all listeners initially performed individual loudness matches before proceeding to the experiment proper. An adaptive two-interval forced choice one-up/one-down paradigm was used to match samples of either low-pass or high-pass filtered noise to the fixed white-noise reference at 67.5 dB SPL. All sounds had a duration of 0.5 s and there were no random fluctuations of the segment levels.

The resulting loudness matches varied somewhat across listeners (up to ~7 dB for the low-pass, and 3 dB for the high-pass noise, see Fig. 3). They required the low-pass noise to be raised in level by approximately 23 dB on average to be equally loud as the broadband noise. The high-pass noise required 3-dB amplification on average to achieve the same loudness. In the experiment proper *individual* matches were used for calibration of the filtered blocks.

B. Results of Experiment 2

As in Experiment 1, performance measures were calculated for each listener in each condition. On average, performance was almost identical in the three experimental conditions (WN, LH, and HL), amounting to approximately 64% correct when based on the SPL criterion. The SPL score varied across listeners with a minimum of 54% for listener VA in the LH condition and maximum of 74% for listener MD in the WN and LH conditions. The value of β was very close to 1.0 for all listeners in all conditions, indicating an equal balance between loud and soft judgments in all conditions. Generally, the performance was worse in Experiment 2 than in Experiment 1, which may for example be caused by the extreme weighting applied by some listeners (see for example GS in Fig. 4) or because of the listeners being less consistent in their judgments (large error bars for VA in Fig. 4). However, there are no indications that spectral-change condtions are harder than the condition with no spectral change.

As in Experiment 1, weighting curves were derived for each individual listener, using logistic regression, separately for the white noise (WN), low-high (LH), and high-low (HL) conditions. The estimated weights are depicted in Fig. 4.

The results of the white-noise condition may be compared to those of Experiment 1, in which identical stimuli were used. A similar trend as in the "no feedback" condition of Experiment 1 is found (compare top rows of Figs. 4 and 2), with relatively greater weights being assigned to the initial sound segments. The results of the two experiments are similar, except that the emphasis on the initial segments is even greater, and there is no evidence for higher weighting of the ending of the sound in the new experiment. As in Experiment 1, the weighting patterns vary across listeners.

When a spectral change is introduced in the middle of the sound (LH and HL in the center and bottom rows of Fig. 4), the weighting curves show distinctly different patterns. For most listeners the sixth segment (for which the spectral change occurs) receives greater weight in the LH and HL conditions. It also appears that the order of the high- and low-pass filtered blocks makes a difference for the weighting strategy applied by the listeners, though in idiosyncratic ways, consider GS for example: In the HL condition his decision is based almost exclusively on the first segment (beginning of low-frequency block), whereas in the LH condition both the first and the sixth segment contribute significantly to the decision. Thus, the start of the low-frequency block is always heavily weighted by this listener, but the beginning of the high-frequency block is only weighted heavily if it is also the onset of the entire sound. Listener SM almost shows the reverse behavior with respect to the weighting in the two spectral conditions. Finally, SA almost seems to ignore the high-frequency part of the sound in both the LH and HL conditions.

As can be seen from the size of the 95%-confidence intervals depicted in Fig. 4, some listeners were clearly more consistent in their weighting than others. Nevertheless, all listeners performed significantly better than chance.

A statistical test as to whether the spectral change (LH or HL) made a difference compared to the nonchanging

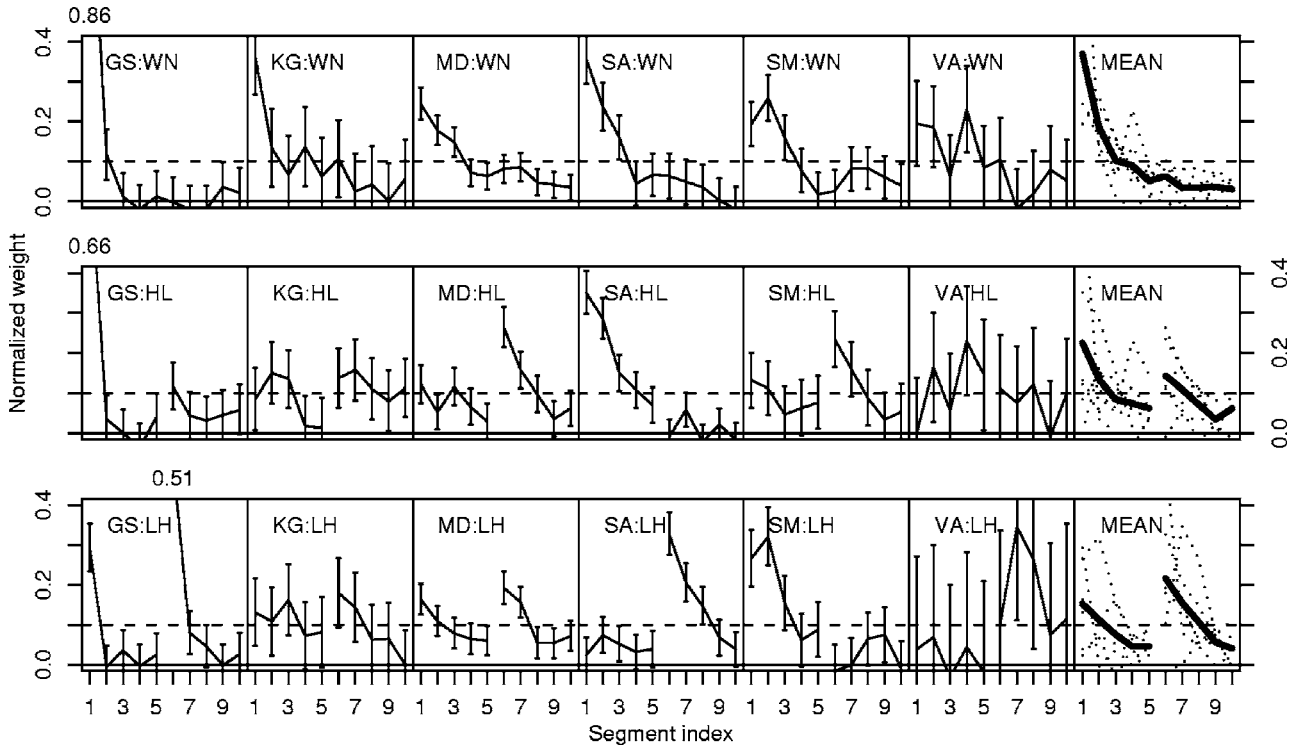


FIG. 4. Weighting curves for all listeners in Experiment 2. Columns represent listeners while rows contain the different experimental conditions. First row: Broadband noise with no frequency change. Second and third rows: Two-event conditions; high-low and low-high changes, respectively. The onset of the spectral change is indicated by a break in the weighting curve. The error bars indicate 95%-confidence intervals for the weights as calculated from the logistic regression. Average weights for the three spectral conditions are depicted in the right column.

(WN) condition was performed in the following way: The sixth and seventh segments were defined as reflecting the onset of the spectral change. By summing each listener's weights for these two segments, a score for the weighting of the spectral change was calculated for each listener in each condition. Using these scores, two-tailed, repeated measures t-tests were performed, between the spectral-change conditions and the nonchanging condition. They revealed the weights for the critical segments (6 and 7) to be significantly greater in the spectral-change conditions, both when comparing LH with WN: $t(5)=3.02$, $p=0.03$, and when comparing HL with WN: $t(5)=2.65$, $p=0.045$. Thus, the increased weighting given to the onset of a new spectral event (see Fig. 4) appears to be statistically significant.

C. Discussion

Experiment 2 showed the temporal location at which a spectral change occurred to receive about as much weight as the initial onset of the composite sound. This is consistent with the idea of perceptual weighting being guided by salient events. These may be onsets, offsets, spectral shifts, or qualitative changes yet to be investigated such as changes in spatial location, etc.

The results of Experiment 2 are not easily reconciled with a memory explanation based on primacy and recency effects, at least not one that requires the entire sound to be stored in memory in a simple sequential way. Whether assumptions about "resetting" the onset detector, or separate storage of the two spectral events might remedy the situation, is doubtful.

IV. FINAL DISCUSSION AND CONCLUSION

That condition of Experiment 1 in which listeners received trial-by-trial feedback based on the generation of the physical levels making up the sound sequence may be considered a straightforward intensity discrimination task. In that task, listeners were encouraged to optimize their performance with respect to inferring which of two level generators produced the auditory percept they had. The experiment showed that—given feedback—the participants could indeed accomplish that and were using the evidence of ten successively presented levels almost optimally, i.e., with a nearly "flat" weighting characteristic (see the bottom part of Fig. 2).

When—by contrast—the trial-by-trial feedback was omitted, as for the other group of listeners participating in Experiment 1, and in Experiment 2, additional effects emerged, such as a higher weighting given to the beginning (and end) of the sound sequence (see the top part of Fig. 2), or an increased emphasis on those portions of the signal that were temporally close to a spectral change (see Fig. 4).

This may be summarized as stating that *salient events* are perceptually emphasized in natural, unbiased listening, as it occurs when no particular feedback scheme is implemented. It may be further hypothesized that that kind of listening is close to our everyday loudness perception. Only when feedback suggests otherwise (as in the pertinent condition of Experiment 1), are the temporal loudness weights adapted to maximize correct performance.

This kind of reasoning, and the fact that most of the data of the present work were collected in an intensity discrimination paradigm that—for the participants—was framed as a

loudness classification task without feedback, suggests to investigate how well current loudness models can explain the peculiar temporal weighting patterns observed. It will be argued that most of the results of the present experiments are incompatible with the notion of an automatic, accumulative integration process as hypothesized by most loudness models (e.g., Glasberg and Moore, 2002; Zwicker, 1977). A major outcome of these loudness models is to generate a continuous loudness curve, which is to account for the results of, e.g., temporal masking experiments and subjective loudness matches of modulated sounds. But to predict how listeners arrive at global loudness judgments requires further stating how this continuous curve is "integrated" to produce a single judgment. The present data address both of these stages.

In the calculation of a continuous loudness curve, all current models operate with some sort of temporal summation with a critical time coefficient in the range from 20 to 50 ms depending on whether the loudness curve is rising or falling (Glasberg and Moore, 2002; Grimm *et al.*, 2002; Zwicker, 1977). It is therefore impossible for loudness determined by these models to fluctuate any faster than the time coefficients allow. The fact that in the present experiment, for some listeners, adjacent segments were weighted very differently (see Fig. 2) implies that their "continuous loudness" must fluctuate at least as rapidly as the segment duration of the sounds (100 ms) or else a particular segment could not be "singled out" receiving extra weight. Thus, though the time coefficients of the models are not in direct contradiction with the observed weighting patterns, there is some indication that the integration taking place is not a simple "smoothing" process. In their loudness model, Glasberg and Moore (2002) introduce a further stage of determining "long-term" loudness, with integration coefficients of approximately 100 ms for rising and 2000 ms for falling loudness. These long time coefficients are not compatible with the results of the present experiments. Other researchers have also found it hard to reconcile the outcome of listening experiments with the predictions of loudness models when different forms of temporal variation were examined (e.g., Grimm *et al.*, 2002; Stecker and Hafter, 2000).

When it comes to integrating the sensory information into a loudness judgment, the present experiments provides further evidence against the operation of simple loudness integration:

(1) Weights derived for the ten temporal segments defined were not uniform, but rather, in the unbiased, nonfeedback conditions of Experiment 1 and 2, provided evidence for perceptual emphasis of onsets and offsets. That is not predicted by any of the current loudness models. Nor is it predicted by practical measurement rules (e.g., Zwicker and Fastl, 1999) that assume values close to the maximum (e.g., the fourth percentile; Grimm *et al.*, 2002; Zwicker and Fastl, 1999) to determine the loudness of a time-varying sound. All of these rules would, for the randomly varying sounds used in the present experiments, imply "flat" weighting curves to result.
(2) When trial-by-trial feedback was provided in Experiment 1, listeners adapted their temporal weights to approach an optimal, uniform weighting of all stimulus segments. Such a "learning effect" is hard to reconcile with the notion of an automatic integration process operating in the auditory periphery with a relatively long time coefficient. Rather, the listener must have access to some representation of the segment loudnesses (prior to integration) with a finer resolution than the segment duration in order to modify weights to maximize the percentage of correct responses.
(3) When a change was introduced into the noise sequence by switching the spectrum from a low-pass to a high-pass characteristic (or vice versa) in Experiment 2, listeners strongly weighted the onset of the "new" sound feature, thus boosting weights in the central portion of the composite stimulus. That is inconsistent with temporal wide-band energy integration which would be "blind" to the spectral change; it is also inconsistent with a memory explanation based on a "primacy" and "recency" advantage.
(4) All of the weighting patterns observed exhibited considerable interindividual variability. That in itself argues against a low-level integration mechanism, which one would not assume to leave degrees of freedom for individual idiosyncrasies. Rather it suggests some cognitive process to be involved, which can be controlled by the listener to some extent.

What then are the alternatives for understanding the weighting of level information, and its adaptability to various listening conditions? It appears that, in the time range of several hundred milliseconds investigated here, different stimulus segments must be individually accessible, granting the listener "multiple looks" (Viemeister and Wakefield, 1991) on a temporal loudness pattern. Depending on the task requirements (Experiment 1) or on stimulus features (the spectral changes in Experiment 2), these "looks" may be weighted differently, under implicit control by the listener. The particular salience of onsets and offsets, as well as qualitative changes in the stimulus, may be due to mechanisms of memory, or more likely to the "distinctiveness" (Neath, 1993; Neath *et al.*, 2006) of these events in relation to other stimulus components, thereby attracting greater perceptual weight.

How could these hypotheses be put to further tests? If memory was a factor, one might expect the timing of the event sequence to play a crucial role. Furthermore, to explore the distinctiveness concept, salient changes other than spectral ones (e.g., spatial lateralization) might be explored, or an event could be generated by switching from coherent to incoherent noise samples of the carrier signal across the two ears. Potentially, the segment levels could also be different across the ears, providing a means to examine both temporal and binaural loudness summation.

Hopefully, based on such research, a clearer picture will emerge, on how perceptual and cognitive processes interact when listeners discriminate time-varying sounds differing in level.

ACKNOWLEDGMENTS

We would like to thank Florian Wickelmaier for helpful hints regarding the statistical analysis of the results. This research was carried out as part of the "Centercontract on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen. Brüel & Kjær, and DELTA Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

[1]The "stability" of the individual weighting patterns was examined by analyzing the data in ten blocks of increasing practice (details in Pedersen, 2007). There was no indication that listeners altered their temporal weighting in the course of the experiment. The observed difference between listeners in the feedback and no-feedback group may thus already emerge in the training trials prior to the experiment proper.

[2]Noninteger degrees of freedom result since a Welch–Satterthwaite approximation was used, which means that equal variance of the weights in the feedback and no-feedback group need not be assumed.

Berg, B. G. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Buus, S. (**1999**). "Temporal integration and multiple looks, revisited: Weights as a function of time," J. Acoust. Soc. Am. **105**, 2466–2475.

Cohen, J. (**2003**). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. (Lawrence Erlbaum, Mahwah, NJ).

Doherty, K. A., and Lutfi, R. A. (**1996**). "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **99**, 1053–1058.

Doherty, K. A., and Lutfi, R. A. (**1999**). "Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **105**, 1831–1840.

Ellermeier, W., and Schrödl, S. (**2000**). "Temporal weights in loudness summation," in *Fechner Day 2000. Proceedings of the 16th Annual Meeting of the International Society for Psychophsics*, edited by C. Bonnet (Université Louis Pasteur, Strasbourg), pp. 169–173.

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.

Grimm, G., Hohmann, V., and Verhey, J. L. (**2002**). "Loudness of fluctuating sounds," Acust. Acta Acust. **88**, 359–368.

Kortekaas, R., Buus, S., and Florentine, M. (**2003**). "Perceptual weights in auditory level discrimination," J. Acoust. Soc. Am. **113**, 3306–3322.

Lutfi, R. A. (**1990**). "Informational processing of complex sound. II. Cross-dimensional analysis," J. Acoust. Soc. Am. **87**, 2141–2148.

Lutfi, R. A. (**1995**). "Correlation-coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," J. Acoust. Soc. Am. **97**, 1333–1334.

McFarland, D. J., and Cacace, A. T. (**1992**). "Aspects of short-term acoustic recognition memory: Modality and serial position effects," Audiology **31**, 342–352.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).

Neath, I. (**1993**). "Distinctiveness and serial position effects in recognition," Mem. Cognit. **21**, 689–698.

Neath, I., Brown, G. D. A., McCormack, T., Chater, N., and Freeman, R. (**2006**). "Distinctiveness models of memory and absolute identification: Evidence for local, not global, effects," Q. J. Exp. Psychol. **59**, 121–135.

Pedersen, B. (**2007**). "Auditory temporal resolution and integration: Stages of analyzing time-varying sounds," Ph.D. thesis, Aalborg University, Aalborg University, Denmark.

Plank, T., and Ellermeier, W. (**2003**). "Discrimination of temporal loudness profiles," in *Fechner Day 2003. Proceedings of the 19th Annual Meeting of the International Society for Psychophysics*, edited by B. Berglund and E. Borg, Stockholm, Sweden, pp. 241–244.

Richards, V. M. (**2002**). "Varying feedback to evaluate detection strategies: The detection of a tone added to noise," J. Assoc. Res. Otolaryngol. **3**, 209–221.

Saberi, K. (**1996**). "Observer weighting of interaural delays in filtered impulses," Percept. Psychophys. **58**, 1037–1046.

Sadralodabai, T., and Sorkin, R. D. (**1999**). "Effect of temporal position, proportional variance, and proportional duration on decision weights in temporal pattern discrimination," J. Acoust. Soc. Am. **105**, 358–365.

Southworth, C., and Berg, B. G. (**1995**). "Multiple cues for the discrimination of narrow-band sounds," J. Acoust. Soc. Am. **98**, 2486–2492.

Stecker, G. C., and Hafter, E. R. (**2000**). "An effect of temporal asymmetry on loudness," J. Acoust. Soc. Am. **107**, 3358–3368.

Stecker, G. C., and Hafter, E. R. (**2002**). "Temporal weighting in sound localization," J. Acoust. Soc. Am. **112**, 1046–1057.

Stellmack, M. A., and Viemeister, N. F. (**2000**). "Observer weighting of monaural level information in a pair of tone pulses," J. Acoust. Soc. Am. **107**, 3382–3393.

Surprenant, A. M. (**2001**). "Distinctiveness and serial position effects in tonal sequences," Percept. Psychophys. **63**, 737–745.

Tang, Z., Richards, V. M., and Shih, A. (**2005**). "Comparing linear regression models applied to psychophysical data," J. Acoust. Soc. Am. **117**, 2597.

Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.

Willihnganz, M. S., Stellmack, M. A., Lutfi, R. A., and Wightman, F. L. (**1997**). "Spectral weights in level discrimination by preschool children: Synthetic listening conditions," J. Acoust. Soc. Am. **101**, 2803–2810.

Zwicker, E. (**1977**). "Procedure for calculating loudnesss of temporally variable sounds," J. Acoust. Soc. Am. **62**, 675–682.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and models* (Springer, Berlin).

Behavioral and physiological correlates of temporal pitch perception in electric and acoustic hearing

Robert P. Carlyon,[a] Suresh Mahendran, John M. Deeks, and Christopher J. Long
MRC Cognition & Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Patrick Axon and David Baguley
Addenbrooke's NHS Trust, Hills Road, Cambridge, United Kingdom CB2 2QQ

Stefan Bleeck[b] and Ian M. Winter
Physiological Laboratory, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 7 March 2007; accepted 15 November 2007)

In the "4–6" condition of experiment 1, normal-hearing (NH) listeners compared the pitch of a bandpass-filtered pulse train, whose inter-pulse intervals (IPIs) alternated between 4 and 6 ms, to that of isochronous pulse trains. Consistent with previous results obtained at a lower signal level, the pitch of the 4–6 stimulus corresponded to that of an isochronous pulse train having a period of 5.7 ms—longer than the mean IPI of 5 ms. In other conditions the IPI alternated between 3.5–5.5 and 4.5–6.5 ms. Experiment 2 was similar but presented electric pulse trains to one channel of a cochlear implant. In both cases, as overall IPI increased, the pitch of the alternating-interval stimulus approached that of an isochronous train having a period equal to the mean IPI. Experiment 3 measured compound action potentials (CAPs) to alternating-interval stimuli in guinea pigs and in NH listeners. The CAPs to pulses occurring after 4-ms intervals were smaller than responses to pulses occurring after 6-ms intervals, resulting in a modulated pattern that was independent of overall level. The results are compared to the predictions of a simple model incorporating auditory-nerve (AN) refractoriness, and where pitch is estimated from first-order intervals in the AN response. [DOI: 10.1121/1.2821986]

PACS number(s): 43.66.Hg, 43.66.Ts, 43.64.Pg [AJO] Pages: 973–985

I. INTRODUCTION

A. Background

In normal, acoustic hearing, the pitch of a complex tone is dominated by the lower-numbered harmonics, which are resolved by the peripheral auditory system (Plomp, 1967; 1985). The reasons for this domination remain unclear, but may arise from one or more of the following: (i) the presence of place-of-excitation cues (ii) the existence of a "match" between the place of excitation and the frequency of phase locking (Moore, 1989; Oxenham *et al.*, 2004; Moore and Carlyon, 2005), (iii) superior phase locking to fine structure than to the envelope (Moore *et al.*, 2006), and (iv) differences in the relative timing of the responses of different auditory nerve (AN) fibers (Loeb *et al.*, 1983; Shamma, 1985; Loeb, 2005; Moore and Carlyon, 2005). These timing differences are reflected by a steep transition in the function relating the phase of AN firing to place along the cochlea, and the place at which it occurs may code the frequency of pure tones or of resolved harmonics (Kim *et al.*, 1980; Loeb *et al.*, 1983; Shamma, 1985; Loeb, 2005; Moore and Carlyon, 2005).

The present article investigates, instead, pitches that can only be conveyed by the temporal response of AN fibers tuned to the frequency components of the stimulus. Stimuli that elicit this "purely temporal" pitch can be produced by bandpass filtering an acoustic pulse train so that it contains only high-numbered, unresolved harmonics, or by presenting an electric pulse train to one channel of a cochlear implant ("CI": McKay and Carlyon, 1999; Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003). As we have pointed out before (Carlyon *et al.*, 2002), such stimuli, although producing a fairly weak pitch, are of both theoretical and practical interest. First, by restricting the number of peripheral cues available, they can provide a simple test of more general models of pitch perception. Second, cochlear implants encode fundamental frequency (F0) using this purely temporal code, and understanding it may provide a basis for improving the generally poor pitch percepts experienced by CI users (McDermott, 1997; Moore and Carlyon, 2005). Third, the extent to which similar patterns of results can be obtained with acoustic and electric stimuli may allow one to develop an accurate simulation of CI hearing using normal-hearing (NH) listeners. Such simulations could, for example, be useful in the development of novel signal-processing strategies and/or experimental produres.

Another feature of the stimuli used to study purely temporal pitch perception is that a comparison of behavioral data with the response of the AN is much more straightforward than when resolved harmonics are present. For acoustic pulse trains lacking resolved harmonics, one can ignore place-of

[a]Author to whom correspondence should be addressed. Electronic mail: bob.carlyon@mrc-cbu.cam.ac.uk.
[b]Current address: Institute of Sound and Vibration Research, University of Southampton, S017 1BJ England.

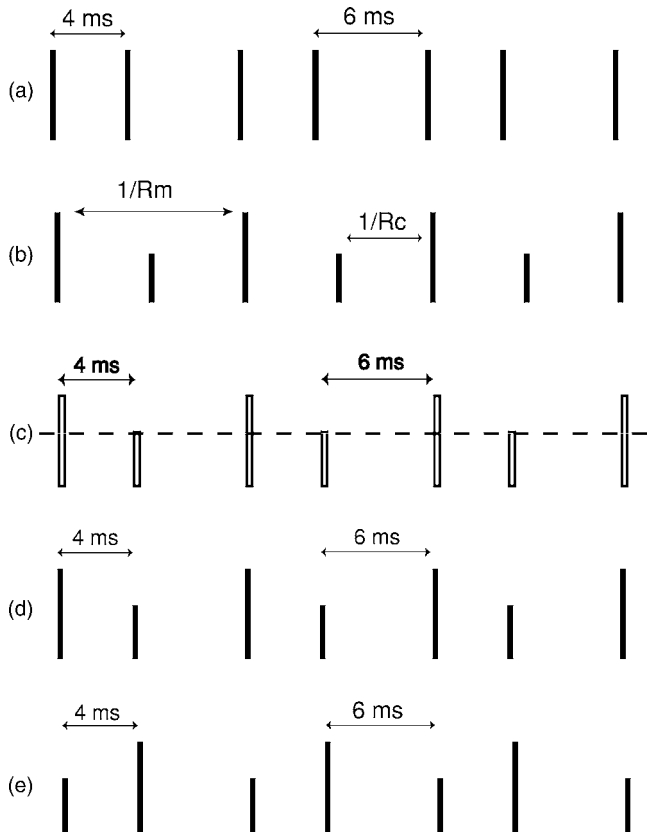


FIG. 1. Solid bars show schematic illustrations of some of the stimuli used in this and other studies. Only the first seven pulses in each train are shown. The open bars in part (c) illustrate a possible pattern of CAP responses.

excitation cues, and it is likely that there is no consistent cue conveyed by the relative timing of the responses of different AN fibers: for example, the steep phase transition observed across the AN fiber array for resolved partials (Kim *et al.*, 1980) is absent, and all fibers that do respond to each pulse probably do so at the same time (Carlyon and Shamma, 2003). Hence, a good estimate of the information conveyed by the AN can be obtained from the whole-nerve response to each pulse, as measured by the compound action potential (CAP). We exploit this fact to compare AN and behavioral responses to very similar stimuli. Such a technique would not be possible when stimuli consist of resolved harmonics; the temporal smoothing produced by peripheral filtering would prevent the measurement of CAPs throughout the stimulus, and potentially important information on place of excitation and on the relative timing of AN responses would be lost. To the extent that the behavioral results are similar for NH and CI listeners, this method also provides an indirect way of studying physiological correlates of temporal pitch perception in electric hearing.

The behavioral experiments that we report here exploit and extend a paradigm previously described by Carlyon *et al.* (2002). They asked both NH listeners and CI users to compare the pitch of a pulse train whose inter pulse intervals (lPIs) alternated between 4 and 6 ms (Fig. 1(a)) with that of a range of isochronous pulse trains, in which the IPI was constant throughout the stimulus. They found very similar results with the two groups of listeners: the pitch of the "4–6" stimulus corresponded to that of an isochronous train having an IPI of about 5.7 ms. They noted that this match was longer than the mean interval (5 ms) of the 4–6 stimulus, and shorter than its 10 ms period. They proposed a model that could account for the pitch of these stimuli, and also for the results obtained by themselves and others using several different paradigms (Carlyon, 1997; Plack and White, 2000; Carlyon *et al.*, 2002). The model assumed that only the first-order intervals in the stimulus determined pitch, and that longer intervals received greater "weights" than shorter ones.

Carlyon *et al.*'s (2002) model successfully accounted for a wide range of findings using a single set of parameters. Furthermore, the idea that the pitches of pulse trains are dominated by first-order intervals is consistent with the conclusions from a number of other recent studies (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001; Yost *et al.*, 2005). However, it has a number of limitations. One of these is illustrated by a study by McKay and Carlyon (1999). They presented both NH and CI listeners with a set of pulse trains, each of which underwent amplitude modulation (AM) by increasing the level of every *n*th pulse. Figure 1(b) illustrates the fact that such stimuli can be characterized as having a carrier rate (R_c) and a modulator rate (R_m). By performing a multi-dimensional scaling (MDS) experiment using stimuli having different combinations of R_m and R_c McKay and Carlyon showed that listeners were sensitive to both the carrier and modulation rates. Carlyon *et al.*'s (2002) model, however, produces a single pitch value, and cannot account for this finding. Indeed, the fact that subjects are at all sensitive to the modulator rate implies that pitch cannot be entirely determined by first-order intervals in the *stimulus*, at least when some pulses have a higher amplitude than others. Not surprisingly, when subjects are forced to match the pitch of a modulated pulse train, then, as the modulation depth increases from 0 to 100%, the match decreases from the carrier to the modulator rate (McKay *et al.*, 1995; McKay and Carlyon, 1999).

Here, we investigate the hypothesis that a similar phenomenon can occur for equal-amplitude pulse trains, provided that the *auditory nerve* response, as measured by the CAP, is amplitude modulated. Specifically, for a 4–6 pulse train, the response to pulses occurring after a 4-ms interval may be smaller than that to those occurring after a 6 ms interval (Fig. 1(c)). A simple model, incorporating this effect, would predict the pitch as follows: If there exists an array of more-central neurons that fire only when the CAP amplitude exceeds a given threshold (e.g., when a threshold number of AN fibers fire synchronously), and if some of those neurons have thresholds higher than the response to pulses occurring after 4-ms intervals (Fig. 1(c), dashed line), then those neurons will fire every 10 ms. The remainder, having thresholds lower than this criterion, will fire after every pulse. Pitch matches may then be obtained by a simple average of the first-order intervals in the responses of these two sets of more-central neurons. For the 4–6 stimulus, this would be a combination of 4 and 6 ms intervals (lower-threshold neurons) and of 10 ms intervals (higher-threshold neurons). This general scheme would be consistent with the results of van Wieringen *et al.* (2003), who used alternating-interval pulse

trains in which the pulses after either the short or the long intervals were attenuated (Figs. 1(d), 1(e)). They found that pitch was lower when the pulses after the shorter intervals were attenuated (Fig. 1(d)), consistent with the neural response to each lower-amplitude pulse occurring after short IPIs being further attenuated by refractory effects originating from the previous, higher-amplitude pulse. This pitch difference decreased with increasing overall IPI (e.g., from 4–6 ms to 8–12 ms), both in NH and CI listeners, consistent with an explanation based on refractoriness, if one further assumes that the recovery function starts to flatten off over this range of delays. This general scheme would also be consistent with McKay and Carlyon's (1999) finding that listeners can perceive both the carrier and modulator rates of AM stimuli, if one assumes that they can selectively "attend" to different subsets of the more-central neurons. At the same time, it would also be consistent with reports that higher-order intervals do not have a large effect on the pitches of pulse trains that do not produce large and/or regular modulations in the AN response (Kaernbach and Demany, 1998; Plack and White, 2000; Kaernbach and Bering, 2001; Yost *et al.*, 2005).

Two further points are worth making. First, Pressnitzer *et al.* (2001; 2004) have argued that higher-order intervals between *pulses* can be transformed into first-order intervals between *spikes* in the response of AN and cochlear-nucleus neurons. As with the simple model proposed here, their research emphasizes the need to consider the input to the pitch mechanism in terms of neural activity, rather than solely considering the statistics of the stimulus. Second, we should stress that our simple model assumes that a summary statistic is derived from the summed response of multiple fibers, rather than assuming that a statistic (such as the inter-spike-interval histogram) is derived from each fiber, with these individual statistics then being combined. This distinction is important, because it has been argued, based on the results of single-unit recordings (Cariani and Delgutte, 1996), that a temporal code based on first-order intervals should depend strongly on overall level. However, Carlyon *et al.* (2002) have argued that this should not necessarily be the case when the responses of several neurons are combined before a summary statistic is derived.

B. Overview of model and experiments

The aim of the experiments described here was to compare behavioral measures of temporal pitch perception in NH and CI users to the neural response as measured by the CAP. In particular, we wished to determine the extent to which the behavioral measures could be accounted for by the type of simple neural model described in the previous subsection. Specifically, we compare the results to the predictions of a model in which the thresholds of the "more central" neurons are uniformly distributed across level, and where the predicted pitch is obtained from an unweighted sum of the first-order intervals at the output of this more-central population. For example, if the CAP to a 4–6 pulse train were amplitude modulated by 10%, then 10% of the "more central" neurons would fire every 10 ms, and 90% would follow the 4–6 pattern. The period corresponding to the pitch would then be $(0.1*10)+(0.45*4)+(0.45*6)=5.5$ ms.

The model described above assumes that the CAP provides an accurate measure of the whole-nerve input to the more-central neural population, and that the time window over which this input is summed corresponds simply to the smoothing inherent in the CAP measurement, which likely derives from the integrative properties of the inner-hair-cell and AN-fiber membranes. As noted in the Introduction, this approach allows a direct comparison of behavioral results to physiological measures obtained using very similar stimuli under conditions where, unlike the case where resolved harmonics are present, the neural code is limited to purely temporal cues.

A "first pass" test of the neural model is, of course, that the CAP response to an alternating-interval pulse train is indeed amplitude modulated. To test this, experiment 3a measured CAPs to bandpass filtered 4–6 pulse trains in five anesthetized guinea pigs (GPs) and from two NH listeners. The results showed that the predicted form of modulation was indeed present. A more stringent test comes from the requirement that pitch be largely level independent. As noted above, we have previously argued that a statistic that is derived from the responses of multiple fibers may be robust to changes in overall level, and this feature would be needed to avoid the prediction that pitch changed markedly with level. The results of experiment 3a showed that the modulation in the CAP to 4–6 pulse trains was indeed largely independent of level over a 50-dB range. In addition, the results of experiment 1 showed that the pitch of 4–6 stimuli was judged by NH listeners to be similar to that of an isochronous pulse train having a period of about 5.6 ms, a result very similar to that obtained previously at a 24 dB lower level (Carlyon *et al.*, 2002).

A yet more demanding test of the model is that, as stimulus parameters are manipulated, the behavioral results should quantitatively follow the predictions. Experiment 1 also included two new conditions in which the 4–6 pulse train was replaced by one with slightly shorter ("3.5–5.5") or longer ("4.5–6.5") intervals. These conditions were originally included because we wished to measure CAPs to alternating-interval pulse trains, and wanted to avoid stimuli whose F0 (which was 100 Hz for the 4–6 train) was a harmonic of the 50 Hz U.K. mains frequency. The inclusion of such stimuli also led to an interesting prediction. One would, of course, expect the matched pitch to correspond to longer intervals for a 4.5–6.5 train than for a 3.5–5.5 train. In addition, if the *relative* change in refractoriness between the long and the short intervals is smaller at longer overall IPIs, then one might also expect a decrease in the proportion of central neurons responding only to every other pulse. This in turn would cause the matched pitch to be closer to the mean of the two intervals in the alternating-interval stimulus; that is, closer to 5.5 ms for the 4.5–6.5 stimulus than to 4.5 ms for the 3.5–5.5 stimulus. This hypothesis was also tested in experiment 2 with CI listeners, using 4–6 and 5–7 pulse trains. The hypothesis was confirmed behaviorally with both groups of listeners. However, the pattern of results was not reflected

by differences in the modulation depth of the CAP response to the 3.5–5.5, 4–6, and 4.5–6.5 stimuli, obtained in Experiment 3b from three additional GPs. We conclude that, although the general form of model described here can account qualitatively for a wide range of data, the physiological data obtained from the GP AN does not account for the effects of varying inter-pulse interval. Two possible explanations for this discrepancy—species differences and the existence of an additional source of refractoriness—are discussed, and, in the latter case, a quantitative estimate of the additional refractoriness needed is presented. The aim of all these studies was not to disprove Carlyon *et al.*'s (2002) model, but rather to see whether the refractory properties of the auditory nerve would allow a more simple explanation of the data that would dispense with the need for a central weighting function.

II. EXPERIMENT 1: TEMPORAL PITCH STUDIED WITH NH LISTENERS

A. Method

All pulse trains were generated digitally in the time domain and played out through a 16 bit digital-to-analog converter (DAC; CED 1401*plus* laboratory interface) at a sampling rate of 50,000 Hz. They were then passed through a reconstruction filter (Kemo VBF25.01; 100 dB/octave) and bandpass filtered between 3900 and 5400 Hz using a lowpass and a highpass eighth-order Butterworth filter in series (Kemo VBF25.03; 48 dB/octave). The duration of each pulse train was 400 ms, including 10-ms raised-cosine ramps. The level of every pulse train was 78 dB SPL. This was higher than the 54 dB SPL used by Carlyon *et al.* (2002), in order to aid comparison with the CAPs to the same stimuli in experiment 3, for which a higher level was considered desirable in order to obtain a more robust neural response. Pulse trains were then attenuated (Tucker-Davis Technologies PA2) and mixed with a pink noise. The noise was gated on and off synchronously with the pulse train and was played out of a second DAC. A fresh 400-ms sample of noise was selected for each presentation by sampling from a random point in a previously generated 2-s wave file (CoolEdit 2000, Syntrillium software Inc). The noise was bandpass filtered between 100 and 3900 Hz (Kemo VBF25.03 highpass and lowpass filters in series; attenuation 48 dB/octave), attenuated (TDT PA2), and mixed with the pulse train. Its spectrum level at 1000 Hz was 42 dB SPL. Stimuli were then presented via one earpiece of a Sennheiser HD250 headset to a listener seated in a double-walled sound-attenuating booth. Calibration was performed with the aid of a B&K type 4153 artificial ear and an HP3561A spectrum analyzer.

There were three conditions, defined by the durations of the IPIs in the alternating-interval stimulus: 3.5–5.5 ms, 4–6 ms, and 4.5–6.5 ms. In each trial of each condition, the listener heard the alternating-interval stimulus and one of five isochronous pulse trains, presented in random order. The IPIs for the isochronous trains in the 3.5–5.5 condition were 2.5, 3.5, 4.5, 5.5, and 6.5 ms. In the 4–6 and 4.5–6.5 conditions these values were increased by 0.5 and 1 ms, respectively. The listener was instructed to identify which of the two stimuli in the trial had the higher pitch by clicking on one of two virtual buttons on a computer monitor. No feedback was provided. Each listener performed ten repeats (X 5 isochronous stimuli=50 trials) for each condition before moving on to the next one. The standard stimulus started with the shorter of its two IPIs for five of these repeats and with the longer IPI for the other five. No differences were observed between the results obtained with these two types of trial, and they were therefore averaged. Each condition was then repeated in the same order until each listener had completed 200 trials per data point, with the exception of listeners NH2 and NH7, who completed 150 and 160 trials, respectively. Seven NH listeners participated in the experiment.

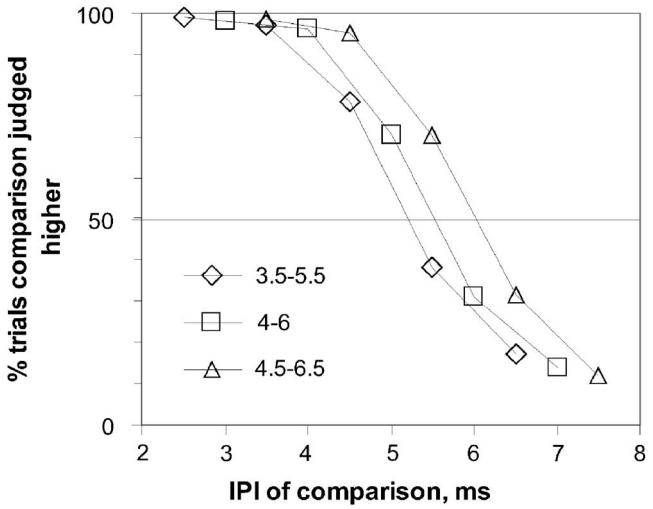


FIG. 2. Psychometric functions showing the percentage of trials in which the isochronous comparison sound, whose period is given on the abscissa, was judged higher than 3.5–5.5 (diamonds), 4–6 (squares), and 4.5–6.5 (triangles) standard stimuli. Data are averaged across the NH listeners of experiment 1.

B. Results

Each psychometric function in Fig. 2 shows the proportion of trials, averaged across listeners, on which each isochronous pulse train was judged higher in pitch than the alternating-interval stimulus in one condition. It can be seen that, for each alternating-interval stimulus, the psychometric function spans the range from below 20% to above 95%. The functions for stimuli with longer IPIs (e.g., 4.5–6.5) are to the right of those with shorter IPIs (e.g., 3.5–5.5), showing that their perceived pitch corresponded to a longer IPI, and was therefore lower. To estimate the period of an isochronous stimulus judged equal in pitch to each standard, the psychometric function for each subject and condition was subjected to a probit analysis. The point of subjective equality ("PSE"), corresponding to the point at which the fitted function passed through 50% intercept on the ordinate, is shown for each subject and condition in Fig. 3(a); mean data are shown by the thick dashed curve with square symbols. Not surprisingly, the PSE increases as the IPIs in the alternating-interval standards increase from 3.5 to 4.5 through 4–6 to 4.5–6.5 ms. For the 4–6 stimulus, the mean

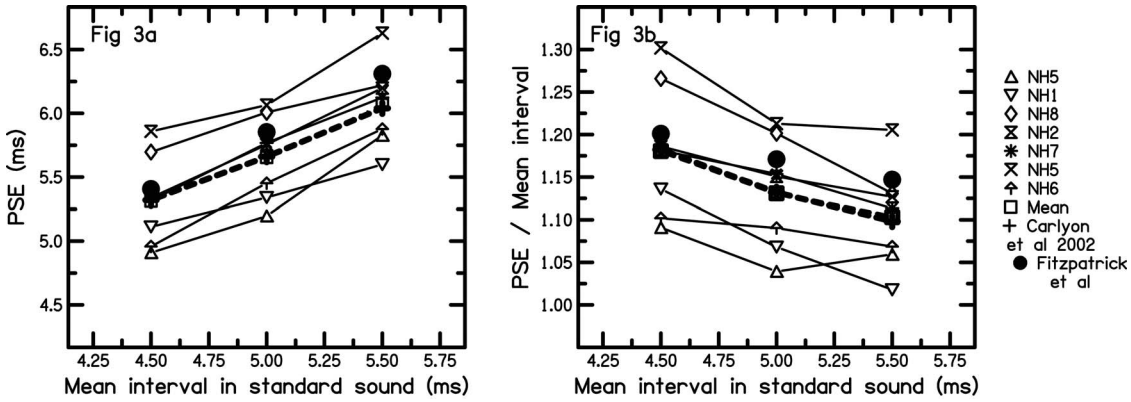


FIG. 3. Part (a) shows the Point of Subjective Equality ("PSE") derived from the psychometric functions of experiment 1, for 7 NH listeners. The abscissa shows mean interval in each of the three standard sounds tested. Mean data are shown by the heavy dashed line joining squares. The prediction of Carlyon *et al.*'s (2002) model is shown by the heavy dashed line joining "plus" signs. These two heavy curves overlap, testifying to the success of the model. Part (b) shows the same data, with the PSEs divided by the mean interval in each standard. In both parts of the figure, predictions based on the recovery function described by Fitzpatrick *et al.* (1999) are shown by filled circles.

PSE is 5.64 ms, close to the 5.7 ms reported by Carlyon *et al.* (2002) for stimuli presented at a softer overall level (54 vs. 78 dB SPL). Figure 3(b) shows the PSE in each condition divided by the average IPI in the standard for that condition. In the absence of refractory effects, our simple, unweighted model, would predict a ratio of one. The ratio decreases from 1.18 for the 3.5–5.5 stimulus to 1.10 for the 4.5–6.5 stimulus, consistent with the change in refractoriness between 4.5 and 6.5 ms being smaller than that between 3.5 and 5.5 ms. This trend was confirmed by a one-way analysis of variance (ANOVA) ($F(2,10)=15.8$, $p<0.001$, Huynh-Feldt sphericity correction). We should also note that, when averaged across listeners, the results are consistent with Carlyon *et al.*'s (2002) model, the predictions of which are shown by the heavy dashed line with "plus" symbols in Figs. 3(a) and 3(b). These lines are superimposed on the mean data (dashed line, squares), reflecting the good fit of the model. The solid circles without lines will be discussed in Sec. V A.

III. EXPERIMENT 2: TEMPORAL PITCH STUDIED WITH CI USERS

A. Method

The method used for experiment 2 was generally similar to that for experiment 1. An important difference is that, instead of presenting filtered acoustic pulse trains to NH listeners, we presented electric pulse trains via a bipolar pair of intracochlear electrodes of a CI. Five listeners took part, all of whom had been implanted with either the CI22 or CI24 device manufactured by Cochlear Corp. The listeners' details, including information on the device used by each of them, are given in Table I. All stimulation was on electrode 17, with electrode 13 serving as the return electrode. This corresponds to so-called "BP+3" mode, with approximately 3 mm between electrodes. Stimuli consisted of 400-ms trains of biphasic pulses, with each pulse having a phase duration of 100 μs and an inter-phase gap of 8 μs. Standard stimuli consisted of 4–6 and 5–7 pulse trains. The isochronous stimuli to be compared to the 4–6 standard had IPIs of 3, 4, 5, 6, and 7 ms; those to be compared to the 5–7 ms standard had IPIs of 4, 5, 6, 7, and 8 ms.

At the start of the experiment, the threshold and most-comfortable ("C") level was obtained for the 4–6 stimulus for each subject, using the same electrodes and configuration as in the main experiment. The 5–7 standard was then loudness balanced to the 4–6 standard using the procedure similar to that described by McKay and Carlyon (1999). One of the two stimuli was presented first, followed 500-ms later by the second stimulus at a level that was ten clinical current units (approx 1.76 dB) lower. The subject could then adjust the level of the second sound to be presented on the next trial, by pressing one of six virtual buttons (labeled '+++', '++', '+', '−', '−−', and '−−−') on a computer screen. This procedure continued until the subject was satisfied that the two stimuli had equal loudness. It was then repeated. The roles of the fixed and standard stimuli were then swapped, and the procedure repeated twice. The average difference between the levels of the two stimuli over these four runs

TABLE I. Details of the cochlear implant patients who took part in experiment 2. CSOM refers to chronic serous otitis media. NI refers to noise-induced hearing loss.

Subject	Age (years)	Etiology	Age of Onset	Implant date	Device
CI 1	58	Familial	47 years	Nov. 1999	CI 24M
CI 2	68	Meniere's/CSOM	Progressive	June 1996	CI 22
CI 3	38	Labyrinthitis	15 years	Nov. 1995	CI 22
CI 4	64	Idiopathic	18 years	April 1996	CI 22
CI 5	77	Otosclerosis / NI	22 years	March 2001	CI 24M

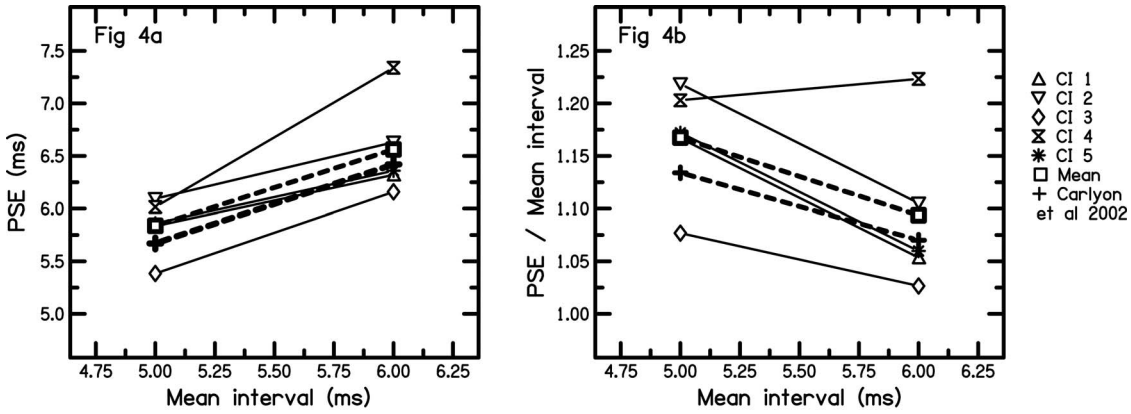


FIG. 4. As in Fig. 3, except for the five CI listeners of experiment 2.

was used to equate their loudness. Each standard was then loudness balanced to the isochronous pulse trains having the longest and shortest IPIs to be compared to it (e.g., 3 and 7 ms for the 4–6 standard). Levels for isochronous stimuli having intermediate IPIs were obtained via linear interpolation in clinical current units (CUs), where one CU is equal to approximately 0.176 dB. This was deemed reasonable because loudness does not vary markedly with level over the range of IPIs studied here (McKay and McDermott, 1998; Carlyon *et al.*, 2002).

B. Results

The PSEs for each condition were obtained in the same way as for experiment 1 and are shown in Fig. 4(a). Consistent with the results of that experiment, we obtained the unsurprising finding that the PSE was longer on average for the 5–7 than for the 4–6 standard. More interesting is the fact that, as shown in Fig. 4(b), the ratio of the PSE to the mean interval in the standard was lower for the 5–7 than for the 4–6 stimulus. This finding was obtained for four out of the five listeners, and was significant overall, as revealed by a paired-samples t test ($df=4$, $p<0.05$, two tailed). This result is consistent with that obtained in experiment 1, and with the idea that refractory effects influence temporal pitch perception in electric and acoustic hearing in a roughly similar way. Again, however, we should note that the results are also roughly consistent with the central weighting function proposed by Carlyon *et al.* (2002) (heavy dashed line and "plus" symbols). Presumably, however, any central mechanism will operate on the AN response rather than on the physical stimulus. The aim of the next experiment was to determine whether refractory properties of the AN would produce a peripheral representation that would allow one to dispense with the need for such a weighting function.

IV. EXPERIMENT 3: CAP MEASUREMENTS

A. Subjects

Experiment 3a measured CAPs to the same stimuli in anesthetized guinea pigs ("GPs") and in (human) NH listeners. The former group consisted of five pigmented GPs with weights between 330 and 585 g and CAP thresholds within 5 dB of the norms obtained in author I.M.W.'s laboratory, where the GP recordings were obtained. The latter group consisted of five normal-hearing adults, including subjects NH1, NH2, and NH3 from experiment 1. However, for reasons that will be discussed in Sec. IV D, it was only possible to record reliable responses to each pulse in a train from listeners NH1 and NH7. Experiment 3b measured CAPs to equal-amplitude and amplitude-modulated pulse trains from an additional three GPs.

B. Recording

1. GPs

The method of recording was as described by Neuert *et al.* Briefly, GPs were anesthetized with urethane [1.5 g/kg, intraperitoneally (ip)]. Hypnorm was administered as supplementary analgesia (1 mg/kg, intramuscularly) i.m.). Anesthesia and analgesia were maintained at a depth sufficient to abolish the pedal-withdrawal reflex (front paw). Additional doses of Hypnorm (1 ml/kg) or urethane (1 ml) were administered on indication. Incisions were preinfiltrated subcutaneously with the local anesthetic Lignocaine (Norbrook Laboratories, Newry, UK). Core temperature was monitored with a rectal probe and maintained at 37 °C using a thermostatically controlled heating blanket (Harvard Apparatus, Holliston, MA). The trachea was cannulated, and the animal was ventilated artificially with a pump if it showed signs of suppressed respiration. Surgical preparation and recordings took place in a sound-attenuated chamber (Industrial Acoustics). The animal was placed in a stereotaxic frame that had ear bars coupled to hollow speculas designed for the guinea pig ear. A midsagittal scalp incision was made, and the periosteum and the muscles attached to the temporal and occipital bones were removed. The bone overlaying the left bulla was fenestrated, and a silver-coated wire was placed on the round window of the cochlea to record the CAPs. The hole was resealed with petroleum jelly. The wire electrode was connected via an amplifier (WPI DAM 50, gain=X 10,000) to an interface box (Hammerfall DSP multiface MIDI 24 Bit 96 kHz multichannel interface), and then stored on a PC via an interface card (RME Intelligent Audio Solutions Hammerfall DSP system-HDSP cardbus interface), for offline analysis. All GP experiments were performed in accordance with terms and conditions of the project licences issued by the U.K. Home Office to author I.M.W.

2. Human NH listeners

Pre-test examination and placement of electrodes for the human subjects was similar to that used in standard clinical electrocochleography. The active (recording) electrode was a soft-tipped electrode (Bio-logic Systems Corp TM-ECochGtrode) placed gently on the surface of the tympanic membrane. Along with single-use common (forehead) and reference (contralateral mastoid) electrodes (SLE diagnostics, Ref. M0872) it was connected to one channel of the head box of a Digitimer D360 8-channel patient amplifier (Digitimer Ltd., U.K.). The Digitimer amplifier was set to a passband of 70–1500 Hz, a gain of 50,000, and the notch filter turned on to eliminate 50 Hz (U.K. mains frequency) hum. The output of the amplifier was then connected to the interface box, and the results stored on the PC for offline analysis. For safety reasons, all equipment with the exception of the patient amplifier was battery powered.

C. Stimulus generation

The method of stimulus generation was similar for both groups of subjects. In experiment 3a it was identical to that described in experiment 1 with the following exceptions: (i) Filtering was carried out in software, using eighth-order low-pass and highpass Butterworth filters, in series. (ii) The sampling rate was 96 kHz, (iii) the duration of each pulse train was 100 ms. (iv) There were no onset and offset ramps, and (v) no pink noise was presented. The pink noise was used in experiment 1 to prevent listeners from using cochlear distortion products, which we did not expect to affect the CAP. Only the 4–6 and 6–4 pulse trains were presented. The stimulus level was 78 dB SPL for the NH subjects. For the GPs, a range of levels between 38 and 88 dB SPL, in 10-dB steps, were tested for the 4–6 stimulus. Additionally, measurements for the 6–4 stimulus were obtained at 88 dB SPL for all five GPs, and at 78 dB SPL for GP2, GP3, and GP4.

In experiment 3b, CAPs were obtained in three further GPs for an additional set of stimuli, presented at a level of 78 dB SPL. These consisted of equal amplitude pulse trains in which the odd- and even-numbered intervals, in ms, were 3.5–5.5, 5.5–3.5, 4–6, 6–4, 4.5–6.5, 6.5–4.5, 8–12, and 12–8. In addition, CAPs were obtained for 4–6 and 8–12 pulse trains in which the pulses occurring after the short (4 or 8 ms) or long (6, 12 ms) intervals were attenuated by 2 or 6 dB (Figs. 1(d) and 1(e)). These amplitude-modulated pulse trains, which resemble a subset of those used by van Wieringen *et al.* (2003) are designated here by the letter "S" (for "short") or "L" (for "long"), indicating the intervals after which pulses were attenuated, followed, optionally, by a number indicating the attenuation in dB. (For example, 4-6S2 is a 4–6 stimulus in which all pulses after the shorter (4 ms) interval are attenuated by 2 dB). The final number is omitted whenever we refer to a general class of stimulus without specifying the amount of attenuation used.

Waveform files were transferred from a PC to the same interface box as used for response collection. For the GPs they were played out via a power amplifier (Rotel RB971 Mk 2) and a custom-built end attenuator before being presented over a speaker (30-1777 tweeter: RadioShack, Fort-Worth, Texas) that was mounted in a coupler designed for the GP ear. Stimuli were acoustically monitored using a condenser microphone (B&K 4134) attached to a calibrated l-mm-diam probe tube that was inserted into the speculum, close to the eardrum. For NH subjects they were presented via a headphone amplifier to one earpiece of an Etymotic ER-3 insert phone. Because the extra-tympanic electrode used for the normal-hearing human subjects was likely to reduce the sound pressure level at the eardrum, the following procedure was adopted to correct for this. Detection thresholds for a 5-kHz pure tone in quiet were measured using a two-interval forced-choice task and the adaptive procedure described by Levitt (1971). Two adaptive runs were obtained and the results averaged. This procedure was performed before and after placement of the electrode, using Sennheiser HD-250 headphones, and the difference between the thresholds obtained (mean=9.4, s.e.=1.7 dB) was used as an estimate of the attenuation produced by the electrode. The sound pressure level delivered by the insert earphone during the CAP recordings was then increased to compensate for this attenuation.

In each recording run, 100-ms presentations of a given pulse train were presented repeatedly, with a 50-ms silent interval between bursts. Every other burst was inverted in polarity, in order to reduce the influence of stimulus artifact and of cochlear microphonics when the responses were averaged. Typically we averaged responses to 600 stimuli for each GP and to 2000 or 3000 stimuli for the NH listeners. For the NH listeners, several additional conditions were also run. Subject NH1's responses were measured under four conditions: the 100-ms 4–6 and 6–4 pulse trains used for the GPs, and, for reasons unrelated to the present study, the same two pulse trains with a 90-ms duration; 2000 responses were obtained in each condition. Subject NH7 was tested with the 100 ms 4–6 and 6–4 stimuli, and with the interval between pulse trains increased to 300 ms; 3000 responses were obtained in each condition.

D. Results

1. Experiment 3a: GPs

CAPs obtained from GPs were very similar across animals. Figure 5(a) shows the CAP to the first pulse in a 4–6 train, obtained from one GP. It shows the typical form (e.g., Murnane *et al.*, 1998) consisting of a negative deflection followed by a positive deflection. Here, the amplitude of the CAP is defined as the difference, in μV, between these negative and positive peaks.

Figure 5(b) shows the response from the same GP to an entire 100-ms 4–6 pulse train at a level of 78 dB SPL. It shows an alternating pattern of 4 and 6 ms intervals, reflecting the IPIs present in the stimulus. The overall amplitude of the response decreases rapidly over the first three whole periods (30 ms), before reaching an asymptote (cf. Eggermont and Spoor, 1973). Averaged across GPs, the CAP amplitudes after 10, 20, and 30 ms were 65, 58, and 56% of that to the first pulse. The response to the last pulse in the stimulus had an amplitude that was 55% of that to the first.

Figure 5(c) shows a close-up of the response shown in

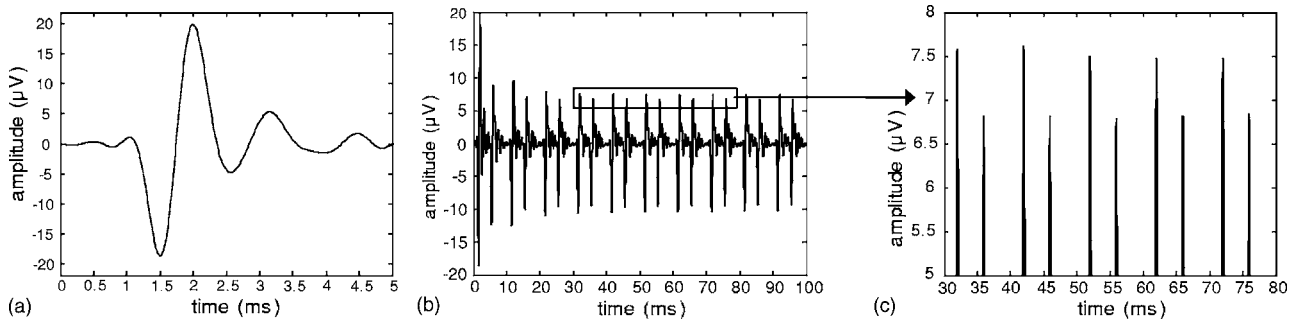


FIG. 5. Part (a) shows the CAP to a single pulse from one GP of experiment 3a. Part b) shows the response to a 78 dB SPL 4–6 pulse train in the same animal. The area shown by the dashed box is expanded and illustrated in part (c).

Fig. 5(b), focusing on the positive deflections between 30 and 80 ms. It can be seen that the responses after 4-ms intervals are smaller than those after 6-ms intervals. To quantify this difference while minimising the influence of short-term adaptation, we first excluded the responses to the first 30 ms of the pulse train.[1] We then separately averaged the response amplitudes of all pulses occurring after 4-ms and after 6-ms intervals. At 78 dB SPL the average response after 6-ms intervals was 11.0% (s.e=1.4%) greater than that after 4-ms intervals, a difference that was statistically significant (t test $df=4$, $p<0.05$)). Figure 6(a) shows that this ratio varied only between 7% and 11% over the range 33–88 dB SPL. The level independence of the AN response is illustrated further in Figs 6(b) and 6(c), which show the CAP waveforms obtained from GP2 at levels of 88 and 38 dB SPL, respectively. Although the CAP is smaller and slightly noisier at the lower stimulus amplitude, both the general form of the waveform and the amount of amplitude modulation are similar at the two levels.

2. Experiment 3a: NH human listeners

Figure 7(a) shows the response to a single pulse in subject NH2. Apart from the smaller amplitude, the CAP is similar to that obtained from GPs. A similar response was also obtained in the other four human subjects. However, when we measured CAPs to the pulse trains, it was only for subjects NH1 and NH7 that we saw a clear response to each pulse. The reasons for this are discussed in a separate report (Mahendran *et al.*, in press). One reason is that the response even to a single pulse is smaller than in the GPs, and adaptation both from previous pulse trains and throughout each pulse train may reduce the amplitude into the noise floor. Another is that, in the majority of subjects, the response to a single pulse was followed by a myogenic response having a latency sufficiently long (e.g., 20 ms) to be missed by the short time window over which CAPs are usually analyzed in clinical practice. Mahendran *et al.* suggested that the CAP to each pulse may be distorted by this longer-latency response to previous pulses. Neither subjects NH1 nor NH7 showed this long-latency response, which was also absent from our GP recordings.

NH7's CAPs to a 100-ms pulse train are shown in Fig. 7(b). The response shows a series of CAPs following each pulse in the stimulus, as is further illustrated by the zoomed-in plot with gridlines in Fig. 7(c). (Note that the vertical gridlines are separated by 4 and 6 ms, following the alternating intervals in the stimulus, and that the peaks in the CAP response are aligned with the gridlines). It can be seen that, as with the GPs, the response to pulses after 6-ms IPIs are larger than those to pulses after 4-ms IPIs. To quantify this difference, we averaged the responses after 4-ms and after 6-ms intervals in the same way as for the GPs. Averaged across the 4–6 and 6–4 stimuli, the responses after 6-ms intervals were 19.7% higher than those after 4 ms intervals. Rather than report a measure of inter-subject variability from only two subjects, we obtained a measure of the variability of the response within each subject. For subject NH7 this was done by, for the 4–6 and 6–4 trains separately, analyzing the responses to pulses n, $n+5$, $n+10$, etc., and obtaining five separate measures corresponding to $n=1,2,3,4$ and 5. Three thousand measures were obtained for each condition, so each submeasure corresponded to the average of 600 responses. Analyzing the data in this way, we obtained 95% confidence limits between 9.2 and 35.5%. Averaged across the four stimuli tested for subject NH1, responses after 6-ms intervals were 10.7% bigger than after 4-ms intervals. When the data from the four different stimuli were further subdivided into three interleaved sets of 666 responses, we obtained 95% confidence limits between 2.1 and 16.6%. It is worth noting that the confidence limits for both subjects are wide but do not encompass any negative values.

It is clear that, even in the two NH subjects from whom we obtained meaningful data, the responses to pulses throughout a train are much noisier than the results obtained from GPs. The wide confidence intervals mean that, for the NH listeners, we cannot obtain an accurate measure of the degree of amplitude modulation in the AN response. What the results do demonstrate is that the same general finding applies *qualitatively* to GPs and NH human subjects.

3. Experiment 3b: GPs

Experiment 1 showed that, for human NH listeners, the pitch match for a 4.5–6.5 stimulus was closer to the average IPI in that stimulus than was the case for the 4–6 and 3.5–5.5 stimulus. In terms of our simple model, this would be consistent with there being less modulation in the AN response at longer overall IPIs. To test this, we measured the average CAP after long vs. short intervals for 3.5–5.5, 4–6, and 4.5–6.5 stimuli in three GPs. We also obtained measures for an

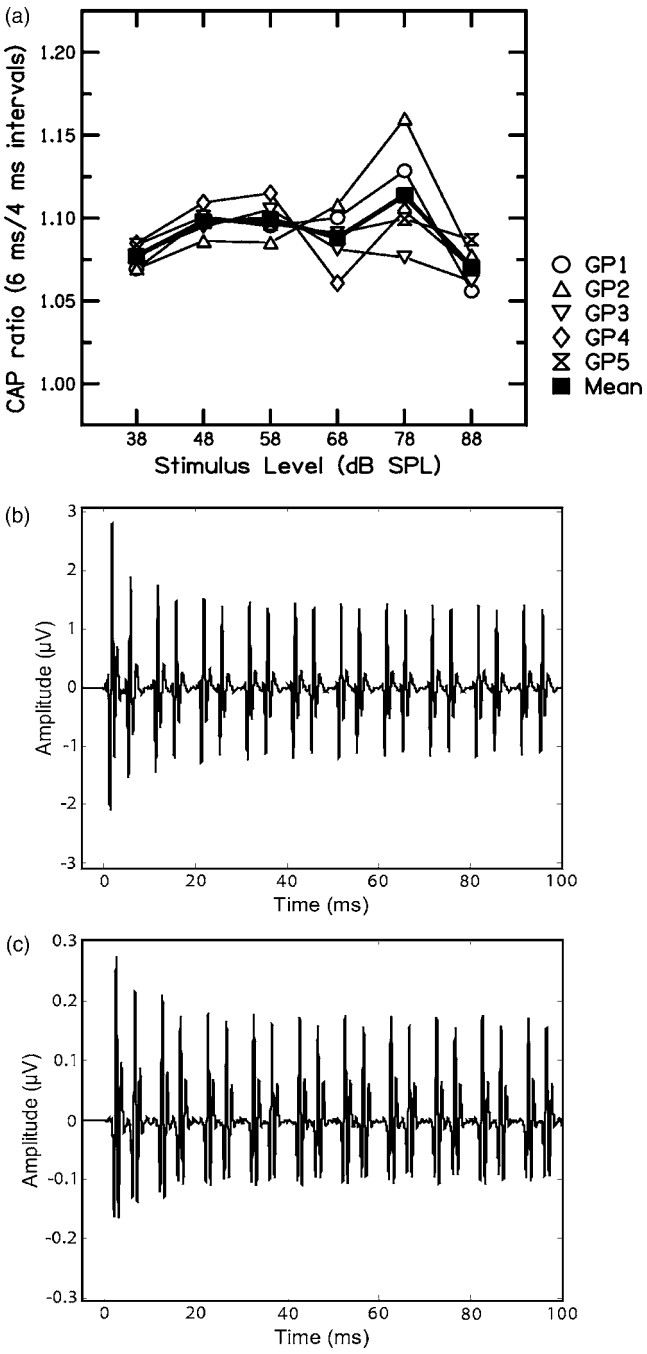


FIG. 6. Part (a): Lines connecting symbols show the ratio of CAP amplitudes after 6- vs. 4-ms intervals in a 4–6 pulse train, for each GP of experiment 3a, as a function of stimulus level. The heavy line without symbols shows the mean data. Parts (b) and (c) show the level independence of the modulation in the AN response by plotting the CAP waveform for GP2 at levels of 88 and 38 dB SPL, respectively.

"8–12" stimulus, in order to maximize the chances of seeing an effect of overall IPI. To minimize effects of short-term adaptation, data from the first three whole periods were excluded for all stimuli except 8–12, for which data from the first two periods (40 ms) were excluded. As in experiment 3a, these measures were also obtained for stimuli starting with the longer IPI (5.5–3.5, 6–4, 6.5–4.5, 12–8), and the results averaged. The ratio of the CAPs after the long vs. short intervals is shown for the three GPs in Table II(a). It can be seen that there is no tendency for the amount of modulation to decrease with increasing IPI over the range studied. Section V A 3 discusses possible reasons for this discrepancy.

Table IIb shows the percentage difference in CAP amplitude for pulses after the 6- vs 4-ms intervals for a subset of the stimuli similar to those used by van Wieringen *et al.* (2003). When the amount of attenuation is increased, the percentage CAP difference increases for the 4–6S stimuli, and decreases for the 4–6L stimuli. In the latter case, the difference becomes negative, reflecting the fact that the attenuation of pulses after the longer intervals overcomes the smaller refractory effects relative to pulses after the shorter intervals. Similar trends are seen for the 8–12 stimulus in Table IIc.

Although the data for the three GPs are quantitatively similar for the unmodulated stimuli (4–6 and 8–12), the effect of attenuating pulses after the longer or shorter intervals was much greater for GP6 than for the other two animals. We therefore restrict our discussion to describing two trends that were apparent in the data of all three GPs. First, as the modulation in the stimulus is increased from 2 to 6 dB, then, for both the 4–6S and 4–6L stimuli, the modulation in the CAP response also increases. This is consistent with van Wieringen *et al.*'s finding that subjects matched to longer periods for stimuli with larger modulation depths. Second, the differential effect of attenuating pulses after the longer vs. the shorter intervals was greater for the 4–6 than for the 8–12 stimuli; a three-way ANOVA (factors=4–6 vs. 8–12, attenuation amount, attenuation on long vs short) revealed a borderline interaction between the overall interval duration (4–6 vs. 8–12) and whether the attenuation was applied to pulses after the longer vs. the shorter interval ($F(1,2)=16.03$, $p=0.057$).

V. DISCUSSION

A. Models of temporal pitch perception

1. Carlyon et al. (2002)

Carlyon *et al.*'s 2002 model assumed that pitch was estimated using a weighted sum of first-order intervals in the stimulus. The model accounted for the fact that a 4–6 stimulus was matched in pitch to an isochronous sound having a period of 5.7 ms—longer than the 5-ms mean IPI in the 4–6 sound—by assuming that the weights increased with increasing IPI up to 10–12 ms. As shown in Fig. 3(b), it can also account for the fact that this tendency to produce a match longer than the mean IPI is greater for the 3.5–5.5 than for the 4–6 stimulus, and smallest of all for the 4.5–6.5 pulse train. It does so because the difference in weights between IPIs of 3.5 and 5.5 ms is greater than that between 4 and 6 ms, which in turn is greater than that between 4.5 and 6.5 ms. The model could also account for the data of Plack and White (2000), who presented listeners with sequences of eight filtered pulses, with an IPI of 4 ms. They found that delaying the last four pulses, thereby increasing one IPI in the stimulus, had a much bigger effect on pitch than was produced by advancing those pulses. The model succeeded

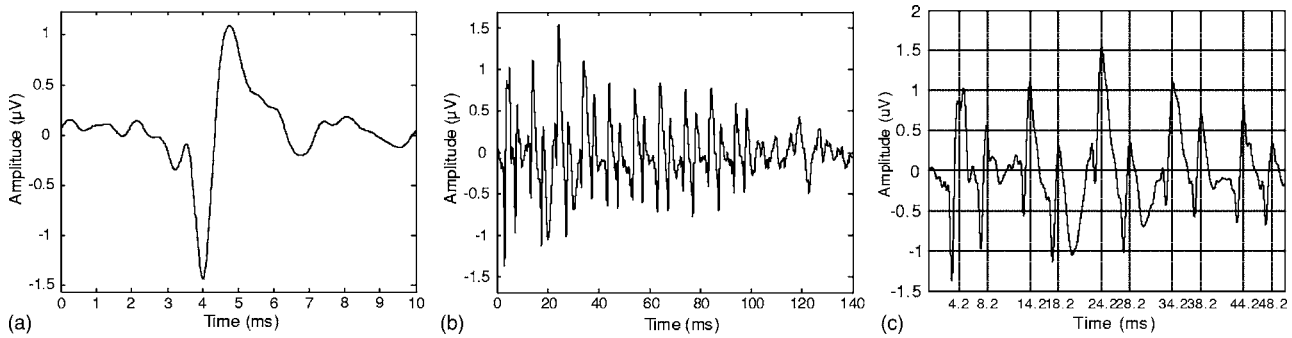


FIG. 7. Part (a) shows the CAP to a stimulus consisting of the first pulse of a 4–6 pulse train, in listener NH2. Part (b) shows the response to part of a 4–6 pulse train in listener NH7. Part (c) shows a zoomed-in portion of part (b). The vertical gridlines are spaced, alternately, by 4 and 6 ms.

because the increased IPI in the "delayed" stimulus received a smaller weight than the shortened IPI in the "advanced" stimulus.

Overall, Carlyon *et al.*'s 2002 model does a good job of predicting the pitch of equal-amplitude pulse trains, such as those used here and in previous experiments (Carlyon, 1997; Plack and White, 2000; Carlyon, 2002). However, as noted in the Introduction, it can account neither for the multiple pitches that can be heard in amplitude-modulated isochronous pulse trains (McKay and Carlyon, 1999), nor for the different effects of attenuating pulses occurring after the longer *vs.* the shorter intervals in alternating-interval pulse trains (van Wieringen *et al.*, 2003). Here we consider whether a simple model based on refractory properties of the auditory nerve can account for such data.

TABLE II. (a) Percentage difference between the amplitudes of CAPs measured after the shorter vs. longer intervals for the unmodulated stimuli of experiment 3b. Data from 3 GPs are shown, and are averaged across conditions where the stimulus started with the shorter of the two possible intervals (e.g., 4-ms in the "4–6" pulse train) and where it started with longer interval (e.g., 6-ms). An exception occurred for GP7 in the 8–12 condition, where, due to an error, only stimuli starting with the 8-ms interval were used. Part (b) is similar to (a) but for the 4–6 stimuli of experiment 3b in which the pulses occurring after the longer or shorter intervals could be attenuated by 2 or 6 dB, thereby producing amplitude modulation. When modulated, only stimuli starting with the shorter of the two intervals were used, and for consistency analysis of the unmodulated stimuli was restricted to those starting with the shorter interval. (This is why the data for the unmodulated stimuli can differ slightly from those shown for the same stimuli in part a). (c) is similar to (b) except for the 8–12 stimuli.

a)	Stimulus			
Animal	3.5–5.5	4–6	4.5–6.5	8–12
GP6	13.0	10.1	9.4	9.5
GP7	7.3	6.1	4.6	14.4
GP8	11.1	10.1	17.0	12.7
Mean	**10.5**	**8.7**	**10.3**	**12.2**

b)	Stimulus				
Animal	4–6L6	4–6L2	4–6	4–6S2	4–6S6
GP6	50.4	−21.3	9.5	57.9	144.2
GP7	−13.4	−8.3	7.4	24.1	35.4
GP8	−5.1	−2.3	7.7	21.7	25.6
Mean	**−23.0**	**−10.6**	**8.2**	**34.6**	**68.4**

c)	Stimulus				
Animal	8–12L6	8–12L2	8–12	8–12S2	8–12S6
GP6	−43.4	−14.9	9.5	44.5	130.8
GP7	9.3	3.3	15.1	26.9	13.8
GP8	4.3	2.6	12.7	23.0	18.2
Mean	**−9.9**	**−3.0**	**12.4**	**31.4**	**54.3**

2. Neural model: General form and level independence

In the Introduction we described a new type of model in which an array of neurons, central to the AN, only respond when the amplitude of the CAP exceeds a certain fixed threshold value. Here we consider the simplest form of this scheme, in which the thresholds are distributed uniformly across the more-central neurons, and where pitch is estimated from an unweighted sum of the first-order intervals in the outputs of these neurons. One implementation of this idea could occur via an array of "synchrony detectors," each responding when a threshold number of input fibers fire in synchrony. Our only assumptions concerning the time window over which synchrony detection occurs is that it includes those smoothing properties—e.g., integration by IHC and AN cell membranes—that are involved in the generation of the CAP, and that it is shorter than the shortest interval between any two successive CAPs described here (e.g., 3.5 ms).

Two aspects of the physiological data obtained in experiment 3 lend general support to the model. First, the CAPs to equal-amplitude, alternating-interval pulse trains are indeed amplitude modulated. Second, the depth of this modulation is largely independent across level. This second finding is important because the model assumes a uniform distribution of thresholds for the "more central" neurons, so that any marked change in modulation depth across level would predict a substantial change in pitch. In fact, as Fig. 6 shows, the modulation in the CAP response differs only from 7–11% over a 50-dB range of input levels. In terms of the model, this would produce matches that ranged only from 5.35 to 5.66 ms. This finding is also of more general theoretical importance as it helps resolve a potential paradox in the literature. A number of studies point to the conclusion that pitch is dominated by the first-order intervals in the

stimulus, and that higher-order intervals have a smaller effect (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001; Carlyon, 2002; Yost *et al.*, 2005). However, it has been argued that models of pitch that rely on first-order intervals is that such representations, when applied to the responses of single AN fibers, are highly level dependent (Cariani and Delgutte, 1996; McKinney and Delgutte, 1999). The CAP measures obtained in experiment 3a support Carlyon *et al.*'s (2002) suggestion that this problem can be overcome if one assumes that the representation of first-order intervals is derived *after* the responses of individual AN fibers are combined. Our results and analysis also suggest that the "first-order interval" approach should be modified such that, when the CAPs to some pulses are larger than those to others, intervals between these larger CAPs may contribute to pitch.

3. Neural model: Effect of inter-pulse interval

Our results also suggest, however, that the refractory properties of the AN, as processed by our simple model, cannot, by themselves, account for all aspects of temporal pitch perception. An important discrepancy can be seen in Fig. 3(b); the downward slope of the line connecting the mean data reflects the fact that the pitch match to a 4.5–6.5 pulse train is closer to the mean IPI (5.5 ms) than is the case for a 3.5–5.5 train (mean IPI=4.5 ms). This would be consistent with the refractoriness model if the slope of the recovery function decreased over this range, so that the amount of modulation in the CAP waveform were smaller for the 4.5–6.5 than for the 3.5–5.5 stimulus. However, the GP recordings from experiment 3a did not reveal such a trend. To quantify this discrepancy we obtained an estimate of the function relating neural response to inter-pulse interval that would be necessary to account for the mean data shown in Figs. 3(a) and 3(b). The procedure adopted was as follows: (i) calculate the *relative* size of the CAPs to pulses occurring after the shorter *vs.* longer intervals that would be needed to account for the pitch data obtained with each of the 3.5–5.5, 4–6, and 4.5–6.5 stimuli. This gives three pairs of values, where the relative CAP amplitude for the two members of each pair are defined relative to each other. It does not constrain the relative amplitude between members of other pairs (e.g., between gaps of 3.5 and 4 ms). (ii) Assume that the CAP amplitude after a 3.5 ms interval has a value of 1, and that the amplitude after gaps of 4 and 4.5 ms is equal to $i1+Add4$ and $1+Add4.5$, respectively, (iii) Assume that the form of the function relating CAP amplitude to inter-pulse interval (Δt) is $y=a.\ln(\Delta t-r)+b$. (iv) Adjust *Add4*, *Add4.5*, *a*, *b*, and *r* to minimize the least-squares error between this function and the data, using the routine "solver.exe" in Microsoft Excel, and with the constraint that $r \geq 0$.[2] The resulting function, $y=0.163\ln(\Delta t-2.5)+1$ is shown by the solid lines connecting squares in Fig. 8, with CAP amplitude plotted relative to that after a gap of 3.5 ms. It is initially steeper than a similar function fit to the *actual* data, obtained from the CAPs to the alternating-interval pulse trains in GPs ($y=0.267\ln(\Delta t)+0.656$; solid lines and triangles; see also Table II), but, unlike the GP function, decreases in slope over the range studied.

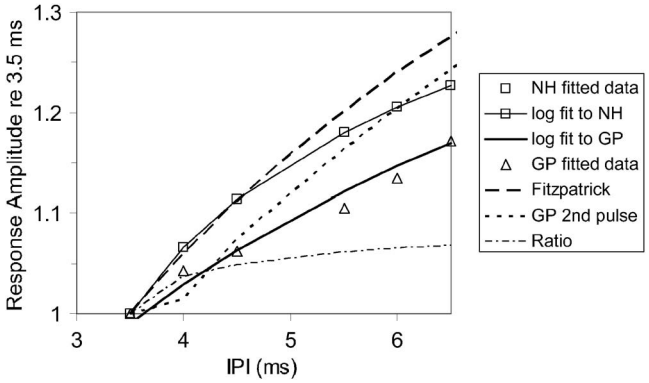


FIG. 8. The squares show the recovery function, expressed as response amplitude *re* that at IPI=3.5 ms, necessary for the neural model to account for the NH pitch data from experiment 1. The solid line passing through these points represents the best fit to these data using a logarithmic function. The triangles show the points derived from the GP data of experiment 3b: the ratio between the amplitudes at 5.5 vs 3.5, 6 vs 4, and 6.5 vs 4.5 ms reflect the depth of AM in the CAP response to the 3.5–5.5, 4–6, and 4.5–6.5 stimuli, respectively. The vertical distances between other delays (e.g., 3.5 vs. 4 ms) were adjusted to provide the best logarithmic fit to the data, shown by the bold solid line. The ratio between these first two curves (faint and bold solid lines) is shown by the dot-dashed line. The bold dashed line shows the two-pulse recovery functions for the cat AN described by Fitzpatrick *et al.* (1999), as fit by Carlyon *et al.* (2002). The dotted line shows the CAP amplitude to the second pulse in each train as a proportion of that to the first, using data obtained from experiment 3b.

One possible reason for the discrepancy stems from species differences. The dashed line in Fig. 8 shows the recovery functions obtained from the cat by Fitzpatrick *et al.* (1999), defined as the probability of a single AN fiber firing in response to the second of two acoustic clicks, as a function of inter-click interval. The probabilities were estimated using the logarithmic fit to Fitzpatrick *et al.*'s data employed by Carlyon *et al.* (2002), and normalized to that at an inter-click interval of 3.5 ms. This function, like that needed to account for the NH data, decreases in slope over the range shown. If one assumes that the CAP modulation in an alternating-interval pulse train is determined by the ratio of the value of Fitzpatrick *et al.*'s recovery function at the two IPIs in that train, then our neural model predicts the pitch matches shown by filled circles without lines in Fig. 3. It can be seen that, although this version of the model corresponds to slightly longer periods than the average obtained in experiment 1, it captures the general trends in the data—most notably the tendency for matches to move towards the mean IPI in the stimulus as the overall IPI is increased (Fig. 3(b)). A *caveat* is that the recovery functions obtained with two-pulse stimuli may not capture the ratio of the CAP amplitudes to pulses occurring after the longer *vs.* shorter inter-pulse intervals in a pulse train. To illustrate this, the dotted line in Fig. 8 shows the amplitude of the GP CAP to the second pulse of a pulse train as a proportion of that to the first, using data obtained in experiment 3b. (These measures should be similar to two-pulse recovery functions provided that the responses to the first two pulses are not strongly influenced by those to the previous pulse trains). Like the function obtained using pulses within a train (solid line), it does not get shallower as IPI is increased from 3.5 to 6.5 ms. It is, however, steeper overall, suggesting that IPI has a

larger effect on CAP amplitude at the start of a pulse train than during it.

An alternative explanation is that pitch is affected by some additional source of refractoriness, over and above that observed in the auditory nerve. This could arise either from the output of the AN passing through a second neural stage prior to the "more-central" neurons, or to refractoriness inherent to those more-central neurons. To quantify the additional refractoriness needed, we divided the function fit to the NH data (squares) from the measured GP function (triangles), and plotted this ratio by the dot-dashed lines. Note that this function describes the "gain" applied to a pulse after a given IPI, *relative to that at an IPI of* 3.5 ms. As refractoriness should reduce the amplitude of neural responses, we would expect the *absolute* value of the gain to be less than 1. The curve flattens off above about 4 ms, and the best-fitting logarithmic fit is $y=0.089\ln(\Delta t)+0.9$

B. More general models of pitch

Our measurements have, for reasons described in the Introduction, been restricted to situations where resolved harmonics are absent. However, the general idea that neural refractoriness can influence "purely temporal" pitch perception could be incorporated into more general models of pitch. As an example, we consider the recent model proposed by Wiegrebe and colleagues (Wiegrebe and Meddis, 2001; Wiegrebe and Winter, 2001). They proposed that pitch may be coded by populations of chop-S neurons in the cochlear nucleus, where each population consists of neurons with a given chopping rate (CR) but a range of characteristic frequencies. They showed, both by computer simulations and reproductions of GP recordings made by others (Wiegrebe and Meddis, 2001; Winter *et al.*, 2001), that when a population is stimulated by a sound whose F0 is equal to CR, the neurons in that population show an enhanced tendency to fire at a rate equal to CR. Importantly, this enhancement also occurs when the F0 is an integer multiple of CR, reflecting the ability of chop-S units to "skip" input spikes. As a result, chop-S units with CRs equal to the period of a harmonic complex will chop at CR=1/F0, even when their CFs are tuned to higher (but still resolved) harmonics of that F0. In the absence of refractory effects, we might expect our (unresolved) 4–6 stimulus to produce enhanced temporal firing in populations with CRs equal to the reciprocals of 4, 6, and 10 ms (the latter representing the ability of chop-S neurons to skip input spikes). If, as we have shown, refractory effects cause the AN input to chop-S neurons to be amplitude modulated, then we might expect this to increase the enhancement in those populations with CRs of 100 Hz (the reciprocal of 10 ms). Pitch might then be judged from a weighted sum of Chop-S populations with CRs that produce the greatest temporal enhancement. We should note that, in fact, Wiegrebe and Meddis propose a refractory period of only 0.75 ms, so it is unlikely that the current implementation of the peripheral stages of their model could capture the CAP modulation observed here. However, this could easily be modified by changing the refractory period of the model.

An additional issue facing the Wiegrebe and Meddis model, shared with our own simple account, is that, when NH listeners are allowed to adjust the period of an isochronous stimulus to match that of a 4–6 pulse train, they never produce matches to a period of 10 ms (Carlyon, 2002). Instead, the resulting distribution of matches is unimodal, suggesting that subjects either do not have conscious access to the separate 4, 6, and 10 ms intervals present in the CAP, or always choose to combine them into a summary measure rather than sometimes matching to one or another of these periods. In contrast, McKay and Carlyon's (1999) MDS study showed that, with *physically* amplitude modulated stimuli (Fig. 1(b)), listeners *were* sensitive to both the modulator and carrier rates. This difference may have been due to differences in the depth of CAP modulation produced by the different stimuli in the two studies, to the fact that in McKay and Carlyon's study the carrier and modulator rates were harmonically related, or to differences in measurement procedure (pitch judgements vs. MDS).

VI. SUMMARY AND CONCLUSIONS

An important topic in the study of pitch perception is the relationship between the representation of the stimulus at the level of the AN and the perceived pitch. The experiments described here compared behavioral measures of "purely temporal" pitch perception with measurement of auditory nerve activity obtained from GPs and humans, using very similar stimuli. The absence of resolved harmonics in our stimuli greatly simplifies this comparison, and allows a quantitative comparison between physiology and behavior. This feature was used to compare pitch judgments to the predictions of a simple model in which pitch is derived from first-order intervals in the combined responses of many AN fibers, as measured by the CAP. According to the model, an array of "more-central" neurons, whose thresholds are uniformly distributed, fire whenever the CAP exceeds threshold. Pitch is then estimated from an unweighted average of the first-order intervals in the outputs of these more-central neurons.

The results show that the CAP to equal-amplitude alternating-interval stimuli is amplitude modulated both in NH humans and GPs, and that this AM is constant over the 50-dB range of levels studied in the GP. The presence of AM can qualitatively account for the finding that the pitch of such stimuli corresponds to a period that is slightly longer than the mean interval present in the stimulus. Importantly, its level independence is consistent with our behavioral finding that the pitch of 4–6 stimuli is similar to that observed in a previous study using a 24 dB lower level. This helps resolve the potentially conflicting findings that temporal pitch is dominated by first-order intervals in the *stimulus* (Kaernbach and Demany, 1998; Plack and White, 2000; Kaernbach and Bering, 2001; Yost *et al.*, 2005), but that codes based on first-order statistics of the responses of individual neurons are strongly level dependent (Cariani and Delgutte, 1996; McKinney and Delgutte, 1999). The resolution occurs because a model based on first-order intervals in the neural response can produce realistic pitch estimates that are level independent, provided that the summary statistic is derived *after* the responses of many neurons have been combined.

There are, however, quantitative discrepancies between the predictions of the model and the variation in pitch between 3.5 and 5.5, 4–6, and 4.5–6.5 stimuli. We discuss this discrepancy in terms of possible species difference and of the effects of refractoriness in neural stages central to the AN.

[1]We waited until the effects of short-term adaptation had started to level off so that, when comparing the response to pulses occurring after 4- and 6-ms intervals, the results would not be overly influenced by the response to the first pulse that we analyzed. To check that we had succeeded in doing so, we compared the ratio of the response amplitude after 6-ms vs. 4-ms intervals for the 4–6 pulse train and for the 6–4, pulse train in the three GPs for whom these data were available. The first pulse analyzed would correspond to a 6-ms interval in the former case and to a 4-ms interval in the latter. If the measured ratio were overly influenced by the first pulse then it would therefore differ between these two stimuli. The two ratios were very similar (1.061 and 1.065, respectively), and did not differ significantly ($t(df=2)$, $p=0.1$).

[2]The choice of a log function was motivated by other data in the literature (e.g., Fitzpatrick *et al.*, 1999), but a fit using a compressive power function yielded similar results.

Cariani, P. A., and Delgutte, B. (**1996**). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol. **76**, 1698–1716.

Carlyon, R. P. (**1997**). "The effects of two temporal cues on pitch judgements," J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

Carlyon, R. P., and Shamma, S. (**2003**). "An account of monaural phase sensitivity," J. Acoust. Soc. Am. **114**, 333–348.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

Eggermont, J. J., and Spoor, A. (**1973**). "Cochlear adaptation in guinea pigs: A quantitative description," Audiology **12**, 193–220.

Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (**1999**). "Responses of neurons to click pairs as simulated echoes: Auditory nerve to auditory cortex," J. Acoust. Soc. Am. **106**, 3460–3472.

Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanisms involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–l047.

Kaernbach, C., and Demany, L. (**1998**). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," J. Acoust. Soc. Am. **104**, 2298–2306.

Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). "Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure," J. Acoust. Soc. Am. **67**, 1704–1721.

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

Loeb, G. E. (**2005**). "Are cochlear implant patients suffering from perceptual dissonance?," Ear Hear. **26**, 435–450.

Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). "Spatial cross-correlation," Biol. Cybern. **47**, 149–163.

Mahendran, S., Bleeck, S., Winter, I. M., Baguley, D. M., Axon, P. R., and Carlyon, R. P. (**2007**). "Human auditory nerve compound action potentials and long latency responses," Acta Oto-Laryngol., **127**, 1273–1282.

McDermott, H. J. (**1997**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.

McKay, C. M., and Carlyon, R. P. (**1999**). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," J. Acoust. Soc. Am. **105**, 347–357.

McKay, C. M., and McDermott, H. J. (**1998**). "Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals," J. Acoust. Soc. Am. **104**, 1061–1074.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). "Pitch matching of amplitude modulated current pulse trains by cochlear implantees: The effect of modulation depth," J. Acoust. Soc. Am. **97**, 1777–1785.

McKinney, M. F., and Delgutte, B. (**1999**). "A possible neurophysiological basis of the octave enlargement effect," J. Acoust. Soc. Am. **106**, 2679–2692.

Moore, B. C. J. (**1989**). *An Introduction to the Psychology of Hearing*, pp. 183–187 (Academic, New York).

Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Springer Handbook of Auditory Research: Pitch Perception*, edited by C. J. Plack and A. J. Oxenham (Springer-Verlag, Berlin), pp. 234–277.

Moore, B. C. J., Glasberg, B. R., and Flanagan, H. J. (**2006**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.

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.

Murnane, O. D., Prieve, B. A., and Relkin, E. M. (**1998**). "Recovery of the human compound action potential following prior stimulation," Hear. Res. **124**, 182–189.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Plack, C. J., and White, L. J. (**2000**). "Pitch matches between unresolved complex tones differing by a single interpulse interval," J. Acoust. Soc. Am. **108**, 696–705.

Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533.

Pressnitzer, D., de Cheveigne, A., and Winter, I. M. (**2001**). "Perceptual pitch shift for sounds with similar waveform autocorrelation," ARLO Online (http://scitation.aip.org/ARLO/top.html) **3**, 1–6. (Last viewed online 7 August 2007).

Pressnitzer, D., de Cheveigne, A., and Winter, I. M. (**2004**). "Physiological correlates of the perceptual pitch shift for sounds with similar waveform autocorrelation," ARLO Online (http://scitation.aip.org/ARLO/top.html) **5**, 1–6. (Last viewed online 7 August 2007).

Shamma, S. (**1985**). "Speech processing in the auditory system: II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.

van Wieringen, A., Carlyon, R. P., Long, C. J., and Wouters, J. (**2003**). "Pitch of amplitude-modulated irregular-rate stimuli in electric and acoustic hearing," J. Acoust. Soc. Am. **114**, 1516–1528.

Wiegrebe, L., and Meddis, R. (**2001**). "The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus," J. Acoust. Soc. Am. **115**, 1207–1218.

Wiegrebe, L., and Winter, I. M. (**2001**). "Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus," J. Neurophysiol. **85**, 1206–1219.

Winter, I., Wiegrebe, L., and Patterson, R. D. (**2001**). "The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea pig," J. Physiol. (London) **537**, 553–566.

Yost, W. A., Mapes-Riordan, D., Shofner, W., Dye, R., and Sheft, S. (**2005**). "Pitch strength of regular-interval click trains with different length "runs" of regular intervals," J. Acoust. Soc. Am. **117**, 3054–3068.

On the influence of interaural differences on temporal perception of noise bursts of different durations

Othmar Schimmel
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

Armin Kohlrausch[a)]
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands and Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

(Received 8 December 2006; accepted 14 November 2007)

The perception of a composite sound's temporal cues, like synchronous onsets, is considered essential to correct perceptual grouping of its constituent components. The processing of a single sound's spatial cues, already present at its onset, may interact with temporal perception of the onset. The current study investigated the influence of interaural differences on temporal perception of a sound's onset. As a measure of temporal perception, the ability to position the onset of a temporally displaced target sound to the regular meter of diotic reference marker sound onsets was measured for various target sound lateralizations, sensation levels, and target and marker sound durations. For target sounds presented in quiet, no influence of interaural differences on temporal positioning of the onset was found. However, increasing a sound's duration systematically shifted the perceived onset position into its "interior." For target sounds presented at low sensation levels in a noise masker, the precision of temporally positioning the onset generally degraded, though faster for dichotic conditions and for longer durations. The level below which temporal perception precision starts to degrade was found to depend on signal-to-noise ratio rather than on sensation level or duration, and is influenced by the presence of interaural differences. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2821979]

PACS number(s): 43.66.Lj, 43.66.Pn, 43.66.Mk [RLF] Pages: 986–997

I. INTRODUCTION

In everyday listening, the correct grouping of signal components from concurrent sounds within an acoustic scene is essential for the perceptual organization of their sources. This grouping of signal components into so-called auditory objects is based on combining signal components that are likely to come from the same source, based on common spectral, temporal, and/or spatial cues (Bregman, 1990).

A strong cue for *initiating* the grouping of signal components into a new auditory object, and implying its segregation from other auditory objects, is the detection of synchronous onsets, i.e., when the onsets fall within some temporal window such that they are perceived as being synchronous (cf. Bregman and Pinker, 1978; Dannenbring and Bregman, 1978; Darwin and Ciocca, 1992; Hukin and Darwin, 1995). Once grouping of signal components is initiated through establishing their common onset, the new auditory object can be monitored across time by exploiting cues that support this grouping and the segregation of this object from concurrent objects. Onset asynchronies as small as about 30 ms were found sufficient to promote alternative grouping of signal components into a different auditory object (cf. Darwin, 1984; Roberts and Moore, 1991). Therefore, the accurate temporal perception of onset synchrony is deemed important for the correct grouping of signal components of concurrent sounds and identification of their sources (Hill, 1994), as for instance required for speech intelligibility in energetic masking conditions. A better understanding of temporal perception in multisource conditions may prove useful to, for example, the perception-based modeling of speech recognition and auditory scene analysis, or efficient audio coding for communication technology.

In a previous study (Schimmel and Kohlrausch, 2006), it was investigated to what extent the temporal perception of a broadband noise's onset was influenced by applying interaural time or level differences to achieve lateralization to the right compared to a condition without interaural differences. For signals presented in quiet and at a high sensation level, no influence of interaural differences on the temporal accuracy or precision for perceiving their onset was found. However, for signals presented at low sensation levels in a diotic noise masker, interaural differences did influence the temporal perception precision. For diotic target conditions, temporal perception precision of the onset was similar at all sensation levels, with a standard deviation between about 10 and 25 ms, down to a level as low as 2 dB above detection threshold. Below this level, the temporal perception precision dropped toward chance performance. This finding indicated precise temporal perception down to low suprathreshold levels when detection of the onset was based on only monaural cues. For corresponding dichotic target conditions, temporal perception precision of the onset already degraded at a level of about 5 dB above threshold, also reaching chance performance at threshold. This finding indicated less precise temporal perception close to detection threshold

[a)]Electronic mail: armin.kohlrausch@philips.com.

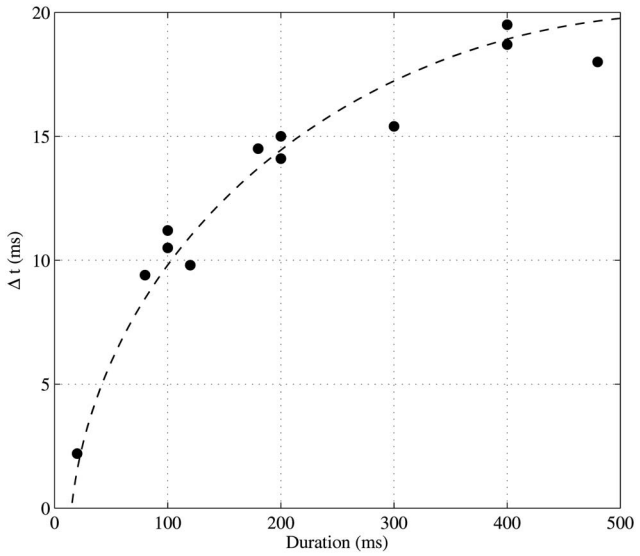


FIG. 1. Temporal shift Δt of the perceived onset for various durations of a signal, using a constant 25-ms reference signal. Each point represents the average of six subjects with six measurements each, and the curve represents the fitted model. Replotted from Schütte (1978b), Fig. 2.

when using binaural cues for onset detection. The observed significant difference in temporal positioning precision between diotic and dichotic conditions at equal low sensation levels is consistent with other studies showing that the binaural hearing system may provide less advantage for suprathreshold perceptual tasks other than detection (cf. Cohen, 1981; van de Par *et al.*, 2005).

In the study of Schimmel and Kohlrausch (2006), the sounds had a fixed duration of 50 ms, and it was implicitly assumed that the *perceived* onset was closely associated, or even coincided, with the *physical* onset. That such an association is not true for all types of signals has been shown by Schütte (1976), Terhardt and Schütte (1976), and Schütte (1978a). Their research on the association of a sound's perceived onset with its physical onset showed that the two can differ by up to 20 ms for sounds with rectangular envelopes. The perceived onset is regarded as a function of a sound's duration and its temporal envelope, following an asymptotic buildup from the physical onset to the maximum of the signal's transient, and can be predicted as the moment at which a certain percentage of the maximum of the transient is reached (Schütte, 1978b; Zwicker and Fastl, 1999, pp. 271–275). Figure 1 shows experimental data and a model fit for the temporal shift Δt of the perceived onset for various durations of a signal, using a constant 25-ms reference signal (replotted from Schütte, 1978b, Fig. 2).

In this interpretation of the transient behavior of the human hearing system, a longer sound with rectangular envelope has a shallower perceived transient than an identical but shorter sound, shifting the moment when the required percentage of the maximum of the transient is reached. With increasing duration, the perceived onset becomes increasingly dissociated from the physical onset and, due to a shallower slope of the perceived transient, is temporally less defined. This transient behavior may also be responsible for the degradation in temporal perception precision when the level of a sound with a rectangular envelope is close to threshold (as observed in Schimmel and Kohlrausch, 2006), due to the relative decrease of the transient's maximum, which results in a shallower slope of the perceived transient.

From previous research, three signal parameters can thus be identified that influence the temporal perception of a sound's onset: its duration (Schütte, 1978b), its spatial configuration, and its level (Schimmel and Kohlrausch, 2006). However, knowledge of possible interactions between these three parameters is currently lacking, because they were investigated in different studies.

The first goal of the current study was to investigate whether the previous conclusions about the influence of interaural differences on temporal perception accuracy and precision from Schimmel and Kohlrausch (2006) can be extended to a wider range of signal durations. Using the same experimental paradigm and setup, experiments 1 and 2 measured the temporal positioning for both diotic and dichotic signal onsets in quiet and at equal low sensation levels in a noise masker for a shorter (5 ms) and a longer signal duration (350 ms) than the previously used 50-ms duration. Based on the transient behavior model of Schütte (1978b), temporal perception precision of the signals' onset is expected to decrease with an increase in duration, independent of its spatial configuration and its sensation level.

The second goal of the current study was to investigate whether the dissociation of the perceived onset from the physical onset, as observed by Schütte *et al.* for diotic signals, is influenced by the lateralization of a dichotic signal. Using the same experimental setup as in experiments 1 and 2, experiments 3 through 5 measured a possible shift of the perceived onset of diotic and dichotic signals for the various durations and sensation levels. Based on the findings of Schimmel and Kohlrausch (2006), no influence of interaural differences on temporal perception and thus on the dissociation of the perceived from the physical onset is expected for sounds presented in quiet. For sounds presented at low sensation levels, no detailed prediction for temporal perception can be derived from the previous research.

II. GENERAL METHOD

The same perceptual center procedure as used in Schimmel and Kohlrausch (2006) was applied for determining the ability to perceive the onset of a sound. The procedure is based on identifying the perceptual center location or perceived moment of occurrence (Marcus, 1976; Pompino-Marschall, 1989). It is comprised of judging the temporal position and adjusting a target sound along the time axis of the stimulus, relative to isochronous reference marker sounds in the stimulus, in order to find the desired temporal alignment. In general, for perceptually identical or very similar target and marker sounds, perceptual isochrony can be accomplished with physical isochrony of target and marker onsets, because their perceptual centers relate to their physical onsets in the same way (or even may coincide with the physical onsets, as assumed in Schimmel and Kohlrausch, 2006). For perceptually different sounds, perceptual isochrony may be accomplished with physical anisochrony of target and marker onsets, where the measured physical aniso-

chrony directly indicates an influence of the acoustical parameters of the target and marker sounds on their respective perceptual centers.

The stimulus structure consisted of a regular series of five clearly audible marker pulses, of which the third marker pulse was omitted. The subjects' task was to align a target's onset, initially positioned randomly around the temporal onset position of the missing third marker pulse, to the meter of the four present marker onsets (see Fig. 2 for a graphical representation of a stimulus). The time difference Δt between the target's adjusted onset position from the physical onset isochrony, induced by the 500-ms period of the marker onsets, was recorded and served as a quantitative measure for temporal positioning performance. From these measures, temporal positioning accuracy and precision of differently processed targets and markers could be compared.

The step size for timing adjustments was set to 5 ms, which roughly matches the just-noticeable difference for changes in the perceptual center (Marcus, 1976; Schütte, 1976), and is about half the 10.7-ms just-noticeable difference for displacement of the fourth tone in an isochronous sequence of six for an interonset interval of 500 ms (Friberg and Sundberg, 1995). The initial random temporal displacements were within 25 steps from physical onset isochrony in both directions, which corresponds to an equal-probability distribution between −125 and 125 ms in steps of 5 ms. The temporal positions for adjustment of the target's onset were limited to the range between −450 and 450 ms.

All sounds were broadband noises with independent Gaussian probability distributions, covering the frequency range up to 22.05 kHz, only limited by the frequency response of the reproduction system. No ramps were applied to the sounds, so they had abrupt onsets and offsets.

The experiment was controlled by the subject via a computer keyboard. In each condition, after initial playback of the stimulus with a randomly positioned target, the target position could be shifted by pressing f (for "forward") or b (for "backward") the number of times the target was to be moved by the step size. By pressing Enter, playback of the stimulus with the target's new temporal position was started, and the procedure could be repeated until the fit of the target onset to the meter of the marker onsets was judged to be as accurate as possible. The final time instant of the target's onset on the stimulus' time line was recorded as its perceived moment of occurrence. Feedback was provided after each condition in the form of a graph, representing the followed track from the initial onset position to the final adjustment.

A within-subject experimental design with counterbalanced complete randomization of all conditions within each experiment was applied, and all conditions were presented twice to each subject. All measured temporal onset positions for each condition were pooled, because the between-subject variability was smaller than the within-subject variability. The pooled data were scanned for severe outliers (onset positions at more than three times the interquartile range of the pooled data), and incidental severe outliers were removed from the data set. Nine male subjects participated voluntarily in the experiment. One subject, however, did not complete experiment 4 and some conditions in experiments 2 and 5.

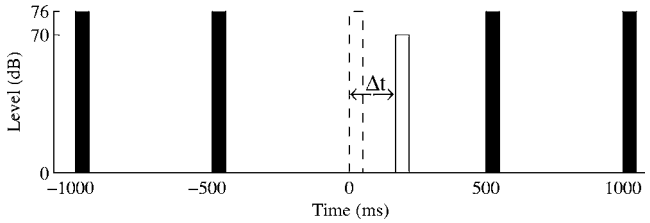


FIG. 2. Example of an acoustic stimulus for the temporal positioning experiment. The onset of a randomly positioned target (in white) was to be temporally aligned to the meter of a regular series of marker onsets (in black). Measured was the dependent variable onset position, quantified as the time difference (Δt) from physical onset isochrony (dashed line at 0 ms). The target was either diotic, i.e., identical at both ears ([C] conditions), or dichotic ([ITD] or [ILD] conditions). Different durations were used for target and markers. Shown is a 50-ms target with 50-ms markers.

Because the averaged results for nine subjects did not change significantly when excluding this subject, the data were included where available and otherwise limited to eight subjects. All subjects had previous experience in listening tests and reported normal hearing. Training was thought not to be required, because the procedure allowed the subjects to adjust the stimuli for as long as necessary to get the preferred temporal alignment of the target to the regular marker pattern before continuing with the next condition.

The measurements were done in an acoustically isolated listening room at the Philips Research Laboratories in Eindhoven. To digitally generate the stimuli for each presentation in real time and automate the experiment and data collection, a computer running MATLAB software was used. Digital stimuli were converted to analog signals by a Marantz CDA-94 two-channel 16-bit digital-to-analog converter at a sampling rate of 44.1 kHz, and were presented over Beyer Dynamic DT990 Pro headphones.

III. EXPERIMENT 1

This experiment investigated the influence of a sound's interaural differences on the ability to temporally position its onset to the meter of a regular pattern of marker onsets in quiet, for equal durations of target and marker sounds.

A. Stimuli

Two kinds of sounds were used: four markers with a level of 76 dB SPL, and a target with a level of 70 dB SPL. The markers were diotic to accomplish lateralization in the center. The target was presented with different interaural parameters to yield lateralization at two different positions, in the center (diotic conditions) and to the far right (dichotic conditions). The center condition ([C]) comprised an unprocessed target, identical at both ears. Lateralization to the far right was done by applying interaural time differences ([ITD]) or interaural level differences ([ILD]). The time delay in the [ITD] condition was set at 29 samples at 44.1 kHz in the left channel, approximating a 660-μs lag at the left ear. The interaural level difference in the [ILD] condition was set at 18 dB: target level +9 dB at the right ear, and target level −9 dB at the left ear.

Target and marker sounds were of two durations, 5 and 350 ms, and were presented in the combinations with equal

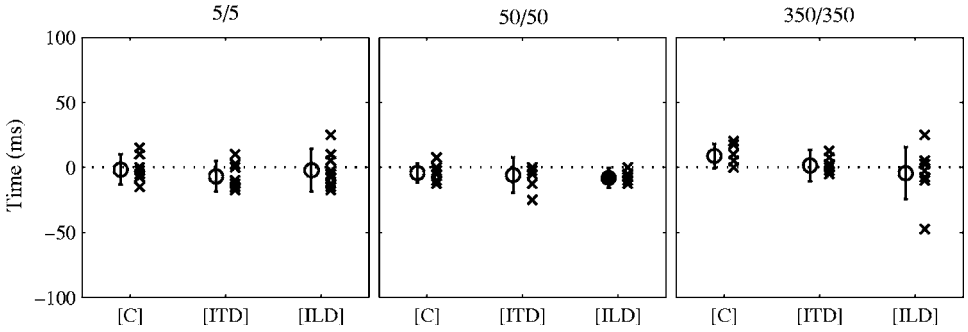


FIG. 3. Means and standard deviations (circles and error bars) and individual subject means (crosses) of the adjusted temporal onset positions in quiet for each lateralization condition and target/marker duration. Closed circles represent conditions with a perceptually relevant bias from physical isochrony.

durations, i.e., a 5-ms target with 5-ms markers (5/5) and a 350-ms target with 350-ms markers (350/350). The combination of a 50-ms target with 50-ms markers (50/50) had been measured earlier under identical experimental conditions (see Schimmel and Kohlrausch, 2006), and results for that condition reported here for the nine subjects participating in both the previous and the current experiments were taken from that data set. For efficiency of reading, duration combinations are hereafter abbreviated in the format of *target/marker* duration, as introduced earlier.

B. Results

Figure 3 shows the means and standard deviations (circles and error bars) and individual subject means (crosses) of the adjusted temporal onset positions (ordinate) for each lateralization condition (abscissa) and for each target/marker duration (panels). As can be seen, mean onset positions for each target/marker duration were similar for the three lateralization conditions, and within a range of two adjustment steps (10 ms) around physical isochrony. An analysis of variance on the temporal onset positions revealed a significant effect at the used 5% level of significance for target/marker duration ($F_{(2,152)}=5.08$, $p=0.007$). No significant interaction between lateralization condition and target/marker duration was found. Tukey HSD post-hoc comparisons showed that the significant differences in mean temporal onset positioning were between the 50- and 350-ms target/marker durations.

Following Schimmel and Kohlrausch (2006), the results of each lateralization condition and target/marker duration were analyzed on the size of relevant biases from physical isochrony in perceptual units (in terms of *d*'), by comparing mean differences with the standard deviations of their underlying distributions. From this analysis, one of the nine mean onset positions was considered to have a perceptually relevant bias different from physical isochrony, i.e., a bias value at a *d*' value of one or higher, because its mean difference was larger than its standard deviation (see Fig. 3).

As a measure of temporal positioning *precision*, the standard deviations of the pooled data of all measurements for each condition were analyzed further. For the three target/marker durations, all standard deviations were between 7 and 20 ms. The temporal positioning was, on average, most precise in [C] conditions and least precise in [ILD] conditions.

To assess temporal positioning precision between conditions, i.e., to quantitatively compare the distributions of temporal onset positions, a statistical analysis of the difference between variances was performed using paired comparisons of *F* ratios, using a 5% level of significance and Bonferroni correction for multiple comparisons. For a fixed target/marker duration, none of the paired comparisons revealed a systematic influence of lateralization condition on temporal positioning precision. Of all nine paired comparisons (three pairs for each of the three durations) only two showed a statistically significant difference: [C] and [ITD] in the 50/50 condition ($F_{(16,17)}=3.44$, $p=0.016$), and [C] and [ILD] in the 350/350 condition ($F_{(17,17)}=4.63$, $p=0.003$). For a fixed lateralization condition, again none of the paired comparisons revealed a systematic influence of duration on temporal positioning precision. The only significant differences resulting from the paired comparisons were observed for the [ILD] condition, where the 50/50 condition had a statistically significant different distribution from the 5/5 and 350/350 conditions ($F_{(16,17)}\geqslant 4.44$, $p\leqslant 0.004$).

C. Discussion

No systematic influence of lateralization condition on mean temporal positioning or temporal positioning precision was found for any of the target/marker durations in quiet. All conditions resulted in mean temporal positioning within two step sizes of physical isochrony (10 ms), just around the just-noticeable difference for changes in the perceptual center. Because very similar sounds are likely to have very similar perceptual centers, finding mean onset positions close to physical isochrony for similar target and marker sounds may not be surprising.

The standard deviations for the 5/5 and 350/350 target/marker duration conditions were similar to those for the previously established 50/50 condition (cf. Schimmel and Kohlrausch, 2006). This means that there was no difference in temporal positioning precision between the shorter and the longer durations. This finding does not support the expectation of a decreasing precision with increasing duration, which was formulated based on the model by Schütte (1978b). Obviously, the steepness of the perceived transients and the dissociation of the perceived onset from the physical onset, which were equal for both target and marker sounds but different for each of the three durations, were not different enough to influence the precision of temporal perception. In conclusion, the similarity in mean temporal positionings close to physical isochrony and in the corresponding standard deviations indicated similar temporal perception of the target, regardless of interaural differences and duration.

TABLE I. Means and standard deviations of the masked threshold levels (in dB, as spectral power ratio between the target and masker) for each lateralization condition. The lower threshold levels of the dichotic conditions ([ITD] and [ILD]) compared to their reference diotic condition ([C]) reflect the binaural masking level difference resulting from the application of interaural differences on the target. Please note that these were different for the three target durations.

	[C]	[ITD]	[ILD]
5 ms	0.6	−3.9	−10.3
	(1.0)	(1.7)	(1.5)
50 ms	−3.9	−13.5	−17.7
	(1.2)	(1.8)	(1.5)
350 ms	−6.3	−17.7	−22.7
	(1.8)	(1.7)	(1.6)

IV. EXPERIMENT 2

This experiment investigated the influence of a sound's interaural differences on the ability to temporally position its onset to the meter of a regular pattern of marker onsets for the durations of target and marker sounds as before, though now at equal low target sensation levels relative to the individual detection thresholds for each condition.

A. Stimuli

Three kinds of sounds were used: a continuous masker with a level of 70 dB SPL, four markers with a level of 76 dB SPL, and a target at two different sensation levels, at 6 dB and at 3 dB above the detection threshold in the masker. The masker and the markers were identical at both ears, to accomplish lateralization in the center. The target had the three lateralizations, referred to as [C], [ITD], and [ILD], as in the previous experiment.

The same two target/marker durations of 5 and 350 ms were used. As before, the 50/50 condition was measured earlier (see Schimmel and Kohlrausch, 2006). The noise masker had a duration of 2500 ms, and started simultaneously with the onset of the first marker, and ended 500 ms after the onset of the last marker.

To enable the temporal positioning measurements at equal low sensation levels in the masker, a three-alternative forced-choice procedure with adaptive level adjustment was applied to estimate the individual detection thresholds of the targets for the various lateralizations and durations (for details, see Schimmel and Kohlrausch, 2006). Table Ishows the mean and standard deviations of the masked threshold levels for each lateralization condition and target duration, based on the pooled data of nine subjects with four measurements each. Again, the results of the conditions with 50-ms targets come from Schimmel and Kohlrausch (2006) for the nine subjects involved. Threshold levels are expressed in decibels as signal-to-overall-noise spectral power ratio between target and masker.

B. Results

Figure 4 shows the means and standard deviations of the adjusted temporal onset positions (ordinate), for each lateralization condition (abscissa) and for each target/marker duration (panels), in the same way as in Fig. 3. The different symbols represent the data measured in quiet (from experiment 1, included for reference), at 6 dB, and at 3 dB target sensation level. For each target/marker duration, mean onset positions at low sensation levels were generally close to physical isochrony. Like the mean onset positions in quiet, mean onset positions at low sensation levels were close to physical isochrony, with the exception of [ITD] at 6 dB SL for the 350/350 condition, which had a considerable mean deviation from physical isochrony. The variability in onset positions was generally larger at equal low sensation levels than in quiet, especially in the condition with the 350-ms sounds, which had much larger standard deviations than conditions with 5- or 50-ms sounds.

An analysis of variance on the means of the adjusted temporal onset positions for all three target sensation levels did not reveal a significant effect for any of the lateralization conditions, target sensation levels, or target/marker durations. As in experiment 1, the results for the two low target sensation levels were analyzed in terms of the size of the perceptual bias different from physical isochrony, by comparing the mean differences with the standard deviations of their underlying distributions. With d' values between 0.0 and 0.6, none of the 18 mean onset positions at the 6 and 3 dB target sensation levels were considered to have perceptually relevant biases from physical isochrony.

Again, standard deviations were taken as the measure for temporal positioning *precision*. Figure 5 shows the standard deviation of the adjusted temporal onset positions (ordinate) at each target sensation level (abscissa), for each lateralization condition (symbols) and each target/marker duration (panels).

As can be seen in Fig. 5, standard deviations increased with a decrease in signal-to-noise ratio for all lateralization conditions and durations. The most precise temporal positioning at low sensation levels was found for 50-ms sounds

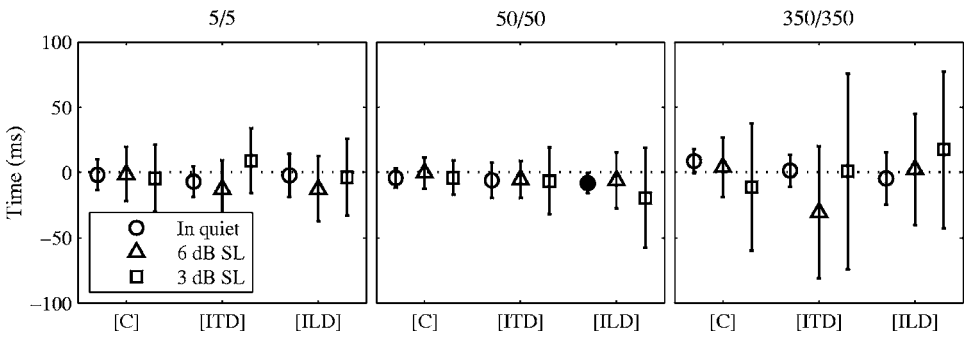


FIG. 4. Means and standard deviations of the adjusted temporal onset positions in quiet (from experiment 1, circles), at 6 dB (triangles), and 3 dB (squares) target sensation level, for each lateralization condition and target/marker duration. Closed symbols represent conditions with a perceptually relevant bias from physical isochrony.

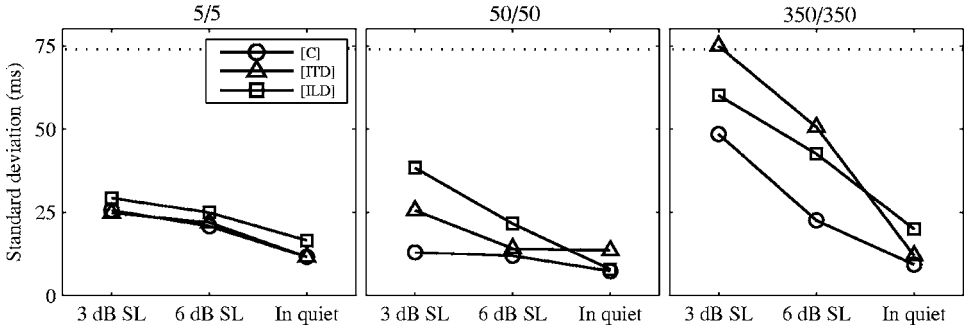


FIG. 5. Standard deviations of the adjusted temporal onset positions vs target sensation levels, for each target/marker duration and lateralization condition. Target sensation levels are relative to the detection threshold in the 70-dB noise masker. The dotted line at 74 ms represents the calculated standard deviation of the random distribution of initial onset positions.

in the [C] condition, which was considerably lower than for 5- and 350-ms sounds in the other two [C] conditions. The least precise performance was found for 350-ms sounds in the dichotic conditions at 3 dB sensation level, which was at or close to the theoretical chance performance of the random distribution of initial onset positions, suggesting that sensitivity to temporal positioning is poor.

The standard deviations for 5- and 50-ms sounds were similar, though with larger differences between lateralization conditions for the 50-ms sounds at the lowest target sensation level. For the 350-ms sounds, standard deviations were considerably larger than for the 5- and 50-ms sounds at the two lower sensation levels. The larger standard deviations for 5- and 350-ms sounds at 6 and 3 dB sensation level for all [ITD] and [ILD] conditions, compared to the corresponding sounds for the [C] condition, suggest that temporal positioning precision decreased more for dichotic conditions than for the diotic condition, and more for longer durations than for shorter durations.

A statistical analysis was conducted by paired comparisons between lateralization conditions for each target/marker duration, using the F ratio of the variances of temporal onset positions. For the 5-ms sounds, no significant influence of lateralization condition on temporal positioning precision was found. For the 50-ms sounds, the paired comparisons found significant differences in temporal positioning precision between the [C] and [ITD] conditions in quiet (cf. experiment 1) and between the diotic and both dichotic conditions at 3 dB sensation level ($F_{(17,17)} \geqslant 3.89$, $p \leqslant 0.008$). For the 350-ms sounds, significant differences were found between the [C] and [ILD] conditions in quiet ($F_{(17,17)}=4.63$, $p=0.003$) and between the diotic and both dichotic conditions at 6 dB sensation level ($F_{(15,15)} \geqslant 3.57$, $p \leqslant 0.019$).

Further analysis was done to investigate whether these observed differences in temporal positioning performance between the three target duration conditions could result from a difference in detectability of the targets for a given sensation level. Using the same procedure as described in Schimmel and Kohlrausch (2006) for the 50-ms targets, psychometric functions were constructed from the detection thresholds measurements for the 5- and 350-ms targets. Similar to detection of the 50-ms targets, the detectability of the 5- and 350-ms targets at 3 dB sensation level was well above 90% correct on the psychometric curves for all lateralization conditions. This analysis indicates that the differences in temporal positioning at sensation levels of 3 and 6 dB cannot be attributed to differences in detectability.

C. Discussion

No systematic deviations of mean onset positions from physical onset isochrony were found for any of the three lateralization conditions, indicating that applying interaural differences to a broadband noise target did not influence its mean temporal position at low sensation levels. The absence of perceptually relevant biases was caused by two facts. First, the mean onset positions remained as close to physical isochrony as in quiet, and second, most standard deviations increased with a decrease in the target's sensation level. These increases in standard deviation showed a reduced temporal positioning precision at the two lower target sensation levels compared to the precision in quiet. The increases were, however, different for the three lateralization conditions and the three target/marker durations.

For the 5-ms sounds, no systematic influence of lateralization condition on temporal positioning precision was found. Standard deviations were similar for all three lateralization conditions at all three target sensation levels, covering a range from 12 to 29 ms. This finding indicates equivalent precision of temporal perception, regardless of the spatial configuration for these short sounds.

For the earlier obtained data for the 50-ms sounds, a statistically significant difference in temporal positioning precision between diotic and dichotic conditions was found at 3 dB target sensation level. When analyzed relative to the detection threshold, the standard deviations for the dichotic conditions increased faster than those for the diotic condition, which meant reduced precision of temporal perception in binaural unmasking conditions as a consequence of the binaural processing required for the unmasking.

For the 350-ms sounds, a statistically significant difference in temporal positioning precision between diotic and dichotic conditions was shown already at 6 dB target sensation level. Standard deviations for the dichotic conditions increased faster than those for the diotic condition, which also meant reduced precision of temporal perception for the dichotic conditions due to binaural processing, though at a higher target sensation level. Compared to the results obtained in Schimmel and Kohlrausch (2006), the temporal positioning precision of 350-ms sounds at 6 dB SL matches the precision of 50-ms sounds at about 2 dB SL.

The temporal positioning accuracy of the three lateralization conditions for each of the three durations indicated systematic differences in temporal perception at equal low sensation levels. In the following, the results for different

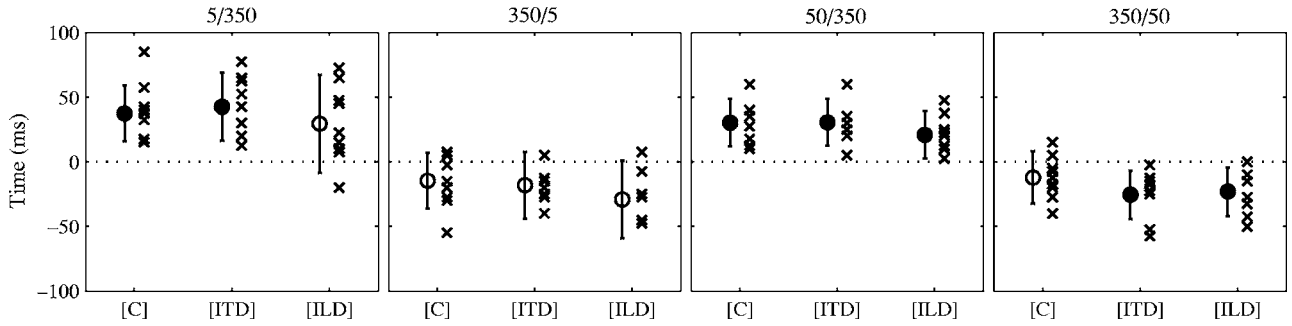


FIG. 6. Means and standard deviations (circles and error bars) and individual subject means (crosses) of the adjusted temporal onset positions in quiet for each lateralization condition and unequal-duration target/marker combination. Closed circles represent perceptually relevant biases from physical isochrony.

durations will be compared in terms of absolute signal level, rather than relative to the duration-dependent masked threshold. Such an analysis emphasizes the intensity contrast at the signal's onset.

For the 50-ms targets, temporal perception precision degraded below a level of about 68 dB SPL (2 dB SL) for the [C] condition and below 63 and 58 dB SPL (6 dB SL) for the [ITD] and [ILD] conditions, respectively. For the 5-ms targets, these absolute signal levels are 2–4 dB *below detection threshold* for all lateralization conditions. In the measurements at 6 and 3 dB SL, the signal onset introduced a level change compared to the masker that was much higher than the critical level derived from the 50-ms targets. In line with the idea that the level change at the onset determines the degradation of temporal precision, no significant change in temporal positioning precision was observed for the 5-ms targets at these sensation levels. For the 350-ms targets, these critical signal levels correspond to sensation levels of about 4 dB for the diotic condition and of about 10-11 dB for the dichotic conditions. Using the same logic, temporal positioning precision is expected to be degraded only at 3 dB SL for the [C] condition, and at both 6 and 3 dB SL for the [ITD] and [ILD] conditions. As shown in Fig. 5, the standard deviations of the 350-ms sounds at 6 and 3 dB SL for the [ITD] and [ILD] conditions are large compared to the small standard deviations of the 5-ms sounds at the same sensation levels. These differences in standard deviations support the hypothesis that the signal-to-noise ratio is a better predictor for temporal perception precision than the sensation level. This finding is in close agreement with the observation from Schütte (1977) that temporal perception precision of a 3-dB level increase of a diotic sinusoidal stimulus is still similar to precision in quiet, while precision degraded for a 2-dB level increase. In the diotic conditions of the current experiment, temporal perception precision remained approximately constant for level increases down to 2.1 dB (68 dB SPL, corresponding to −2 dB signal-to-noise ratio) and decreased strongly for level increases of 1.8 dB and less.

Given the equal durations of the target and the marker sounds, the current results did not provide any information about the actual *size* of a temporal shift from the physical onset to the perceived onset. Using a comparable perceptual center procedure in quiet, Schütte (1976), Terhardt and Schütte (1976), and Schütte (1978a) observed systematic deviations from the objective equidistance of pulse onsets of up to 20 ms for targets and references with rectangular envelopes that had different durations. The experiments of Schütte were restricted to diotic conditions, and his results do not provide any information about temporal perception in dichotic conditions. The following experiments were designed to investigate whether the established dissociation of the perceived onset from the physical onset was influenced by interaural differences. The temporal positioning experiments 1 and 2 were repeated for various combinations of unequal durations of target and marker sounds.

V. EXPERIMENT 3

This experiment investigated the influence of a sound's interaural differences on the ability to temporally position its onset to the meter of a regular pattern of marker onsets in quiet, for unequal durations of target and marker sounds.

A. Stimuli

The same two kinds of sounds as in experiment 1 were used, with the same target lateralization conditions. In contrast to experiment 1, the durations of target and marker were different. The target and markers were presented in two pairs of different unequal-duration combinations, all involving 350-ms sounds. The first pair had the largest ratio in durations, i.e., a combination of a 5-ms target with 350-ms markers (5/350) and the "opposite" combination (350/5). The second pair had the smallest ratio in duration, i.e., a combination of a 50-ms target with 350-ms markers (50/350) and also the "opposite" combination (350/50).

B. Results

Figure 6 shows the means and standard deviations (circles and error bars) and individual subject means (crosses) of the adjusted temporal onset positions (ordinate) for each lateralization condition (abscissa) and for each unequal-duration target/marker combination (panels). As can be seen, mean onset positions for each unequal-duration target/marker combination were similar for the three lateralization conditions, but biased from physical isochrony, with the direction of the bias dependent on the ratio of the target-to-marker duration. All of the 5/350 and 50/350 conditions produced positive values, i.e., the target onset was positioned *later* than physical isochrony as defined by the marker onsets, while the 350/5 or 350/50 conditions all produced

negative values, i.e., the target onset was positioned *earlier* than physical isochrony.

An analysis of variance on the temporal onset positions revealed significant effects of lateralization condition ($F_{(2,203)}=3.67$, $p=0.027$) and target/marker duration combination ($F_{(3,203)}=83.30$, $p<0.001$). No significant interaction between lateralization condition and target/marker duration combination was found. Tukey HSD post-hoc comparisons showed that the significant differences in temporal onset positioning for the lateralization conditions were between the [C] and the [ILD] condition. The post-hoc comparisons also showed that all target/marker duration combinations in which the target was shorter than the marker (5/350 and 50/350) produced mean onset positions that were significantly different from all combinations in which the target was longer than the marker (350/5 and 350/50).

Analysis of the results of each lateralization condition and unequal-duration target/marker combination found that 7 of the 12 mean onset positions yielded perceptually relevant biases from physical isochrony (i.e., bias values at a d' value of one or higher; see the closed symbols in Fig. 6). With five d' values higher than one (on average $d'=1.6$) for the 5/350 and 50/350 conditions, indicating average positioning later than physical isochrony, and only two d' values higher than one (on average $d'=1.3$) for the 350/5 and 350/50 conditions, indicating average positioning earlier than physical isochrony, an asymmetry in temporal positioning for the opposite target/marker duration combinations was observed.

Again the standard deviations of the pooled data, as a measure for temporal positioning *precision*, were taken for further analysis. Temporal positioning was most precise in the conditions with the smallest difference in duration between target and markers (standard deviations 18 ms and 22-38 ms), and least precise in the conditions with the largest difference in duration (standard deviations between 18 and 20 ms).

A statistical analysis of paired comparisons between lateralization conditions for each unequal-duration target/marker combination by the F ratio of the variances of temporal onset positions was done. For a fixed target/marker combination, none of the paired comparisons revealed a significant difference between distributions of temporal positions, indicating that there was no influence of lateralization condition on temporal positioning precision. For a fixed lateralization condition, significant differences between the distributions of temporal positions for the 5/350 condition and those of the 50/350 and 350/50 conditions were found in the [ILD] condition ($F_{(17,17)}\geqslant 3.94$, $p\leqslant 0.007$). All other distributions were statistically identical, suggesting that there was no systematic influence of target/marker combination on temporal positioning precision either.

C. Discussion

For the unequal-duration combinations presented in quiet, no systematic influence of interaural differences on mean temporal positioning or temporal positioning precision could be established. A systematic shift in the perceptual center of the 350-ms sounds was observed. In contrast to the equal-duration combinations, three out of four unequal-duration combinations had perceptually relevant biases from physical isochrony, with similar temporal positioning precision. In line with Schütte (1976), Terhardt and Schütte (1976), and Schütte (1978a), this finding suggests a systematic shift of the perceptual center for the 350-ms sounds. The mean target onset positions for the unequal-duration combinations with 350-ms *targets* were, on average, about 20 ms earlier than physical isochrony. The mean target onset positions for the opposite unequal-duration combinations with 350-ms *markers* were, however, on average about 30 ms later than physical isochrony. Consistent with this, the d' analysis showed larger biases (in perceptual units) in the conditions with 350-ms markers than in the conditions with 350-ms targets. The switching of target and marker durations did not result in a similarly-sized shift of the mean temporal position to the other side of physical isochrony.

Closer examination of the symmetry around physical isochrony of the mean onset positions for these opposite unequal-duration combinations (5/350 vs 350/5, and 50/350 vs 350/50) showed that mean onset positions of the [ILD] conditions were symmetrical around zero *and* that the positioning precision was equal. In contrast, both the [C] and [ITD] conditions showed an asymmetry in mean onset positions of 350/5 and 350/50 from, respectively, 5/350 and 50/350, although the positioning precision was also equal in these cases. Therefore, the explanation for this difference in the temporal positioning must be in the order of presentation of the sounds and in a difference between the [ILD] and the [C] and [ITD] conditions.

An important reason to suspect a role of loudness in this asymmetry of mean temporal positioning of opposite unequal-duration combinations is that when an accent is perceived for a tone in a pure tone sequence because it has a higher intensity, the perceived duration of the interonset interval preceding the higher intensity tone is affected (Tekman, 2001). This interval was then perceived as being longer than it was in reality, and a shorter interval was judged as being closer to physical isochrony.

There were several factors in the experimental procedure that influenced the loudness of the stimulus components. First, there was a 6-dB level difference between target and markers due to the procedure adopted. Second, perceived loudness is influenced by duration (Florentine *et al.*, 1996). Therefore, the perceived loudness difference between a 70-dB target and 76-dB markers also depends on their durations. This difference is enlarged for 5- or 50-ms targets paired with 350-ms markers, and reduced for a 350-ms target paired with 5- or 50-ms markers.

To address the influence of level cues on the mean temporal positioning for the opposite unequal-duration combinations, a control experiment was devised. If level cues played a role in mean temporal positioning, then choosing the perceived loudness of the target and marker sounds to be more similar should make the positioning of opposite unequal-duration combinations more symmetric around physical isochrony compared to the asymmetric positioning of opposite unequal-duration combinations as observed.

TABLE II. Mean onset positions and standard deviations of the adjusted temporal onset positions (in ms) for unequal target/marker duration combinations, for conditions with different (experiment 3) or similar loudness for target and marker (experiment 4). The reported values are based on 18 (different loudness) or 16 measurements (similar loudness) for each condition.

	5/350	350/5	50/350	350/50
Different loudness	37.4 (22.4)	−14.7 (22.3)	30.3 (18.7)	−12.2 (21.0)
Similar loudness	27.8 (31.8)	−31.3 (33.6)	20.9 (16.5)	−22.2 (20.5)

VI. EXPERIMENT 4

This control experiment investigated the influence of level cues on mean temporal positioning of a target onset to the meter of a regular pattern of marker onsets in quiet, for opposite unequal-duration target and marker combinations.

A. Stimuli

The same target and marker sounds as in experiment 3 were used, with the same target lateralization conditions and unequal-duration combinations of 5- and 50-ms sounds paired with 350-ms sounds. In contrast to experiment 3, the levels of the individual sounds were adjusted to achieve similar perceived loudness. For all 5-, 50-, and 350-ms targets and markers, the levels were fixed at 70, 67, and 63 dB SPL, respectively, according to the level differences in detection threshold between the three durations as described for the [C] condition in experiment 2.

B. Results

Table II displays the means and the standard deviations of the adjusted onset positions for the conditions with similar perceived loudness for target and markers (lower part). Included are the results from experiment 3, with different perceived loudness for target and markers, for reference (upper part). Of particular interest is the comparison between opposite unequal duration combinations. The 5/350 and 350/5 conditions have means of 27.8 and −31.3 ms, respectively, with statistically identical distributions ($F_{(15,15)}=1.12$, $p=0.835$). The 50/350 and 350/50 conditions have means of 20.9 and −22.2 ms, respectively, also with statistically identical distributions ($F_{(15,15)}=1.55$, $p=0.407$). In line with the results of experiment 3, the precision for conditions with the smallest ratio in target/marker duration (50/350 and 350/50) is consistently higher than for conditions with the largest ratio in target/marker duration (5/350 and 350/5).

C. Discussion

The asymmetry in mean temporal positioning around physical isochrony found for the opposite unequal-duration combinations in the previous experiment showed that there were differences in the perceived onsets for 350-ms targets and for 350-ms markers. The symmetry around physical isochrony for opposite unequal-duration combinations when level cues were minimized, unlike the situation in which level cues were available, led to the conclusion that level indeed influences mean temporal onset positioning. Although the observed bias of about 30 ms was larger than the 20-ms bias found by Schütte (1978a), this value shows a similar trend as function of stimulus duration.

The finding of an influence of level cues on mean temporal positioning is in contrast to the findings of Schütte, in which sound pressure level of the sounds presented in quiet did not influence the mean temporal positioning. The fact that Schütte did not find such an influence may result from his procedure. Unlike the current method, in which only one target was "accented" by a level difference, continuous presentation of alternating targets and references as used by Schütte may have reduced the effect of accenting, and allowed more precise judgments about the interonset interval for targets and references with different levels.

Because level cues did influence temporal perception, increasing the target level in the right ear in the [ILD] condition may have influenced perception of the interonset intervals. In the conditions with short-duration targets and long-duration markers, perceptual dissimilarity caused by the increased difference in loudness was reduced by the level increase, thereby counterbalancing the distorted perception of the interonset interval. This reduced biases in temporal positioning of the [ILD] condition compared to those of the [C] and [ITD] conditions. In the conditions with long-duration targets and short-duration markers, perceptual dissimilarity was increased by the level increase, resulting in larger biases in temporal positioning than for the [C] and [ITD] conditions. In contrast to the asymmetry found for the [C] and [ITD] conditions, symmetry was found for temporal positioning in the [ILD] condition. This symmetry may be the result of a shift along the dimension of perceptual (dis)similarity induced by level cues.

VII. EXPERIMENT 5

This experiment investigated the influence of a sound's interaural differences on the ability to temporally position its onset to the regular pattern of marker onsets, for unequal durations of target and marker sounds, and equal low target sensation levels.

A. Stimuli

The same three kinds of sounds as in experiment 2, with the same target lateralization conditions, and the same four unequal-duration combinations of 5- and 50-ms sounds with 350-ms sounds as in experiment 3 were used.

B. Results

Figure 7 shows the means and standard deviations of the adjusted temporal onset positions (ordinate), for each lateralization condition (abscissa) and for each unequal-duration target/marker combination (panels). The different symbols represent the data measured in quiet (from experiment 3, included for reference), at 6 dB, and at 3 dB target sensation level. For each unequal-duration target/marker combination, mean onset positions at low sensation levels deviated from physical isochrony in the same direction as the mean onset

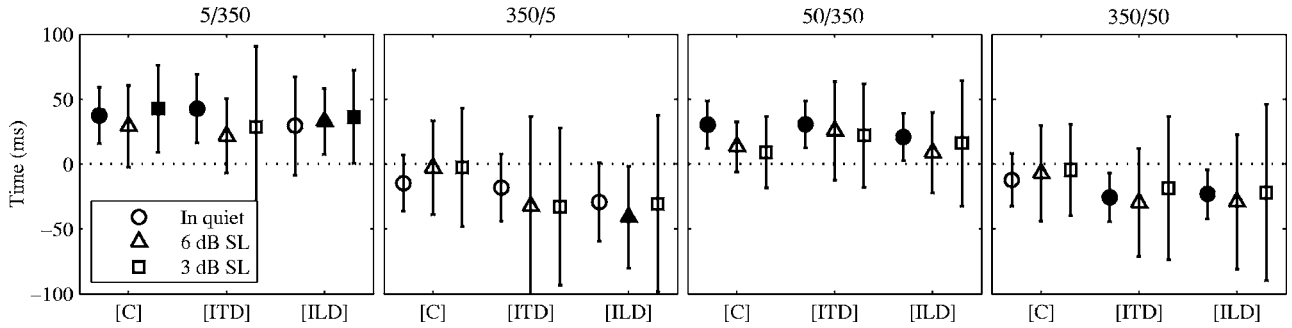


FIG. 7. Means and standard deviations of the adjusted temporal onset positions in quiet (from experiment 3, circles), at 6 dB (triangles), and 3 dB (squares) target sensation level, for each lateralization condition and unequal-duration target/marker combination. Closed symbols represent conditions with a perceptually relevant bias from physical isochrony.

positions in quiet. However, some of them closely approached physical isochrony (see the [C] conditions in the 350/5 and the 350/50 combinations). Again, larger variability in onset positions was observed at low target sensation levels than in quiet.

An analysis of variance on the means of the adjusted temporal onset positions for all three target sensation levels revealed significant effects of both the lateralization condition ($F_{(2,586)}=5.04$, $p=0.007$) and the target/marker duration combination ($F_{(3,586)}=73.68$, $p<0.001$). No significant interaction between the lateralization condition and the target/marker duration combination was found. Tukey HSD post-hoc comparisons showed that the [C] and the [ILD] conditions were significantly different from each other. All target/marker duration combinations were significantly different from each other, except for 350/5 and 350/50.

A d' analysis found that none of the unequal-duration combinations with the smallest ratio in duration showed a perceptually relevant bias from physical isochrony. The unequal-duration combinations with the largest ratio in duration showed four biases that were larger than the response variability at low sensation levels, of which three occurred in the [ILD] condition. Of these three biases, two had a standard deviation that was within 0.5 ms of the mean bias, resulting in (rounded) d' values of one.

As before, standard deviations were taken as the measure for temporal positioning *precision*. Figure 8 shows the standard deviation of the adjusted temporal onset positions (ordinate) for each sensation level (abscissa), lateralization condition (symbols) and duration combination (panels).

As shown in Fig. 8, standard deviations generally increased with a decrease in signal-to-noise ratio for all lateralization conditions and duration combinations. Standard deviations at the two low sensation levels were generally larger for the dichotic conditions than for the diotic condition, again suggesting that temporal positioning precision depended more strongly on sensation level for dichotic conditions than for the diotic condition. At 3 dB sensation level, standard deviations were larger for the combinations with 350-ms targets than for the combinations with 350-ms markers, indicating less precise temporal positioning of the long-duration targets compared to the short-duration targets.

A statistical analysis of paired comparisons between lateralization conditions was performed for each unequal-duration target/marker combination using the F ratio of the variances. Across all duration combinations together, the 36 paired comparisons showed only 5 significant differences, which all occurred between the [C] condition and either the [ITD] or [ILD] condition at 6 and 3 dB sensation level ($F_{(14-17,17)}\geq 3.13$, $p\leq 0.024$). All other distributions were statistically identical, indicating that, although visual inspection of the standard deviations for the dichotic conditions may suggest otherwise, there was no systematic influence of interaural differences on temporal positioning precision for each of the unequal-duration target/marker combinations.

C. Discussion

Because standard deviations generally increased with a decrease in target sensation level, while mean target onset

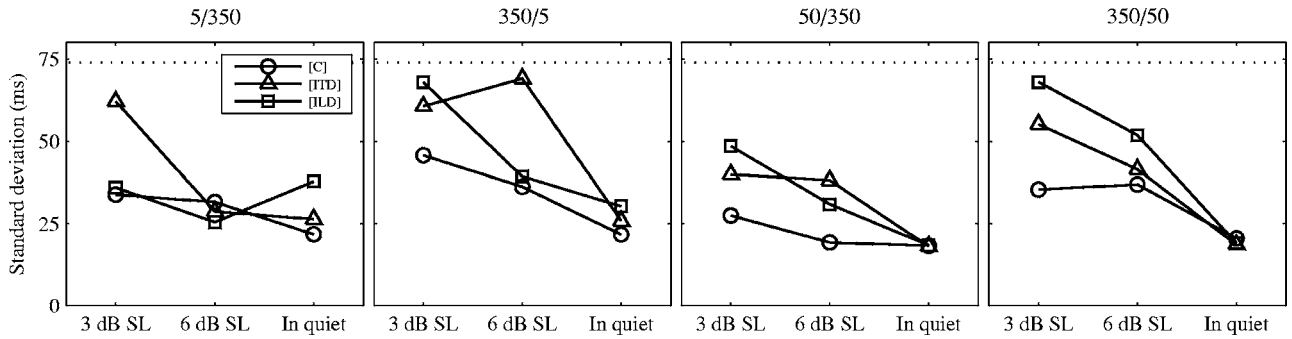


FIG. 8. Standard deviations of the adjusted temporal onset positions vs target sensation levels, for each unequal combination of target and marker durations and lateralization condition. Target sensation levels are relative to the detection threshold in the 70-dB noise masker. The dotted line at 74 ms represents the calculated standard deviation of the random distribution of initial onset positions.

positions remained rather close or even got closer to physical isochrony, the systematic biases from physical isochrony found for the unequal-duration combinations in quiet were not found in noise. In addition, two of the four observed biases had d' values of only around 1.0. In general, the obtained perceptually relevant biases did not show a systematic influence of unequal target and marker durations on mean temporal positioning or positioning precision, indicating that having unequal durations for the target and marker sounds did not significantly affect temporal perception of the target at low sensation levels.

The overall increases in standard deviation did show reduced temporal positioning precision at the two lower target sensation levels compared to the temporal positioning precision in quiet, but these increases were not systematically different for the three applied lateralization conditions, as observed in experiment 2. The larger standard deviations at equal low sensation levels for the unequal-duration combinations with 350-ms targets compared to the unequal-duration combinations with 350-ms markers may also be explained by the level of the target relative to the masker and the following energy contrast at the target's onset, as argued in the discussion of experiment 2.

With regard to the second goal of this study, to investigate whether the established dissociation of the perceived onset from the physical onset is influenced by spatial cues, the results of the experiments with unequal-duration combinations of target and markers did not reveal a significant influence of interaural differences on the perceived onsets for sounds of various durations and sensation levels.

VIII. GENERAL DISCUSSION

The current study addressed whether the previously established conclusions on the influence of interaural differences on temporal perception accuracy and precision from Schimmel and Kohlrausch (2006) can be extended to a wider range of signal durations. From the experiments described in Schimmel and Kohlrausch (2006) for 50-ms sounds, an influence of interaural time or level differences (yielding lateralization to the right) on temporal perception precision is likely to occur for the 5- and 350-ms targets at low sensation levels. Results of experiments 1 and 2 confirmed that applying the interaural differences to the 350-ms target degraded the precision with which they could be temporally positioned close to the detection threshold. Such degradation, however, could not be established for the 5-ms target. This finding is in line with the idea that temporal perception precision of a signal depends on its signal-to-noise ratio and its spatial configuration, regardless of its duration. Because the target's detection threshold varied with interaural differences and duration, the degradation of temporal perception precision of its onset occurred at different sensation levels for each of the lateralization conditions and durations. This argument could indeed explain the improvement and degradation in temporal perception precision at equal low sensation levels in experiment 2 for both shorter and longer targets than those used in Schimmel and Kohlrausch (2006).

The significant differences in temporal perception precision between the diotic and both dichotic conditions found at low suprathreshold levels are consistent with other studies showing that the binaural hearing system provides less advantage for specific perceptual tasks just above threshold than for detection. For example, Cohen (1981) measured just-noticeable differences of interaural time differences for a diotic 250-Hz sinusoidal signal in diotic, interaurally phase reversed, and statistically independent noise maskers, for signal sensation levels of 9 through 18 dB. She found that, for the conditions in which binaural unmasking improved detection, the just-noticeable differences for lateralization were higher, indicating a degraded ability to discriminate the laterality of the signal. Also, van de Par *et al.* (2005) investigated the ability of subjects to discriminate signals with different envelope structures presented in a diotic masking noise. In a diotic condition, discrimination was possible already at the detection threshold. In a corresponding dichotic condition, however, where the detection threshold was about 10 dB lower, discrimination was only possible at a level of 6 dB above detection threshold. Thus, the advantage introduced by interaural differences was only 4 dB for a discrimination task, compared to 10 dB for a detection task.

From the model of the transient behavior of the human hearing system (Schütte, 1978b), an increase in a signal's duration was expected to have two effects on its temporal perception. First, because a longer sound with rectangular envelope has a shallower perceived transient than a shorter but otherwise identical sound, the moment at which the required percentage of the maximum of the transient for perceiving the onset of the sound is reached shifts, i.e., the perceived onset becomes dissociated from the physical onset. Second, due to a shallower slope of the perceived transient for longer sounds and/or a shift in perceived onset to a shallower part of the perceived transient, this perceptual center should be less precisely defined in time. The results of experiment 3, for sounds presented in quiet with unequal durations of target and marker, showed the expected systematic shift of the perceptual center for the 350-ms sounds into their "interior" compared to the 5- or 50-ms sounds. No influence of interaural differences on the shift of the perceptual center was observed. In contrast to the findings of Schütte, level cues were found to influence mean temporal positioning. The results of both experiments 1 and 3, for sounds presented in quiet with either equal or unequal durations of target and marker, did, however, not reveal the expected difference in temporal positioning precision for the longer target and marker durations. As argued, the steepness of the perceived transients and the dissociation of the perceived onset from the physical onset, which were equal for both target and marker sounds but different for each of the three durations, were probably not different enough to reduce the precision of temporal perception. In experiments 2 and 5, for sounds presented at low sensation levels with either equal or unequal durations of target and marker, temporal positioning precision was too imprecise to identify perceptually relevant shifts in mean temporal positioning.

The fact that temporal perception precision degrades below a particular signal-to-noise ratio suggests that grouping

based on common onset may be weakened for these situations. Assuming that the temporal window for grouping signal components' onsets into an auditory object is enlarged due to the degradation of temporal precision, such a broader time window has an increased probability to include one or more increases in energy from other sources in the acoustic scene. Signal components then become more difficult to group by their temporal cues. For instance, it may take more time before conclusive evidence for their grouping is gathered and an auditory object can be formed. The signal components of a new source may be prevented from grouping, and remain part of the acoustic scene until other evidence is gathered for the existence of the source. This may be a contributing factor to the decrease of speech intelligibility in noisy environments, where accurate temporal perception of transients is required for the bottom-up processing of speech signals. Also, signal components of multiple sources that exhibit a considerable increase in energy within close temporal proximity may be incorrectly grouped into a single auditory object, i.e., result in an unsuccessful segregation. However, having different interaural differences or a longer duration of a sound lower its detection threshold and improve its sensation level. Thus, even though temporal precision is degraded below a certain signal-to-noise ratio, the interaural differences that lead to binaural unmasking still contribute to improved detection and the perceptual organization of auditory objects within an acoustic scene.

IX. CONCLUSION

The first question the current study addressed was whether the previous conclusions from Schimmel and Kohlrausch (2006), which found an influence of interaural differences on temporal perception precision, can be extended to a wider range of signal durations. The signal-to-noise ratio and the difference in spatial configuration between sound and masker were found to strongly affect the level below which the precision of the sound's temporal perception is degraded, regardless of its duration. This finding is consistent with previous conclusions about the influence of interaural differences on temporal perception precision.

The second question the current study addressed was whether the dissociation of the perceived onset from the physical onset, as observed by Schütte for diotic signals, is influenced by interaural differences. For sounds presented in quiet, the perceptual center for the 350-ms sounds shifted systematically into the interior of the sound relative to the perceptual center for the 5- or 50-ms sounds, independent of spatial configuration. Thus, it appears that the perceptual center is an intrinsic property of a sound that is independent of its location or lateralization.

ACKNOWLEDGMENTS

We would like to thank Steven van de Par, Barbara Shinn-Cunningham, Tammo Houtgast, Tom Goossens, the associate editor, and both anonymous reviewers for their constructive contributions to this paper. Furthermore, we would like to thank Professor Hugo Fastl for providing us with a copy of the Ph.D. thesis by Schütte.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Bregman, A. S., and Pinker, S. (**1978**). "Auditory streaming and the building of timbre," Can. J. Psychol. **32**, 19–31.

Cohen, M. F. (**1981**). "Interaural time discrimination in noise," J. Acoust. Soc. Am. **70**, 1289–1293.

Dannenbring, G. L., and Bregman, A. S. (**1978**). "Streaming vs. fusion of sinusoidal components of complex tones," Percept. Psychophys. **24**, 369–376.

Darwin, C. J. (**1984**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.

Darwin, C. J., and Ciocca, V. (**1992**). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.

Friberg, A., and Sundberg, J. (**1995**). "Time discrimination in a monotonic, isochronous sequence," J. Acoust. Soc. Am. **98**, 2524–2531.

Hill, N. I. (**1994**). "Binaural processing in a multi-sound environment: The role of auditory grouping cues," Doctoral dissertation, University of Sussex, Sussex, UK.

Hukin, R. W., and Darwin, C. J. (**1995**). "Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification," Percept. Psychophys. **57**, 191–196.

Marcus, S. M. (**1976**). "Perceptual centers," Doctoral dissertation, King's College, Cambridge, UK.

Pompino-Marschall, B. (**1989**). "On the psychoacoustic nature of the P-center phenomenon," J. Phonetics **17**, 175–192.

Roberts, B., and Moore, B. C. J. (**1991**). "The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony," J. Acoust. Soc. Am. **89**, 2922–2932.

Schimmel, O., and Kohlrausch, A. (**2006**). "On the influence of interaural differences on temporal perception of masked noise bursts," J. Acoust. Soc. Am. **120**, 2818–2829.

Schütte, H. (**1976**). "Wahrnehmung von subjektiv gleichmässigem rhythmus bei impulsfolgen" ("Perception of subjectively uniform rhythm in pulse trains"), Fortschritte der Akustik, DAGA '76, Heidelberg, VDI-Verlag GmbH Düsseldorf, pp. 597–600.

Schütte, H. (**1977**). "Bestimmung der subjektiven ereigniszeitpunkte aufeinanderfolgender Schallimpulse durch psychoakustische messungen" ("Determination of the perceived onset time of sequential sound impulses by psychoacoustic measurements"), Doctoral dissertation, Technical University Munich, Germany.

Schütte, H. (**1978a**). "Subjektiv gleichmässiger rhythmus: ein beitrag zur zeitlichen wahrnehmung von schallereignissen" ("Subjectively uniform rhythm: A contribution to the temporal perception of sound events"), Acustica **41**, 197–206.

Schütte, H. (**1978b**). "Ein Funktionsschema für die wahrnehmung eines gleichmässigen rhythmus in schallimpulsfolgen" ("A functional model of the perception of uniform rhythm in sequences of sound impulses"), Biol. Cybern. **29**, 49–55.

Tekman, H. G. (**2001**). "Accenting and detection of timing variations in tone sequences: Different kinds of accents have different effects," Percept. Psychophys. **63**, 514–523.

Terhardt, E., and Schütte, H. (**1976**). "Akustische rhythmus-wahrnehmung: Subjektive gleichmässigkeit" ("Perception of rhythm: Subjective equability"), Acustica **35**, 122–126.

van de Par, S., Kohlrausch, A., Breebaart, J., and McKinney, M. (**2005**). "Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 398–404.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, 2nd ed. (Springer, Berlin).

Gap detection in modulated noise: Across-frequency facilitation and interference

John H. Grose,[a] Emily Buss, and Joseph W. Hall III
Department of Otolaryngology—Head & Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7070

(Received 13 March 2007; revised 19 November 2007; accepted 1 December 2007)

This study tested the hypothesis that a detection advantage for gaps in comodulated noise relative to random noise can be demonstrated in conditions of continuous noise and salient envelope fluctuations. Experiment 1 used five 25-Hz-wide bands of Gaussian noise, low-fluctuation noise, and a noise with increased salience of the inherent fluctuations (staccato noise). The bands were centered at 444, 667, 1000, 1500, and 2250 Hz, with the gap signal always inserted in the 1000-Hz band. Results indicated that a gap detection advantage existed in continuous comodulated noise only for Gaussian and staccato noise. Experiment 2 demonstrated that the advantage did not exist for gated presentation. This experiment also showed that the advantage bore some similarity to comodulation masking release. However, differences were also noted in terms of the effects of the number of flanking bands and the absence of a detection advantage in gated conditions. The detrimental effect of a gated flanking band was less pronounced for a comodulated band than for a random band. This study indicates that, under some conditions, a detection advantage for gaps carried by a narrow band of noise can occur in the presence of comodulated flanking bands of noise. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2828058]

PACS number(s): 43.66.Mk [RLF] Pages: 998–1007

I. INTRODUCTION

The detection of a temporal gap carried by only one of the components in a complex stimulus is typically more difficult than if many of the components carry the gap synchronously. For example, Green and Forrest (1989) found that sensitivity to a gap imposed upon a single tone in a 21-tone complex was poorer than if the gap was imposed upon 11 of the 21 tones simultaneously. In a similar vein, Hall *et al.* (2007) observed that sensitivity to a gap carried by one 50-Hz-wide narrow band of noise in a concurrent complex of four such bands was poorer than if all four bands carried the gap synchronously. Other work has used as a baseline the detectability of a gap in a single noise band presented in isolation. Relative to this baseline, gap detection performance typically improves if additional noise bands are added that also contain the synchronous gap (Grose and Hall, 1988, 1997; Grose, 1991; Hall *et al.*, 1996). At the same time, gap detection performance typically declines if additional noise bands are added that do not contain the gap (Grose and Hall, 1993a; Moore *et al.*, 1993).

Although some of this work has focused on the manner in which temporal cues are combined across frequency (e.g., Grose and Hall, 1997; Hall *et al.*, 2007), other work has focused on the modulation characteristics of the multiple noise bands comprising the stimulus complex. This interest stems in part from the phenomenon of comodulation masking release (CMR), wherein a detection advantage is conferred upon a signal masked by a narrow band of noise when additional bands of noise are present that share the same modulation pattern as the on-signal masker (for review, see Grose *et al.*, 2005a). One interpretation of CMR has been that the presence of the comodulated flanking bands allows for a detection process based upon sensitivity to across-frequency envelope decorrelation. If this interpretation is correct, it follows that such a process might confer a detection advantage on gaps in comodulated noise, relative to random noise, as well as tones. That is, if a gap is imposed on one of several otherwise comodulated bands of noise, then an across-frequency envelope decorrelation would occur for the duration of the gap, including the transition ramps. Sensitivity to this decorrelation could signal the presence of the gap. This hypothesis was tested by Grose and Hall (1993a) using a single flanking band of noise, and by Moore *et al.* (1993) using eight flanking bands of noise. Both studies found that the presence of additional bands of noise not carrying the gap—either comodulated or random—was detrimental to detection of a gap in a single target band of noise. This finding is not consistent with predictions based on CMR findings. The general conclusion was that the across-frequency effects observed were more akin to modulation detection interference (e.g., Yost and Sheft, 1989; Moore and Jorasz, 1992), where the presence of amplitude modulation at a remote frequency degrades sensitivity to temporal features of the target stimulus envelope. Although it might be argued that the signal duration in the CMR paradigm is longer than the typical duration of temporal gaps at threshold, this duration difference is unlikely to account for the failure to observe a gap detection benefit in the presence of comodulated flanking bands: CMR has been observed for signal durations as short as 25 ms (Schooneveldt and Moore,

[a]Author to whom correspondence should be addressed. Electronic mail: jhg@med.unc.edu

1989a), and thresholds for detecting gaps in narrow bands of noise can be over twice as long as this (e.g., Grose *et al.*, 1989).

The purpose of the present investigation was to reassess the generality of the conclusion that comodulated flanking bands do not facilitate the detection of a gap presented in a target band. The study was motivated by two converging lines of thought. The first line was the consideration that all of the gap detection studies noted previously used gated stimuli. In typical gap detection studies, the stimuli are gated on for the duration of each observation interval and, in the target observation interval, a gap is imposed at some point in the waveform. The magnitude of CMR is dependent upon the gating characteristics of the comodulated noise bands—at least when the number of bands is relatively few—such that CMR is greater when the comodulated noise bands are presented continuously than when the bands are gated on and off synchronously during the observation intervals (Hatch *et al.*, 1995). Pilot listening comparing gap detection in similar conditions of gated versus continuous noise indicated that a detection advantage appeared to be present for gaps in the presence of comodulated flanking bands for the continuous mode of presentation. The second converging line was the notion that the limiting factor in the sensitivity to gaps carried by a narrow band of noise is the perceptual similarity between the imposed gap and the on-going fluctuations inherent to the noise band, as expanded on below. If the detectability of the gap is limited by confusion with the inherent random dips in the noise, then the presence of comodulated flanking bands might disambiguate the imposed gap from the on-going fluctuations. A formal experiment was therefore undertaken to test the hypothesis that gap detection benefits from the presence of comodulated flanking bands under conditions of perceptual confusion between imposed gap and inherent fluctuation, and that the benefit is more likely to occur in continuous, rather than gated, modes of presentation.

As just noted, a necessary premise of this hypothesis is that gap threshold in the case of the signal band alone must be limited by factors other than the absolute temporal acuity of the auditory system. That is, any reduction in gap detection threshold brought about by the addition of comodulated flanking bands must imply that temporal processing is not operating at the limits of acuity in the single-band case. It has long been recognized that detection of a gap in a narrow band of noise is likely limited by a perceptual confusion between the imposed gap and the on-going, perceptually salient, fluctuations inherent to narrow bands of noise (Shailer and Moore, 1983; Eddins *et al.*, 1992). If so, it follows that manipulation of the magnitude of the confusion effect should affect the degree of benefit conferred by the presence of the comodulated flanking bands. In a study of the effect of envelope fluctuations on gap detection, Glasberg and Moore (1992) demonstrated that, as the degree of envelope fluctuation of a narrow band of noise is systematically varied by raising the envelope to powers greater or less than unity, the detectability of a gap imposed in the noise band correspondingly varies. This supports the notion that variation in sensitivity to the gap reflects the perceptual salience of the on-going fluctuations in the noise and, hence, the magnitude of the confusion effect. In the present study, the magnitude of the confusion effect was also manipulated by varying the salience of the on-going envelope fluctuations.

In summary, the purpose of this study was to test the hypothesis that a detection advantage for gaps in comodulated noise can be demonstrated in conditions of continuous noise and perceptually salient envelope fluctuations. Two main experiments were undertaken, and one supplementary control experiment. Experiment 1 employed continuous presentation of noise stimuli with differing properties of envelope fluctuation. Experiment 2 focused on gated noise presentation, but included some continuous conditions designed to test whether observed effects could be attributed to CMR-like mechanisms. The supplementary experiment tested a subset of gated conditions analogous to those used in Experiment 2 but employed gap characteristics more similar to those of previous studies.

II. EXPERIMENT 1. GAP DETECTION IN CONTINUOUS NOISE BANDS

A. Method

1. Observers

Six normal-hearing observers participated in Experiment 1, ranging in age from 26 to 56 years (mean=45 years). All had audiometric thresholds ≤20 dB HL at the octave frequencies 250–8000 Hz (ANSI, 1996). All were experienced observers in psychophysical tasks, including temporal processing tasks.

2. Stimuli

The stimuli were 25-Hz-wide bands of noise centered at 444, 667, 1000, 1500, and 2250 Hz; these frequencies are equally spaced on a logarithmic scale. The 1000-Hz band was the signal band in all cases. In the baseline condition this band was presented in isolation. In the remaining conditions, either a single flanking band was present—always centered at 1500 Hz—or all four flanking bands were present. When present, the flanking bands either had the same temporal envelope as the signal band, or all bands had independent envelopes. In all cases the gap was introduced by gating the 1000-Hz signal band off and on with a half cycle of a raised-cosine function, 42 ms in duration. The gap duration was quantified as the time between the initiation of gate offset and the initiation of gate onset. With this metric, for example, a 42-ms gap would be generated by initiating the onset ramp immediately after completion of the 42-ms offset ramp. The 42-ms ramps were chosen in part to minimize spectral splatter associated with the introduction of the gap, but also to avoid a perceptual distinction between the imposed gap and the inherent fluctuations of the signal noise band. Given the approximately 16-Hz average fluctuation rate associated with a 25-Hz-wide Gaussian noise (Rice, 1954), it was desirable to use a ramp time that was not shorter than half the 62-ms period. The notion was that, by avoiding a sharp perceptual contrast between the imposed

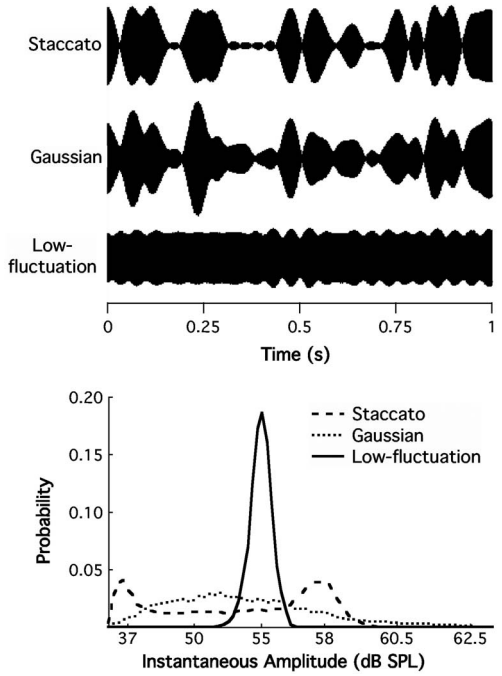


FIG. 1. Upper panel: Sample waveforms of the three types of 25-Hz wide noise bands. Lower panel: Distribution of instantaneous amplitudes for the three noise types. Relative to the Gaussian noise band, with its Raleigh envelope distribution, the staccato noise band was processed to accentuate the interrupted quality of the noise, giving it a bimodal envelope distribution. The low-fluctuation noise band was processed to minimize envelope fluctuations, giving it a narrow envelope distribution centered at 55 dB SPL. See the text for stimulus generation procedure.

gap and the inherent fluctuations of the noise, the confusion effect would be heightened by a "blending" of the gap with the on-going fluctuations.

Three sets of comodulated and random noise bands were generated that differed with respect to their modulation characteristics. The first set consisted of bands of Gaussian noise (GN), the second set consisted of bands of low-fluctuation noise (LFN), and the third set consisted of bands of staccato—or "choppy"—noise (SN). Exemplers of each type of noise are shown in Fig. 1 (upper panel: waveforms, lower panel: distribution of the instantaneous amplitudes). Each random GN band was constructed by generating a Gaussian noise sample and filtering it in the frequency domain by means of multiplication with a pair of boxcar functions placed symmetrically around the Nyquist frequency. To generate the comodulated GN bands, the 1000-Hz signal band was first generated, and then the real and imaginary components associated with this band were replicated at the spectral locations of the other four bands. This procedure results in noise bands with perfectly correlated envelopes across frequency. The LFN bands were generated following a procedure outlined by Kohlrausch *et al.* (1997, Method 1). Briefly, a Gaussian noise signal band was generated as described earlier. The Hilbert envelope was then calculated and the time-domain stimulus subsequently divided by this envelope on a point-by-point basis. This modified time-domain waveform was then transformed into the frequency domain, refiltered to the original 25-Hz passband, and transformed back into the time domain. This entire procedure was repeated eight times, resulting in a 25-Hz-wide noise band with a markedly reduced crest factor. For random LFN bands, the generation process was repeated independently for each of the four flanking bands. For comodulated LFN bands, the real and imaginary components characterizing the final rendition of the signal band were used to generate the flanking bands. The SN bands were generated using procedures analogous to those used for the LFN bands. First, a band of Gaussian noise was generated as described previously and the Hilbert envelope calculated. This envelope was used to create a target envelope comprised of zeros and ones: The target envelope was set to zero for each point in the Hilbert envelope that fell at or below 0.8 of the mean Hilbert envelope value, and the target envelope was set to one for each point above this criterion. The original GN band was multiplied by this target envelope and then refiltered in the frequency domain to the original 25-Hz bandwidth. This process was repeated 20 times to result in a 25-Hz-wide band of noise that had an accentuated interrupted quality. Random and comodulated flanking bands for the SN set were generated in the same manner as described for the GN and LFN sets.

All stimuli were generated in MATLAB (Mathworks) and stored on disk as a waveform library. Each stimulus array was 2^{18} points long, which corresponds to a 21.5 s sample when played out at a digital-to-analog conversion rate of 12 207 Hz. Because stimuli were converted from the frequency to the time domain via an inverse Fourier transform, the noise band waveforms could be output cyclically and continuously with seamless repetition. Gating was applied in software [RPvds, Tucker-Davis Technologies (TDT)]. Stimuli were played out of a real-time DSP processor (RP2, TDT) at a level per noise band of 55 dB SPL, similar to the level used in Grose and Hall (1993a). All stimuli were presented to the left earphone of a Sennheiser HD265 Linear headset.

3. Procedure

A three-alternative forced-choice (3AFC) paradigm was employed, with each of the 500-ms listening intervals separated by 450 ms. To prevent a gap from exceeding the 500-ms listening interval, a ceiling value of 499 ms was imposed on the threshold estimation procedures. The gap was temporally centered in one of the three intervals, selected at random. Thresholds were estimated using a three-down, one-up adaptive track that converged on the 79% correct point on the psychometric function. Initially, gap durations were adjusted by a factor of 1.41. This was reduced to 1.19 after the second track reversal. The track continued until eight reversals had been obtained, and the threshold estimate for a track was computed as the geometric mean of the gap duration at the last six track reversals. Observers indicated their responses by means of a handheld response box. Listening intervals were marked by box-mounted lights, and correct interval feedback was provided by the same lights.

Each observer participated in five conditions for each of the three noise band sets, for a total of 15 conditions. One condition in each set was a baseline condition using the 1000-Hz signal band presented alone. Two conditions in

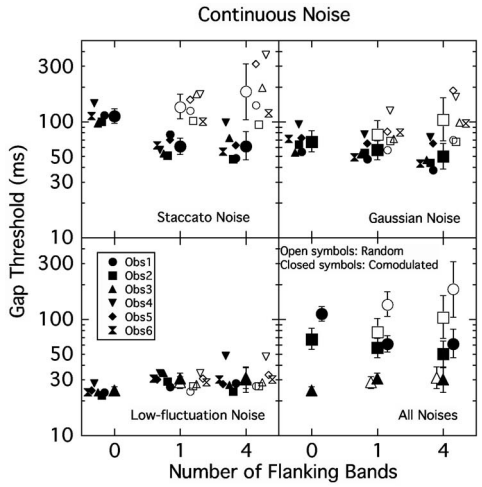


FIG. 2. Gap detection thresholds as a function of the number of flanking bands for continuous stimulus presentation. Data are shown in separate panels for each type of noise (SN, GN, and LFN), with the group means overlaid in the lower right panel (circles: SN, squares: GN, and triangles: LFN). In each panel group means are shown by large symbols, ±1 standard deviation, with filled symbols indicating comodulated conditions and open symbols indicating random conditions. Individual data points are shown clustered to the side of the respective mean in each panel.

each set used the signal band plus a single, 1500-Hz, flanking band; in one condition the flanking band was comodulated with the signal band, while in the other condition it was independent. The single flanking band used by Grose and Hall (1993a) was also centered at 1500 Hz. The final two conditions in each set used the signal band plus all four flanking bands; again, in one of these conditions all bands were comodulated, whereas in the other condition each band was independent of all the others. For each observer and condition, at least three estimates of gap threshold—but often four or five, if time permitted—were measured. The observer's final threshold for each condition was taken as the geometric mean of all estimates collected for that condition. Conditions were blocked, but the order of blocks was random across observers.

B. Results and discussion

The results of Experiment 1 are shown in Fig. 2. Three of the panels show the individual and group mean data for each of the three noise types separately; the fourth panel shows the group means for all three noise types overlaid (circles: SN, squares: GN, triangles: LFN). Three key data patterns are apparent: (1) for the two noise types with the most pronounced envelope fluctuations (SN and GN), gap thresholds in the presence of comodulated flanking bands appear lower than when the signal band is presented in isolation (i.e., a detection advantage is evident); (2) for SN and GN, gap thresholds appear lower for comodulated flanking bands than random flanking bands (i.e., the detection advantage is specific to comodulated bands); and (3) neither of the previous patterns is evident in the minimally fluctuating LFN. Before examining these patterns quantitatively, a key preliminary question is whether gap detection performance deteriorated as the degree, or salience, of the modulation increased. Recall that a premise of the experiment was that gap detection is limited by a perceptual confusion between the imposed interruption and the inherent fluctuation pattern of the stimulus such that the more salient the fluctuation pattern, the higher the gap detection threshold. Mean gap thresholds in the signal band alone condition for SN, GN, and LFN were 111, 67, and 24 ms, respectively. [For comparison, Grose *et al.* (1989) measured an average gap threshold of 70 ms for a (gated) 25-Hz-wide Gaussian noise band, with the gap bounded by 33-ms ramps.] Thresholds in the signal band alone condition were compared for the three types of noises by means of a repeated measures analysis of variance (RANOVA). Note that in this, and all analyses in this study, the log transforms of the data were used to ensure homogeneity of variance. The analysis revealed a significant effect of noise type ($F_{2,10}=515.65$; $p<0.01$), and post-hoc contrasts indicated that thresholds for the GN band were significantly higher than for the LFN band and significantly lower than for the SN band ($F_{1,5}=277.66$; $p<0.01$ and $F_{1,5}=104.74$; $p<0.01$, respectively). As expected, therefore, gap detection was best in the noise with minimal envelope fluctuation (LFN) and poorest in the noise with accentuated fluctuation (SN).

The prediction associated with the experimental hypothesis was that detection of the gap in the 1000-Hz signal band would improve with the addition of comodulated flanking bands (*sans* gap) in those cases where performance in the single band case was limited, not by absolute temporal acuity *per se*, but by a perceptual confusion effect. To test this prediction, RANOVAs were performed on the data for each of the noise types separately to determine the effects of the flanking bands within each set. For the LFN set, the RANOVA indicated a significant effect of condition ($F_{4,20}=7.10$; $p<0.01$), and post-hoc contrasts revealed that this effect was due to all conditions with flanking bands—both random *and* comodulated—yielding higher gap thresholds than the baseline signal band alone condition ($F_{1,5}$ ranging from 10.24 to 47.66; p ranging from 0.001 to 0.024). Thus, for noise bands with minimal fluctuations, where gap detection performance in the baseline condition was relatively acute, the presence of comodulated flanking bands was disadvantageous to performance, as was the presence of random flanking bands. The finding that flanking bands, even with minimal fluctuations, can be disruptive to gap detection is in line with the study of Grose and Hall (1993a) who showed that, in gated conditions, detection of a gap in a narrow band of noise can be disrupted by the presence of a pure tone not carrying the gap.

For the GN set, a RANOVA indicated a significant effect of condition ($F_{4,20}=25.16$; $p<0.01$), and post-hoc contrasts indicated that all conditions with flanking bands resulted in gap thresholds significantly different from baseline ($F_{1,5}$ ranging from 11.85 to 23.75; p ranging from 0.005 to 0.018). It is evident from Fig. 2 that these differences from baseline are due to the comodulated flanking band conditions yielding thresholds *lower* than baseline, whereas random flanking band conditions yielded thresholds *higher* than baseline. A RANOVA examining only the two factors associated with the flanking band conditions (modulation coherence and flanking band number) indicated a significant effect of modu-

lation coherence ($F_{1,5}=89.75$; $p<0.01$), but no effect of flanking band number ($F_{1,5}=1.40$; $p=0.29$). However, the interaction between these two factors was significant ($F_{1,5}=20.67$; $p<0.01$), indicating that the difference in gap threshold associated with comodulated versus random flanking bands increased as the number of flanking bands increased. This pattern of results indicates that, for narrow bands of continuous Gaussian noise, gap detection performance improves in the presence of comodulated flanking bands and deteriorates in the presence of random flanking bands.

Finally, for the SN set, a RANOVA indicated a significant effect of condition ($F_{4,20}=23.59$; $p<0.01$), but post-hoc contrasts indicated that only the conditions with comodulated flanking bands resulted in gap thresholds significantly different from baseline ($F_{1,5}=61.92$ and 44.71, respectively; $p<0.01$). To assess the effects of modulation coherence and flanking band number, a RANOVA was performed on the conditions with flanking bands. The analysis indicated a significant effect of modulation coherence ($F_{1,5}=69.69$; $p<0.01$), but no effect of flanking band number ($F_{1,5}=1.49$; $p=0.28$). The interaction between these two factors was also not significant ($F_{1,5}=3.78$; $p=0.11$). This pattern of results indicates that in the noise bands with accentuated fluctuation, the presence of comodulated flanking bands was advantageous to gap detection but performance was unaffected by the presence of random flanking bands.

The results of this experiment support the hypothesis that gap detection benefits from the presence of comodulated flanking bands under conditions of continuous presentation. However, this gap detection advantage is restricted to noise types with salient envelope fluctuations where performance is likely to be limited by a confusion effect between the imposed gap and the on-going fluctuations. In further support of this interpretation, it is interesting to note that the magnitude of the gap detection advantage in the presence of comodulated flanking bands was greater for the SN noise type than for the GN noise type. For example, the improvement in gap detection threshold when four comodulated bands were added to the signal band was 48.6 ms for the SN bands and 16.6 ms for the GN bands, a difference that is significant ($t_5=4.71$; $p<0.01$).

The finding that the presence of comodulated flanking bands can aid gap detection in continuous noise is in contrast with previous findings indicating no such effect for gated noise. It is possible that this dissimilarity is related to the difference in CMR magnitude seen for continuous versus gated presentation in conditions where relatively few flanking bands are present. However, although the gap detection advantage observed here in the presence of comodulated flanking bands is, in some respects, analogous to CMR, the pattern of data is not consistent with CMR findings. Most studies of CMR show an asymptotic dependence of the magnitude of masking release on the number of flanking bands (Carlyon *et al.*, 1989; Schooneveldt and Moore, 1989b; Haggard *et al.*, 1990; Hall *et al.*, 1990; Hatch *et al.*, 1995). In the present study, where a gap detection advantage was evident, the advantage did not consistently depend on the number of comodulated flanking bands present. Further, at a phenomenological level, the gap detection effect also appears qualitatively different from CMR: In CMR, the masked signal is generally perceived as continuous during its presentation, and the presence of comodulated flanking bands generally enhances the salience of the signal. In the gap detection task, the occurrence of the gap in the presence of the comodulated flanking bands was often described by observers as being signaled by its offset—i.e., by the onset of the signal band at the termination of the gap. In other words, the gap itself is not more salient; rather, the onset of the noise band carrying the gap became more prominent at the termination of the gap. (A typical description was that the signal band "pings on" at the conclusion of the gap.) This description is reminiscent of the auditory enhancement effect noted by Viemeister and Bacon (1982), and others. To examine the relation between comodulation and gap detection further, and to explore the relation between gap detection advantage and CMR, a second experiment was undertaken that used gated presentation and also included conditions designed to test for the presence of CMR-like behavior. Only the two types of noise that resulted in a gap detection advantage (GN and SN) were used in Experiment 2.

III. EXPERIMENT 2. GAP DETECTION IN GATED NOISE BANDS

Existing studies that have examined the detection of gaps in single noise bands in the presence of other noise bands—either comodulated or random—have found no indication of a detection advantage (Grose and Hall, 1993a; Moore *et al.*, 1993; Hall *et al.*, 2007). All of these studies used gated presentation. A goal of Experiment 2 was to determine whether the detection advantage observed in Experiment 1 was restricted to conditions that used a continuous presentation mode. A second goal was to determine whether the detection advantage associated with comodulated flanking bands in Experiment 1 was diminished under conditions that have been shown to reduce CMR. Two stimulus manipulations were implemented to test for similarities with CMR.

The first manipulation—gating asynchronies among components—has been used in studies of CMR that have focused on auditory grouping effects (e.g., Grose and Hall, 1993b; Grose and Hall, 1996; Hall *et al.*, 1997; Dau *et al.*, 2004). Because auditory grouping behavior is generally disrupted by gating asynchronies among components, reductions in CMR associated with this manipulation have been interpreted as indicating a role of perceptual fusion in the phenomenon. To determine whether the gap detection advantage associated with comodulated flanking bands shows a similar sensitivity to gating asynchronies among components, this experiment tested for the effect of an asynchrony between the signal band and the flanking bands. The second stimulus manipulation was that of embedding the gated 'core' of comodulated bands within an on-going stream of random bands. Importantly, the random bands in this temporal surround were in all respects the same as the comodulated bands in terms of frequency, bandwidth and level, but differed in that the envelope patterns were independent across bands (see Fig. 3). The use of such a random temporal surround, or fringe, as an acoustic context for comodulated

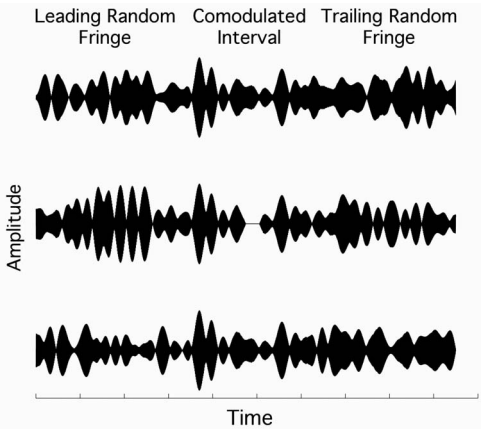


FIG. 3. Sample waveforms of three noise bands showing random temporal fringe surrounding a comodulated core. The gap is carried only by the center band.

noise bands has been used in the characterization of CMR by Grose *et al.* (2005b). They demonstrated that the magnitude of masking release for a tone due to the presence of comodulated flanking bands could be markedly reduced if the comodulated bands were embedded in a random temporal fringe. One way to view this finding is that the random temporal fringe diminishes the weight placed upon the regularity of the spectrotemporal coherence occurring during the listening intervals; i.e., spectrotemporal irregularities outside the listening intervals reduce the weight placed on across-channel cues, and this reduction persists into the listening intervals. The present experiment incorporated a similar temporal fringe within the context of a gap detection paradigm. A similar behavior of the gap detection advantage relative to CMR in this manipulation would be consistent with a common underlying mechanism.

A. Method

The same waveform library from Experiment 1 was used in Experiment 2. For gated conditions, the stimuli were presented only during the 500-ms observation intervals. The stimuli were gated on and off for the intervals using raised-cosine ramps, 50 ms in duration. The same six observers participated in Experiment 2. For each noise type (GN and SN), nine corresponding conditions were constructed. Five of these were similar to the continuous conditions of Experiment 1; that is, gap detection for gated stimuli was measured for the signal band alone, and then in the presence of either one or four synchronously gated flanking bands. These flanking bands were either comodulated with the signal band or were independent. The next three conditions were designed to probe the mechanisms underlying the detection advantage observed in Experiment 1. The first of these conditions tested for asynchrony effects by presenting the 1000-Hz signal band continuously but gating on the comodulated flanking bands only during the observation intervals. The other two conditions tested the effect of a temporal fringe by embedding the core comodulated complex consisting of the signal and four flanking bands into a temporal fringe consisting of five noise bands having the same center frequencies and bandwidths as the core bands. In one condition these fringe

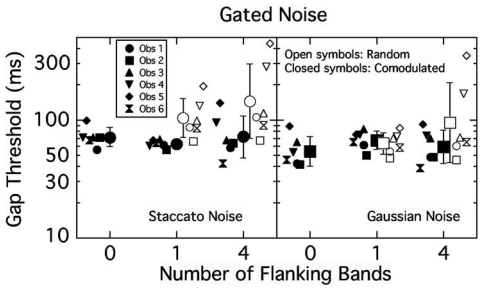


FIG. 4. Gap detection thresholds as a function of the number of flanking bands for gated noise. Data are shown in separate panels for staccato noise and Gaussian noise. Otherwise, as Fig. 2.

bands were also comodulated with respect to each other; in the other condition they were statistically similar to the core bands but were independent. This is illustrated in Fig. 3 for three noise bands, showing the comodulated core surrounded by the random fringe. The fringe bands were presented continuously except during the observation intervals that contained the core comodulated bands. The transitions from the temporal fringe to the observation intervals were accomplished by means of overlapping 50-ms ramps that rendered the transitions perceptually seamless; i.e., as the temporal fringe was gated off at the start of an observation interval, the core comodulated bands were gated on, and vice versa at the end of an observation interval. The condition where the temporal fringe consisted of five comodulated bands therefore reduced to one of continuous comodulated noise presentation. The final condition was a retest of the continuous baseline condition of Experiment 1 (signal band alone). In the rare event that an observer could not generate a reliable threshold below the ceiling value of 499 ms in a particular threshold estimation track, an estimate of 499 ms was entered for that track. This occurred in two conditions for a total of three tracks for Observer 5.

B. Results and discussion

The results of the five gated conditions that complement the continuous conditions of Experiment 1 are shown in Fig. 4 for each noise type. Two key data patterns are evident: (1) unlike Experiment 1, no gap detection advantage is evident in the presence of comodulated flanking bands and (2) there is some indication that gap detection performance is better with comodulated flanking bands than random flanking bands. Before examining these key patterns more quantitatively, a preliminary question concerns the effect of gating on gap detection performance in the baseline single band alone condition. Comparing the upper two panels of Fig. 2 to the corresponding panels of Fig. 4, there is a trend for the baseline thresholds to be lower in the gated conditions than in the continuous conditions. To test this, a two-factor RANOVA was undertaken with factors mode of presentation (gated versus continuous) and noise type (GN versus SN). Note that the data for the continuous mode of presentation used in this analysis were those collected as part of Experiment 2, so that all data in the analysis came from the same testing sessions.[1] The analysis indicated a significant effect of noise type

($F_{1,5}$=72.47; $p<0.01$), but no effect of presentation mode ($F_{1,5}$=4.27; p=0.09) and no interaction between these two factors.

To determine the effect of flanking band presence on gap detection thresholds, separate RANOVAs were performed on the GN and SN datasets. For the GN stimuli, there was a significant effect of condition ($F_{4,20}$=4.21; p=0.01). Post-hoc contrasts indicated that the presence of a single flanking band—either comodulated or random—raised thresholds significantly above baseline ($F_{1,5}$=7.48; p=0.04 and $F_{1,5}$=9.16; p=0.03, respectively), as did the presence of four random flanking bands ($F_{1,5}$=7.64; p=0.04). The presence of four comodulated flanking bands did not shift threshold from baseline ($F_{1,5}$=1.99; p=0.22). To determine the separate effects of number of flanking bands (1 or 4) and modulation type (comodulated or random), the four conditions with flanking bands were submitted to a separate RANOVA. The results indicated no significant main effect of either factor ($F_{1,5}$=0.90; p=0.39 and $F_{1,5}$=3.18; p=0.13, respectively), but the interaction between them was significant ($F_{1,5}$=11.28; p=0.02). The interaction was due to thresholds being equivalent for the case of a single flanking band—comodulated or random—but thresholds being higher with four random flanking bands than with four comodulated bands. For the GN stimuli, therefore, thresholds were generally elevated relative to baseline by the introduction of flanking bands, particularly four random flanking bands. The presence of four comodulated flanking bands did not affect threshold appreciably.

A similar analysis of data for the SN stimuli revealed a slightly different pattern of results. A RANOVA indicated a significant effect of condition ($F_{4,20}$=8.62; $p<0.01$), and post-hoc contrasts indicated that the presence of random flanking bands—either one or four—significantly elevated threshold relative to baseline ($F_{1,5}$=10.98; p=0.02 and $F_{1,5}$=7.96; p=0.04, respectively). The presence of comodulated flanking bands had no effect on threshold. A RANOVA on the conditions with flanking bands indicated no effect of number of flanking bands ($F_{1,5}$=2.59; p=0.68), but a significant effect of modulation type ($F_{1,5}$=16.74; p=0.01). The interaction term was not significant. These results indicate that, for the SN stimuli, the presence of comodulated flanking bands was not beneficial for gap detection relative to baseline; however, the presence of random flanking bands resulted in poorer performance.

Overall, the pattern of results for the gated GN and SN noise types provides an unambiguous answer to the question of whether the detection advantage seen for comodulated flanking bands in Experiment 1 was restricted to the continuous presentation mode. No detection advantage was measured for comodulated flanking bands gated synchronously with the signal band. The second question examined in this experiment was whether the detection advantage associated with continuous comodulated flanking bands in Experiment 1 could be attributed to CMR-like mechanisms. Figure 5 displays the results of the conditions designed to test this for each of the two noise types. The relevant reference condition in each panel is the continuous comodulated configuration (All Com.), which can be thought of as the core complex of

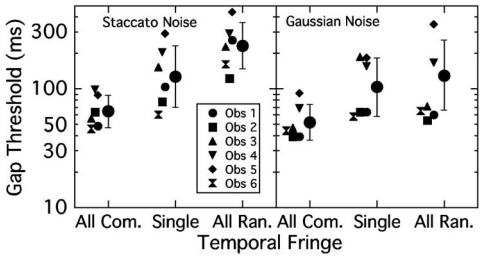


FIG. 5. Gap detection thresholds as a function of the characteristics of the temporal fringe. The core, consisting of the synchronously gated signal band and four comodulated flanking bands, had a temporal fringe of either the same five comodulated bands (All Com.), the signal band alone (Single), or five random bands (All Ran.). Data are shown in separate panels for staccato noise and Gaussian noise. In each panel group means are shown by large symbols, ±1 standard deviation, with individual data points clustered to the side of the respective mean.

five comodulated bands embedded in a temporal fringe consisting of the same five comodulated bands. Relative to this reference, gap thresholds appear elevated when the temporal fringe consists of either the signal band alone (Single) or the five random bands (All Ran.).

The effect of temporal asynchrony was measured by comparing thresholds in the continuous comodulated condition (All Com.) to those for the condition where the signal band was presented continuously whereas the comodulated flanking bands were gated on only for the observation intervals (Single). A RANOVA on the factors of synchrony status (All Com., Single) and noise type (GN, SN) indicated a significant effect of synchrony ($F_{1,5}$=22.92; $p<0.01$) and noise type ($F_{1,5}$=10.61; p=0.02), but no interaction between these factors. The elevation in threshold due to the introduction of temporal asynchrony for both noise types supports the interpretation that the gap detection advantage seen in the continuous comodulated condition bears some similarity to CMR.

To further test for commonalities with CMR mechanisms, a comparison was made between the two conditions where the core comodulated complex was embedded in a temporal fringe consisting either of comodulated bands (All Com.) or of random bands (All Ran.). The resulting RANOVA indicated a significant effect of temporal fringe type ($F_{1,5}$=82.77; $p<0.01$) and a significant effect of noise type ($F_{1,5}$=16.55; p=0.01), but no interaction between these factors. This result supports the interpretation that the gap detection advantage seen for continuous comodulated noise is sensitive to the characteristics of the temporal fringe, as is CMR for tonal signals (Grose *et al.*, 2005b). Despite this general commonality, it should be stressed that a key difference between CMR for tonal signals and the gap detection advantage noted here is that, in CMR, the detection advantage is seen for both gated and continuous presentation, and is larger for continuous presentation, whereas for gaps it is observed only for continuous presentation and is absent for gated presentation.

Although the results of Experiment 2 address the two specific questions that the experiment was designed to test, one aspect of the results remains puzzling. Previous experiments that have measured gap detection in a single band of

noise accompanied by other bands of noise not carrying the gap have consistently shown an elevation in gap threshold due to the flanking bands and, moreover, have generally found this elevation to be more pronounced for comodulated flanking bands than for random flanking bands (Grose and Hall, 1993a; Moore *et al.*, 1993). In contrast, the results of Experiment 2 showed no consistent elevation of gap threshold in the presence of comodulated flanking bands and, overall, gap thresholds were more elevated for random flanking bands than for comodulated flanking bands. There are at least two possibilities that might account for this discrepancy. The first is that, because the observers in Experiment 2 had all participated first in Experiment 1, they may have learned some cue in the continuous presentation mode of Experiment 1 that they then transferred to the gated mode of Experiment 2. If this cue is such that observers who do not get prior exposure to a continuous presentation mode are unlikely to learn it, then a different pattern of results may emerge from "unexposed" observers. This might account for the different data patterns between the Experiment 2 and other published studies (e.g., Grose and Hall, 1993a; Moore *et al.*, 1993). To assess this possibility, three naive observers were recruited and tested in a subset of the gated conditions of Experiment 2. These conditions were, for the GN noise type only: (1) signal band alone; (2) single comodulated flanking band; and (3) single random flanking band. The pattern of results from these three new observers was very similar to the data set of Experiment 2. That is, relative to the signal band alone baseline, thresholds were not affected by the presence of a comodulated flanking band, whereas thresholds were higher in the presence of a random flanking band than a comodulated flanking band. Transfer of a learned cue from continuous to gated presentation modes, therefore, does not appear to account for the data pattern of Experiment 2.

A second possibility that might account for the difference between the results of Experiment 2 and other published findings is that some of the stimulus parameters—in particular the gating characteristics of the gap—were quite different across the studies. For example, in the Grose and Hall (1993a) study, gap detection was measured in a 25-Hz-wide band of (non-Gaussian) noise presented in a broadband background noise, and gaps were imposed with 10-ms fall/rise times. Here, the signal band was presented in quiet, with the gap imposed with a 42-ms fall/rise time. Differences across studies in the gating parameters of the imposed gap, coupled with the characteristics of the noise-band carrier, may have resulted in differences in the degree of perceptual confusion between the gap and the inherent fluctuations of the noise. As discussed earlier, it would be expected that parameter settings that minimized the perceptual confusion would result in less apparent benefit from comodulated flanking bands (cf. LFN conditions of Experiment 1). In order to determine whether such stimulus characteristics underlay the performance differences, a supplementary experiment was carried out using a subset of the conditions from Experiment 2 but having stimulus characteristics more similar to those of Grose and Hall (1993a).

IV. SUPPLEMENTARY EXPERIMENT

A. Method

1. Observers

Eight normal-hearing observers participated, ranging in age from 19 to 49 years (mean=32.5 years). All had audiometric thresholds $\leqslant$20 dB HL at the octave frequencies 250–8000 Hz (ANSI, 1996). Three of the observers had participated in Experiments 1 and 2; the remaining observers were new and were trained on the gap detection task until performance was stable.

2. Stimuli

As in Experiments 1 and 2, the stimuli consisted of narrow bands of noise, each 25-Hz wide. The signal band containing the gap was always the band centered at 1000 Hz, and the remaining flanking bands, which did not contain the gap, were centered at 444, 667, 1500, and 2250 Hz. Unlike the previous experiments, the bands were generated by a method of quadrature multiplication. For each band, two independent Gaussian noises were low-pass filtered at 12.5 Hz and then multiplied by a pair of tones of the same frequency but in quadrature phase. The frequency of the tone pair determined the center frequency of the resulting band. The two independent multiplied noise bands were then summed to produce the final narrow band of noise. For random bands, independent pairs of low-pass noises were used for each noise band; for comodulated bands, the same pair of low-pass noises was used for all noise bands. The signal band could be presented either alone, or in the presence of the flanking random or comodulated bands. All stimuli were gated, and were presented in a background noise consisting of a Gaussian noise low-pass filtered at 4000 Hz and containing spectral notches centered at each of the noisebands. The notches, designed to limit spectral splatter associated with the more abrupt gap boundaries, were one normal equivalent rectangular bandwidth (ERB_N) in width (Glasberg and Moore, 1990); for the center frequencies of 444, 667, 1000, 1500, and 2250 Hz, the notch widths were 72, 94, 128, 182, and 270 Hz, respectively. The notched noise was created by means of a digital filter (TDT PD1) and presented at a level of 10 dB/Hz. The noisebands were presented at a level of 55 dB/band.

3. Procedure

The gap detection method was similar to that described in Experiment 2. A 3AFC procedure was employed, and each observation interval was 500 ms in duration; the intervals were separated by 500 ms. The gap was imposed on the signal band using a 10-ms raised-cosine fall/rise ramp, and gap duration was defined as the interval between the initiation of gate closure and the initiation of gate opening. The maximum allowable duration of the gap was 400 ms. The gap was temporally centered in the target interval, which was selected at random. Each observer participated in three conditions: (1) baseline, 1000-Hz signal band presented alone; (2) signal band plus the four comodulated flanking bands; and (3) signal band plus the four random flanking bands. At least four

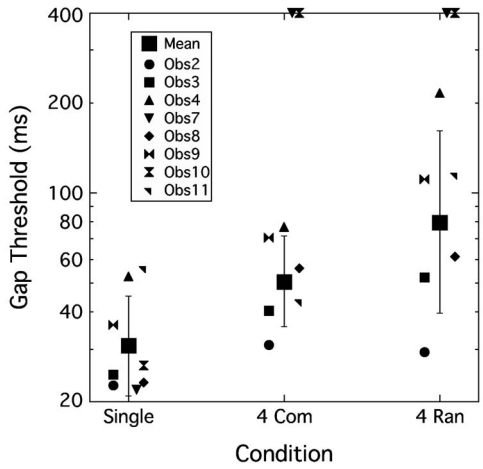


FIG. 6. Gap detection thresholds as a function of the presence and type of flanking bands: Single—no flanking bands present; 4 Com—four comodulated flanking bands; 4 Ran—four random flanking bands. Group means are shown by large squares, ±1 standard deviation, with individual data points clustered around the respective mean. Symbols on the top edge of the graph indicate ceiling performance at 400 ms.

estimates of gap threshold—but often five, if time permitted—were measured. The observer's final threshold for each condition was taken as the geometric mean of all estimates collected for that condition. The thresholds were collected first for the single band, and then for the comodulated or random conditions. The order of these latter two conditions was randomized across observers.

B. Results and discussion

The results of the supplementary experiment are shown in Fig. 6. Thresholds were lowest for the signal band alone and were elevated in the presence of flanking bands. Two of the observers could not reliably detect gaps below the 400-ms ceiling when flanking bands were present—either comodulated or random. Such individual variability in gap detection in the presence of flanking bands is striking, but individual patterns of flanking band disruption are evident also in the results of Grose and Hall (1993a) (four observers) and Moore *et al.* (1993) (two observers). The data for the six observers who generated valid thresholds for all three conditions were submitted to a RANOVA, and the analysis indicated a significant effect of condition ($F_{2,10}=13.41$; $p<0.01$). Post-hoc contrasts indicated that thresholds in the presence of flanking bands were significantly elevated with respect to that for the signal band alone for both the comodulated bands ($F_{1,5}=6.91$; $p=0.047$) and the random bands ($F_{1,5}=29.28$; $p=0.003$). However, threshold differences between the two flanking band conditions failed to reach significance ($F_{1,5}=6.12$; $p=0.056$). The finding that the presence of the gated comodulated flanking bands was detrimental to performance (unlike in Experiment 2) supports the notion that the relative effect of comodulation depends in part on the gating characteristics of the gap. Further work is required to address how the relative effect of comodulated flanking bands depends on specific parameter settings.

V. SUMMARY AND CONCLUSION

The purpose of this study was to revisit earlier conclusions that gap detection in a narrow band of noise is consistently disrupted by the presence of one or more additional bands of noise that do not contain the gap. The hypothesis tested here was that this disruption is not necessarily the general case but, rather, is restricted to conditions of gated presentation. It was hypothesized that a detection advantage for gaps could be demonstrated in the presence of comodulated noise bands under conditions where the fluctuations of the noise bands were salient and, following from the characteristics of CMR, such a benefit would be most likely observed when the mode of presentation was continuous. Experiment 1 demonstrated that, indeed, detection of a gap in a continuous narrow band of noise improved in the presence of flanking bands of comodulated noise that did not contain the gap. This benefit was restricted to noise types that exhibited salient envelope fluctuations, and was not evident for low-fluctuation noise. No gap detection benefit was conferred by flanking bands of random noise. Experiment 2 indicated that no detection advantage was evident in the gated mode of presentation. It is possible that this lack of a detection advantage in the gated mode is due in part to baseline performance occurring closer to the limits of temporal acuity, *per se*; this possibility has the associated implication that the imposed gap is, for some reason, more perceptually distinct from the inherent fluctuations of the noise in the gated mode. Recall that there was a trend for baseline performance to be better in the gated conditions than in the continuous conditions. If thresholds are not elevated due to a confusion effect in the baseline conditions, then no detection advantage can be demonstrated, as exemplified by the LFN results of Experiment 1.

The results of Experiment 2 also demonstrated that the detection advantage seen in the continuous presentation mode bore some similarity to CMR in that the advantage was sensitive to gating asynchronies across bands. Further, it was sensitive to the characteristics of the temporal fringe surrounding the core comodulated bands. However, an important difference was also noted: Whereas CMR for tones can be observed in gated conditions, a detection advantage for gaps has not been observed under these conditions. This finding influences the interpretation of the effect of the random temporal fringe. In CMR, the masking release observed in the gated, comodulated condition is eliminated by the presence of the random temporal fringe; in gap detection, the random temporal fringe is detrimental to performance in that thresholds are elevated relative to those for the comodulated temporal fringe, but this does not constitute elimination of a detection advantage since there is no detection advantage in the gated, comodulated condition to begin with.

In summary, this study has demonstrated that a detection advantage for a gap carried by a single band of fluctuating noise can be conferred by the presence of comodulated flanking bands of noise that do not contain the gap. This detection advantage is observed when the noise bands are presented continuously and when performance in the baseline condition is likely limited by a confusion effect. Al-

though it is possible that the mechanisms underlying the detection advantage are similar to those of CMR, significant differences exist between the two types of detection advantage (gap versus tone).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the comments of Associate Editor Richard Freyman, Brian Moore, and two other anonymous reviewers on a previous version of this paper. This work was supported by an award from the National Institutes of Health [NIDCD R01-DC01507].

[1]For the baseline continuous conditions that were replicated between Experiments 1 and 2, the group mean threshold for the GN band was 67 ms in Experiment 1 and 63 ms in Experiment 2. For the SN band, the corresponding thresholds were 111 and 97 ms, respectively.

ANSI. (**1996**). "American National Standards Specification for audiometers," ANSI S3-1996, American National Standards Institute, New York.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulators as a function of modulation rate," Hear. Res. **42**, 37–46.

Dau, T., Ewert, S. D., and Oxenham, A. J. (**2004**). "Effects of concurrent and sequential streaming in comodulation masking release," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer, New York).

Eddins, D. A., Hall, J. W., and Grose, J. H. (**1992**). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," J. Acoust. Soc. Am. **91**, 1069–1077.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Glasberg, B. R., and Moore, B. C. J. (**1992**). "Effects of envelope fluctuations on gap detection," Hear. Res. **64**, 81–92.

Green, D. M., and Forrest, T. G. (**1989**). "Temporal gaps in noise and sinusoids," J. Acoust. Soc. Am. **86**, 961–970.

Grose, J. H. (**1991**). "Gap detection in multiple narrow bands of noise as a function of spectral configuration," J. Acoust. Soc. Am. **90**, 3061–3068.

Grose, J. H., Eddins, D. A., and Hall, J. W. (**1989**). "Gap detection as a function of stimulus bandwidth with fixed high-frequency cutoff in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1747–1755.

Grose, J. H., and Hall, J. W. (**1988**). "Across-frequency processing in temporal gap detection," in *Basic Issues in Hearing*, edited by H. Duifhuis, H. P. Wit, and J. P. Horst (Academic, New York), pp. 308–316.

Grose, J. H., and Hall, J. W. (**1993a**). "Gap detection in a narrow band of noise in the presence of a flanking band of noise," J. Acoust. Soc. Am. **93**, 1645–1648.

Grose, J. H., and Hall, J. W. (**1993b**). "Comodulation masking release: Is comodulation sufficient?," J. Acoust. Soc. Am. **93**, 2896–2902.

Grose, J. H., and Hall, J. W. (**1996**). "Cochlear hearing loss and the processing of modulation: Effects of temporal asynchrony," J. Acoust. Soc. Am. **100**, 519–527.

Grose, J. H., and Hall, J. W. (**1997**). "Multi-band detection of energy fluctuations," J. Acoust. Soc. Am. **102**, 1088–1096.

Grose, J. H., Hall, J. W., and Buss, E. (**2005a**). "Across-channel spectral processing," Int. Rev. Neurobiol. **70**, 87–119.

Grose, J. H., Hall, J. W., Buss, E., and Hatch, D. R. (**2005b**). "Detection of spectrally complex signals in comodulated maskers: effect of temporal fringe," J. Acoust. Soc. Am. **118**, 3774–3782.

Haggard, M. P., Hall, J. W., and Grose, J. H. (**1990**). "Comodulation masking release as a function of bandwidth and test frequency," J. Acoust. Soc. Am. **88**, 113–118.

Hall, J. W., Buss, E., and Grose, J. H. (**2007**). "Spectral integration and "wideband analysis" of complex stimuli," J. Acoust. Soc. Am. **122**, 3598–3608.

Hall, J. W., Grose, J. H., and Dev, M. B. (**1997**). "Auditory development in complex tasks of comodulation masking release," J. Speech Lang. Hear. Res. **40**, 946–954.

Hall, J. W., Grose, J. H., and Haggard, M. P. (**1990**). "Effects of flanking band proximity, number, and modulation pattern on comodulation masking release," J. Acoust. Soc. Am. **87**, 269–283.

Hall, J. W., Grose, J. H., and Joy, S. (**1996**). "Gap detection for pairs of noise bands: effects of stimulus level and frequency separation," J. Acoust. Soc. Am. **99**, 1091–1095.

Hatch, D. R., Arné, B. C., and Hall, J. W. (**1995**). "Comodulation masking release (CMR): Effects of gating as a function of number of flanking bands and masker bandwidth," J. Acoust. Soc. Am. **97**, 3768–3774.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, S., van der Par, S., and Oxenham, A. J. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acust. Acta Acust. **83**, 659–669.

Moore, B. C. J., and Jorasz, U. (**1992**). "Detection of changes in modulation depth of a target sound in the presence of other modulated sounds," J. Acoust. Soc. Am. **91**, 1051–1061.

Moore, B. C. J., Shailer, M. J., and Black, M. J. (**1993**). "Dichotic interference effects in gap detection," J. Acoust. Soc. Am. **93**, 2130–2133.

Rice, S. O. (**1954**). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York).

Schooneveldt, G. P., and Moore, B. C. J. (**1989a**). "Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth and signal duration," J. Acoust. Soc. Am. **85**, 273–281.

Schooneveldt, G. P., and Moore, B. C. J. (**1989b**). "Comodulation masking release (CMR) for various monaural and binaural combinations of the signal, on-frequency, and flanking bands," J. Acoust. Soc. Am. **85**, 262–272.

Shailer, M. J., and Moore, B. C. J. (**1983**). "Gap detection as a function of frequency, bandwidth and level," J. Acoust. Soc. Am. **74**, 467–473.

Viemeister, N. F., and Bacon, S. P. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**, 1502–1507.

Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," J. Acoust. Soc. Am. **85**, 848–857.

The influence of spread of excitation on the detection of amplitude modulation imposed on sinusoidal carriers at high levels[a)]

Rebecca E. Millman and Sid P. Bacon[b)]
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 3 October 2006; revised 31 October 2007; accepted 31 October 2007)

The improvement in amplitude modulation (AM) detection thresholds with increasing level of a sinusoidal carrier has been attributed to listening on the high-frequency side of the excitation pattern, where the growth of excitation is more linear, or to an increase in the number of "channels" via spread of excitation. In the present study, AM detection thresholds were measured using a 1000-Hz sinusoidal carrier. Thresholds for modulation frequencies of 4–64 Hz improved by about 10–20 dB as the carrier level increased from 10 dB SL (14.5 dB SPL on average) to 80 dB SPL. To minimize the use of spread of excitation with an 80-dB carrier, tonal "restrictors" with frequencies of 501, 801, 1210, and 1510 Hz were used alone and in combination. High-frequency restrictors elevated AM detection thresholds, whereas low-frequency restrictors did not, indicating that excitation on the high side is more important for detecting AM. Results of modeling suggest that the improvement in AM detection thresholds at high levels is likely due to the use of a relatively linear growth of response on the high-frequency side of the excitation pattern.
 [DOI: 10.1121/1.2816575]

PACS number(s): 43.66.Mk, 43.66.Dc [JHG] Pages: 1008–1016

I. INTRODUCTION

The temporal modulation transfer function (TMTF) describes the threshold for detecting sinusoidal amplitude modulation (SAM) as a function of modulation frequency. TMTFs have been measured for sinusoidal carriers (e.g., Zwicker, 1952; Viemeister, 1976; Yost and Sheft, 1997; Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001), broadband noise carriers (e.g., Viemeister, 1979; Bacon and Viemeister, 1985), and narrowband noise carriers (e.g., Fleischer, 1982; Dau *et al.*, 1997a,b, 1999). The different carrier types have certain advantages and disadvantages in terms of measuring the TMTF (for a recent review see Moore and Glasberg, 2001). Because the inherent fluctuations in noise carriers can mask the modulations present in the signal (e.g., Fleischer, 1982; Dau *et al.*, 1997a,b), Moore and Glasberg (2001) argued that sinusoidal carriers provide a better measure of the inherent temporal resolution of the auditory system, at least for modulation frequencies where the modulation sidebands fall within the critical band for a particular carrier frequency.

One interesting difference between sinusoidal and noise carriers is how amplitude modulation (AM) detection thresholds are influenced by carrier level. Thresholds for AM applied to broadband noise carriers only improve for carrier levels up to about 25–30 dB SL, beyond which they are invariant with level (e.g., Bacon and Viemeister, 1985). Using a narrowband noise carrier (127-Hz wide centered at 1000 Hz) Maiwald (1967) found only a small (about 2 dB) decrease in AM detection thresholds for 4-Hz AM over a wide range of carrier levels (20–100 dB SPL). On the other hand, previous studies using sinusoidal carriers (e.g., Zwicker, 1952; Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001) have shown that performance generally improves with increasing carrier level up to the highest level tested. Zwicker (1952) described extensive measurements of AM detection thresholds for sinusoidal carriers over a wide range of carrier frequencies, carrier levels, and modulation frequencies. In all cases, he reported a decrease in AM detection thresholds of about 6 dB with every 30-dB increase in carrier level. Other studies have found similar results (for a review see Kohlrausch, 1993).

The improvement in AM detection thresholds for sinusoidal carriers presented at high levels has been ascribed to two mechanisms. One involves listening at a region on the high-frequency side of the excitation pattern, where the response is less compressive than it is at the peak (e.g., Zwicker, 1970; Strickland and Viemeister, 1997; Moore and Oxenham, 1998; Kohlrausch *et al.*, 2000). An alternative mechanism is an increased number of "channels" available for listening due to spread of excitation at higher carrier levels (Florentine and Buus, 1981; Buus *et al.*, 1986; Moore and Sek, 1994).

It is often assumed that listening on the high-frequency side of the excitation pattern is the dominant mechanism when subjects are listening to a modulated carrier at high levels in quiet, and that the low-frequency side of the excitation pattern makes little or no contribution (e.g., Zwicker, 1970; Strickland and Viemeister, 1997; Kohlrausch *et al.*, 2000). Contrary to this notion, Moore and Sek (1994) found that a band of noise on either side of the signal's excitation

[a)]Portions of this work were presented at the 148th Meeting of the Acoustical Society of America, San Diego, California, November 2004.
[b)]Electronic mail: spb@asu.edu

pattern had an adverse effect upon AM detection, although the band on the high-frequency side had a larger effect. These results suggest that listeners can integrate excitation from frequency regions both above and below the carrier. Moore and Sek (1994) did not investigate the effect of simultaneously presenting noise bands on both sides of the carrier, nor did they attempt to evaluate to what extent excitation on both sides of the peak of the excitation pattern was necessary to explain the improvement in AM detection threshold with increasing carrier level. Thus it is not possible to determine from their results whether the improvement in AM detection at high levels is due to listening on one or both sides of the signal's excitation pattern.

Using noise carriers, Eddins (1993) and Strickland (2000) showed that AM detection thresholds decrease as the bandwidth of the noise carrier is increased on the low-frequency side (i.e., the upper cutoff frequency of the carrier was constant). These results indicate that increasing the number of channels on the low-frequency side *can* result in improved AM detection, and thus in some ways are consistent with the results of Moore and Sek (1994). They do not, however, preclude the possibility that it is the excitation on the high-frequency side that accounts for the effects of level with sinusoidal carriers.

The primary goal of the present study was to determine whether listening *only* on the high-frequency side of the excitation pattern is sufficient to explain the improvement in AM detection with increasing sinusoidal carrier level. A secondary goal was to determine whether the advantage of listening on the high-frequency side reflects listening over a narrow region of the excitation pattern (where the response growth is linear or nearly so) or integrating over a relatively broad region (thereby increasing the number of channels). These goals were addressed using tonal "restrictors" to minimize listening in selected regions of the excitation pattern. These restrictors were presented on the high-frequency side, the low-frequency side, or both sides of the signal.

II. EXPERIMENT 1: EFFECT OF CARRIER LEVEL

The purpose of experiment 1 was to replicate previous work on the effect of carrier level on AM detection performance for sinusoidal carriers (e.g., Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001).

A. Subjects

Four subjects (with ages ranging from 19 to 27 years) participated. All had absolute thresholds better than 15 dB HL (ANSI, 1996) for octave frequencies from 250 to 8000 Hz in the ear tested. Subject S1 (first author) had prior experience in AM detection tasks and the other subjects were naive. All subjects were trained until their performance appeared to have stabilized. Subjects S2, S3, and S4 were paid for their services.

B. Apparatus and stimuli

All stimuli were digitally generated using a Tucker-Davis Technologies (TDT) system. Stimuli were produced at a 50-kHz sampling rate using a digital array processing card (TDT AP2) and one channel of a digital-to-analog converter (DAC) (TDT DD1). The output of the DAC was low-pass filtered at 20 kHz (Kemo VBF8), attenuated (TDT PA4 and/or Wilsonics PATT) and passed through a headphone buffer (TDT HB6) to one earpiece of a Sennheiser HD250 headphone. The carrier level was 80 dB SPL, 30 dB SL (approximately 35 dB SPL), or 10 dB SL (approximately 15 dB SPL). The stimulus duration was 500 ms, including 20-ms (10%–90%) raised cosine rise/fall times. The interstimulus interval was 500 ms.

On each trial, standard and target stimuli were presented successively in a random order to the subject. The standard was a 1000-Hz sinusoid. The target was a 1000-Hz sinusoid that was sinusoidally amplitude modulated. The carrier (f_c) was gated with the modulation. The equation describing the target $T(t)$ was:

$$T(t) = A_0(c(1 + m\sin(2\pi f_m t))\sin(2\pi f_c t)),$$

where A_0 is the mean amplitude of the signal; m is the modulation depth ($0 \leqslant m \leqslant 1$); and f_m is the modulation frequency (f_m=4, 8, 16, 32, 64, 128, or 256 Hz). The term c is a multiplicative compensation term (Viemeister, 1979) set such that the overall power was the same in all intervals. The expression for c is

$$c = (1 + m^2/2)^{-0.5}.$$

Absolute thresholds for a 1000-Hz sinusoid were measured in the test ear in quiet. The same stimulus duration, rise/fall times, and interstimulus interval were used as described above. The procedure was the same as described in Sec. II C, except that the task was to choose the interval containing the 1000-Hz signal and the level of this signal was varied adaptively.

C. Procedure

The experiments took place in a double-walled, sound-attenuating booth. AM detection thresholds were measured using an adaptive two-interval forced-choice procedure with a three-down, one-up stepping rule that estimates the modulation depth (20 log m) required for 79.4% correct detection. The subject had to choose the target interval, which contained the modulated carrier. Visual feedback was provided after each trial. Each run consisted of ten reversals and the threshold estimate for that run was taken as the mean of the last eight reversals. The step size (defined in terms of 20 log m) was initially 5 dB and this was reduced to 2 dB after the second reversal. A run was only included if the standard deviation of the threshold estimate was 5 dB or less. Each threshold reported is the mean of at least three runs. If the standard deviation of this threshold was greater than 3 dB, additional runs were completed and averaged until the mean threshold had a standard deviation of 3 dB or less. Out of the total 689 thresholds measured, 11 (1.5%) required additional runs to reduce the standard deviation to $\leqslant$3 dB.

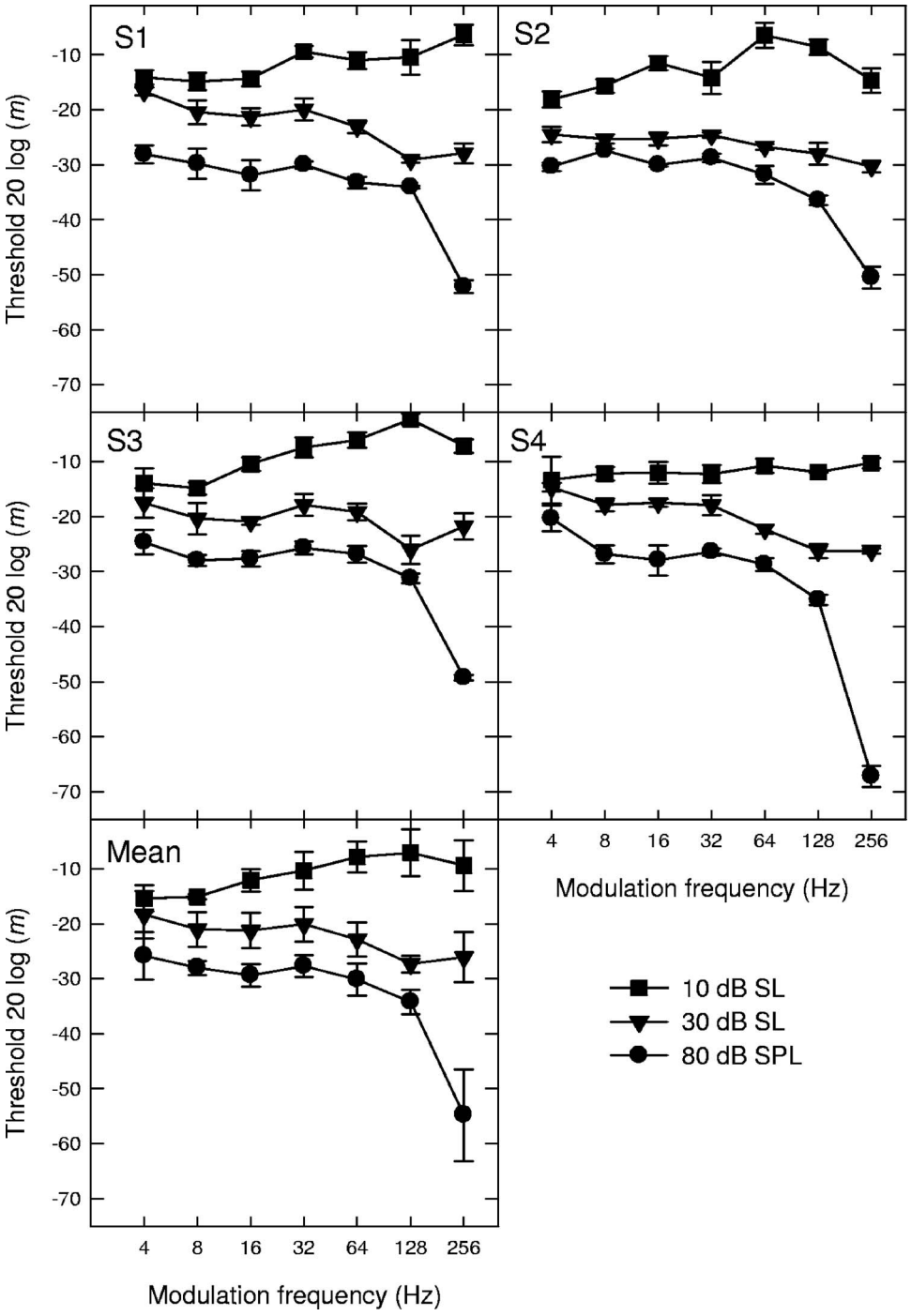


FIG. 1. Individual and mean modulation detection thresholds for a 1000-Hz sinusoidal carrier, plotted as a function of modulation frequency. The level of the carrier was 10 dB SL (squares), 30 dB SL (triangles), or 80 dB SPL (circles).

D. Results and discussion

The individual and mean TMTFs for a 1000-Hz sinusoidal carrier presented at 10 dB SL (squares), 30 dB SL (triangles), and 80 dB SPL (circles) are shown in Fig. 1. The TMTFs are roughly parallel at the two higher levels, where AM detection thresholds are more or less invariant with modulation frequency up to about 32–64 Hz. At the lowest level, however, AM detection thresholds increase above about 8 Hz. In other words, the TMTF has a much lower cutoff frequency at 10 dB SL. This may reflect a narrower auditory filter at this level (see Strickland, 2000). At the two highest carrier levels, AM detection thresholds decreased at modulation frequencies of 64 Hz and above. This suggests that listeners were detecting sidebands. At the lowest carrier level, there is less evidence for sideband detection presumably because the sidebands are inaudible (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). The description of the results will therefore concentrate on modulation frequencies below 64 Hz. Overall, the changes in the shape of the TMTF with level in Fig. 1 are consistent with those seen by Kohlrausch *et al.* (2000) for changes in the level of a 1000-Hz carrier from 20 to 75 dB SPL.

As the carrier level increased from 30 dB SL (about 35 dB SPL) to 80 dB SPL, the mean thresholds decreased by about 7–8 dB for modulation frequencies from 4 to 32 Hz. These improvements in AM detection thresholds are consis-

tent with the results of previous studies (e.g., Zwicker, 1952; Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). The aim of the remainder of this study was to evaluate the role of spread of excitation in this measured improvement.

III. EXPERIMENT 2: EFFECT OF TONAL RESTRICTORS

As noted in Sec. I, two different mechanisms have been proposed to explain the improvement in AM detection threshold with increasing carrier level: listening on the high-frequency side of the excitation pattern where the response is less compressive than at the peak, and increasing the number of listening channels due to spread of excitation (possibly above *and* below the place corresponding to the signal frequency). The objective of experiment 2 was to test the extent to which these mechanisms could explain this level effect. To do so, "restrictors" were used to minimize listening in certain regions of the excitation pattern. The effectiveness of the restrictors was measured by their ability to eliminate the 8-dB improvement in performance, measured in experiment 1, as the carrier level was increased from 30 dB SL (35 dB SPL) to 80 dB SPL.

Tonal restrictors of different frequencies and levels were used to restrict listening on the high-frequency side, the low-frequency side, or both sides of the excitation pattern. The frequencies and levels of the restrictors were chosen based on excitation pattern analysis (Glasberg and Moore, 1990) and pilot data that indicated which restrictor frequencies and levels were effective in masking AM detection for the 1000-Hz carrier.

A. Method

The same subjects participated as in experiment 1. The 1000-Hz signal carrier was presented at 80 dB SPL. The AM detection thresholds were measured as in experiment 1. Modulation frequencies were restricted to between 8 and 32 Hz because the resolution of spectral sidebands presumably played a role in the detection of the 64-Hz AM (as discussed above) and a stimulus duration of 500 ms is not optimal for the detection of 4-Hz AM (e.g., Viemeister, 1979). The only methodological difference was the addition of restrictors, presented at the frequencies and levels described below.

Tonal restrictors were created using a waveform generator (TDT WG1). They were presented continuously in order to minimize potential distracting influences on the listening task caused by gating. The restrictors were mixed with the signal (TDT SM3) and routed via the TDT HB6 to one earpiece of the headphone. The restrictor conditions were as follows:

(i) Restrictor only on the high-frequency side of the excitation pattern; the frequency was either 1210 or 1510 Hz, and it was presented at a level of 90 dB SPL;
(ii) Restrictor only on the low-frequency side of the excitation pattern; the frequency was either 501 or 801 Hz, and it was presented at a level of 70 dB SPL. These frequencies and levels were chosen as a compromise between restricting listening and minimizing upward spread of masking; and
(iii) Restrictors on both the high-frequency and low-frequency sides of the excitation pattern; the 1210- and 501-Hz restrictors, the 1210- and 801-Hz restrictors, or the 1510- and 801-Hz restrictors were presented simultaneously.

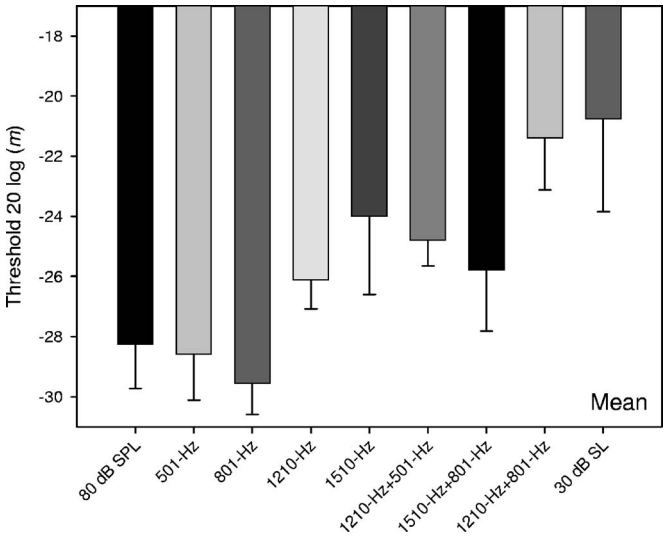


FIG. 2. Mean AM detection thresholds for the 80 dB SPL 1000-Hz carrier in the presence of a low-frequency restrictor alone, a high-frequency restrictor alone, or two restrictors presented simultaneously. The AM detection thresholds are also shown for a 1000-Hz sinusoidal carrier in quiet, presented at a level of either 80 dB SPL (left-hand side) or 30 dB SL (right-hand side). These thresholds are averaged across modulation frequency (8, 16, and 32 Hz) and listeners (N=4).

B. Results and discussion

The effects of the restrictors were similar across modulation frequency and listeners, and therefore AM detection thresholds were averaged across modulation frequency and listener. Figure 2 shows the mean results. The AM detection thresholds in quiet at carrier levels of 80 dB SPL (left-hand side) and 30 dB SL (right-hand side) are plotted to provide a reference for evaluating the effect of the restrictors on the AM detection thresholds for the 80 dB SPL carrier.

Statistical analyses were performed on the data from experiment 2. The thresholds for each listener were collapsed across modulation frequency and a repeated-measures analysis of variance (ANOVA) was performed to determine whether there were any statistically significant effects of adding the restrictors on AM detection thresholds. The main effect of the restrictor condition was significant [$F(8,304) = 26.4$, $p < 0.001$]. *Post-hoc* Student–Newman–Keuls tests were used to compare specific restrictor conditions.

First consider the effects of adding a restrictor on the low-frequency side alone. Neither the 501- nor the 801-Hz restrictor changed the mean AM detection threshold ($p > 0.3$) relative to that measured in quiet for the 80 dB SPL carrier. This is despite the fact that excitation from both low-frequency restrictors likely spreads at least somewhat into the signal carrier's excitation and that each produces some

masking of the signal carrier: The group mean detection thresholds for a 500-ms 1000-Hz tone were elevated by 21.5 dB by the 501-Hz restrictor and 24 dB by the 801-Hz restrictor. The results for the low-frequency restrictors are therefore inconsistent with the results presented by Moore and Sek (1994), which showed that a band of noise on the low-frequency side of the signal carrier reduced the detectability of AM. It is unclear whether the differences between studies are due to the type of restrictor (noise versus tone) or to differences in how restrictive the restrictors were in the two studies. The latter possibility seems likely to explain at least some of the difference. Moore and Sek (1994) used noise bands that were roughly equal in level to the carrier, and the upper cutoff frequency of their low-frequency restrictor was 826 Hz. An excitation pattern analysis (Glasberg and Moore, 1990) suggested that these noise bands were more effective in reducing the listening bandwidth of the signal carrier than were the tonal restrictors used in the present experiment.

Now consider the effects of adding a restrictor on the high-frequency side alone. The AM detection thresholds increased on average by 2 dB in the presence of the 1210-Hz restrictor and 4 dB in the presence of the 1510-Hz restrictor. The 1210- and 1510-Hz restrictors were equally effective at elevating the AM detection thresholds, as the thresholds in the presence of those two restrictors were not significantly different from one another ($p=0.09$). Thresholds in the presence of the 1210-Hz restrictor and in the presence of the 1510-Hz restrictor were significantly different from the AM detection thresholds in quiet for carrier levels of 80 dB SPL ($p<0.02$) and 30 dB SL ($p<0.002$). Thus, as seen in Fig. 2, neither restrictor eliminated the entire 8-dB improvement in AM detection thresholds as the carrier level was increased from 30 dB SL to 80 dB SPL. The high-frequency restrictors should have prevented listeners from using excitation on the high-frequency side of the signal, especially in the region of linear growth. For example, the 1210-Hz restrictor would have almost completely restricted listening on the high-frequency side of the signal's excitation pattern, apart from a small region near the signal frequency.

Finally, consider the effects of presenting a low-frequency and high-frequency restrictor simultaneously. The combination of the 1210- and 501-Hz restrictors resulted in an increase in AM detection thresholds of about 3.5 dB, but this change in performance was not significant ($p=0.29$) compared with the increase in AM detection thresholds in the presence of the 1210-Hz restrictor alone. The addition of the 801-Hz restrictor to the 1510-Hz restrictor elevated AM detection thresholds by about 4.5 dB, but the two restrictors together did not significantly change ($p=0.11$) AM detection threshold relative to that in the presence of the 1510-Hz restrictor alone. However, the addition of the 801-Hz restrictor to the 1210-Hz restrictor elevated AM detection thresholds by 7 dB and the effect of the combined 801- and 1210-Hz restrictors was significantly greater ($p<0.001$) than that for the 1210-Hz restrictor alone. Moreover, this was the only restrictor condition where the thresholds were not significantly different ($p=0.46$) from those measured with a 30 dB SL carrier. Thus, this combination of restrictors eliminated the decrease in AM detection threshold resulting from an increase in carrier level from 30 dB SL to 80 dB SPL.

In summary, the results from experiment 2 indicate that the high-frequency side is more-important than the low-frequency side for detecting AM at high carrier levels. The results, however, do not distinguish between listening at a single location versus combining across the excitation pattern to increase the number of channels. The results also suggest that listeners *can* use spread of excitation on the low-frequency side when the high-frequency side of the excitation pattern is generally unavailable, although performance is not as good as it is when using just the high-frequency side. Finally, only the combination of the 1210- and 801-Hz restrictors resulted in AM detection thresholds that were similar to the thresholds measured with a 30 dB SL carrier. These data extend previous results and suggest that the improvement in AM detection thresholds at high carrier levels can be eliminated when listening is restricted to a relatively narrow listening bandwidth around the carrier.

IV. EXPERIMENT 3: THE INFLUENCE OF CUBIC DIFFERENCE TONES

A disadvantage of using narrowband restrictors (e.g., tones or narrowband noises) is that distortion products can be generated as a consequence of the interaction between the unmodulated or modulated carrier and restrictor. In particular, modulated cubic difference tones (CDTs) could be produced by the interaction of the modulated 1000-Hz carrier with the restrictors under certain restrictor conditions. Due to the nonlinearity of CDT generation, the relative level of the sidebands around the CDT may be higher than the relative level of the sidebands around the signal carrier itself (i.e., the modulation depth of the CDT may be higher than the depth of the 1000-Hz signal). The modulation frequency of the CDT will also be twice that of the signal modulation frequency. The potential influence of CDTs provides an alternative explanation for some of the results in experiment 2 because listeners may have detected modulation at the CDT frequencies in addition to modulation at the 1000-Hz signal carrier.

In the presence of the 1210-Hz restrictor and the 1000-Hz carrier, a CDT may have been generated at 790 Hz. The observation that this high-frequency restrictor did not completely eliminate the improvement in performance measured when the carrier level was increased from 30 dB SL to 80 dB SPL could have been due to listeners detecting a modulated CDT at 790 Hz. This CDT would be far enough away from the 1210-Hz restrictor to remain unmasked by the restrictor. When the 801-Hz restrictor was combined with the 1210-Hz restrictor, it may have masked the modulated CDT at 790 Hz, thereby increasing the AM detection threshold. Thus, the effect of combining the 801-Hz restrictor with the 1210-Hz restrictor may have been mediated via masking of a CDT rather than via masking the spread of excitation on the low-frequency side of the carrier.

The presence of combination tones could also explain why the 801-Hz restrictor had no effect on AM detection threshold. The 801-Hz restrictor could combine with the 1000-Hz signal carrier to generate a CDT at 602 Hz. Thus,

even if the 801-Hz restrictor were masking spread of excitation on the low-frequency side of the signal carrier, any interference caused by that may have been offset by the presence of the CDT.

The aim of experiment 3 was to mask potential CDTs to prevent listeners from using them as a cue. Thus, low-pass noise was used in combination with the 1210-Hz restrictor and the 801-Hz restrictor.

A. Method

Data from Zwicker and Fastl (1999) suggest that the 1210-Hz restrictor and the 1000-Hz carrier would generate a CDT at 790 Hz with a level of about 45 dB SPL. For the 801-Hz restrictor and the 1000-Hz carrier, the CDT at 602 Hz would have a level of about 40 dB SPL. These may be overestimates of the CDT levels, as Zwicker and Fastl used the method of cancellation to estimate CDT levels. Shannon and Houtgast (1980) argued that the cancellation method overestimates the level of CDTs, as it does not take into account the effects of suppression of the cancellation tone by the lower primary tone, f_1, or the attenuating effect of the stapedius reflex on the external tone used to cancel the CDT (Zwicker, 1981), at least for primary frequencies greater than a few hundred Hz and primary levels greater than 80 dB SPL. Using the conservative estimates from Zwicker and Fastl (1999), we assumed the level of the CDT sidebands around 790 and 602 Hz to be about 30 dB SPL.

Detection thresholds for 8-Hz AM were remeasured in the presence of continuous low-pass noise with a spectrum level of either 25 or 30 dB. This noise should be sufficient to mask the carriers and especially the sidebands associated with the CDTs (e.g., Hawkins and Stevens, 1950). The output of a waveform generator (TDT WG1) was passed through a low-pass filter (TDT PF1) to create low-pass noise. In the 1210-Hz restrictor condition, the cutoff frequency of the low-pass filter was 790 Hz (rolloff of 30 dB/oct) and the spectrum level of the noise was 30 dB. In the 801-Hz restrictor condition, the cutoff frequency of the low-pass filter was 610 Hz (rolloff of 30 dB/oct) and the spectrum level of the noise was 25 dB. The low-pass noise was mixed (TDT SM3) with the signal and the restrictors.

Only two of the original listeners (S1 and S4) were available to participate. Their data from the appropriate conditions in experiments 1 and 2 were used here. Five additional listeners were recruited to participate in this experiment.

B. Results and discussion

The pattern of results was similar across subjects, and thus Fig. 3 shows only the thresholds averaged across the seven listeners. The AM detection thresholds in quiet at carrier levels of 80 dB SPL (left-hand side) and 30 dB SL (right-hand side) are plotted to provide a reference for evaluating the effect of the restrictors and low-pass noise.

A repeated-measures ANOVA was performed to determine whether there were any statistically significant effects of the restrictor conditions on AM detection thresholds. The main effect of the restrictor condition was significant

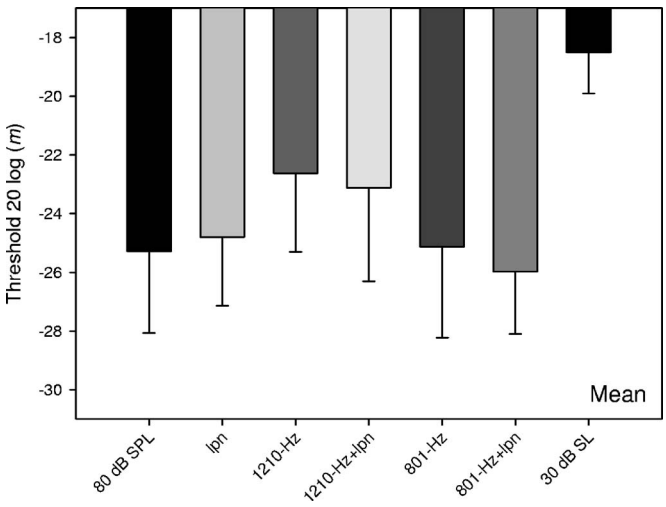


FIG. 3. Same as in Fig. 2, except low-pass noise is also present in some of the conditions. Mean AM detection thresholds are shown in the presence of low-pass noise alone, the 1210-Hz restrictor alone, the low-pass noise plus the 1210-Hz restrictor, the 801-Hz restrictor alone, and the 801-Hz restrictor plus low-pass noise. AM detection thresholds are also shown for a 1000-Hz sinusoidal carrier in quiet, presented at a level of either 80 dB SPL (left-hand side) or 30 dB SL (right-hand side). These thresholds are for a modulation frequency of 8 Hz, and are averaged across listeners ($N=7$).

[$F(6,303)=15.8$, $p<0.001$]. *Post-hoc* Student–Newman–Keuls tests were used to compare specific restrictor conditions.

The mean AM detection threshold measured with the 80 dB SPL carrier was about −26 dB and that with the 30 dB SL carrier was about −18.5 dB. The addition of the low-pass noise alone did not significantly change the mean AM threshold for the 80 dB SPL carrier ($p=0.62$). The AM detection threshold in the presence of the 1210-Hz restrictor was elevated by about 3 dB. The AM detection threshold in the presence of the 1210-Hz restrictor was significantly different from that in quiet for a carrier level of 80 dB SPL ($p=0.04$) and 30 dB SL ($p<0.001$). Adding the low-pass noise to the 1210-Hz restrictor did not produce an additional effect beyond that for the 1210-Hz restrictor alone ($p=0.61$). The AM detection threshold in the presence of the 1210-Hz restrictor plus low-pass noise was significantly different from that in quiet for a carrier level of 30 dB SL ($p<0.001$), although it was not significantly different from that for the carrier at 80 dB SPL ($p=0.07$). The AM detection thresholds in the presence of the 801-Hz restrictor alone ($p=0.65$) and in the presence of the 801-Hz restrictor plus low-pass noise ($p=0.48$) were not significantly different from AM detection thresholds measured for the 80 dB SPL carrier in quiet.

Because the AM detection thresholds in the presence of either the 1210-Hz restrictor or 801-Hz restrictor were unaffected by the addition of low-pass noise, these data strongly suggest that listeners did not use a modulated CDT at 790 Hz as a cue for AM detection in experiment 2.

V. MODEL PREDICTIONS

As discussed in the Sec. I, the improvement in AM detection thresholds for sinusoidal carriers presented at high levels has been ascribed either to listening at a region on the

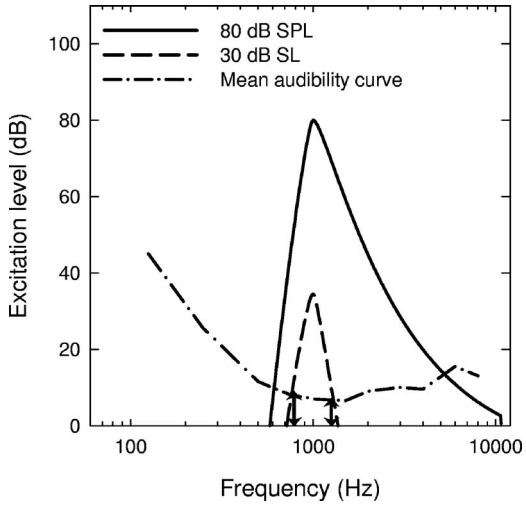


FIG. 4. Excitation patterns for the 1000-Hz sinusoidal carrier presented at 30 dB SL (heavy dashed line) and 80 dB SPL (heavy solid line). The dashed-dotted line represents the mean audibility curve. The bidirectional arrow indicates the "usable" bandwidth of excitation for the 30 dB SL carrier.

high-frequency side of the excitation pattern where the response is less compressive than it is at the peak, or an increase in the number of channels available by virtue of spread of excitation. The results of the present study clearly indicate that the high-frequency side is more important than the low-frequency side for detecting AM at high carrier levels. These results alone, however, do not support one mechanism over the other. That is, the improvement at high carrier levels could be due to the use of a fairly restricted place on the high-frequency side of the excitation pattern or due to integrating across several channels on the high-frequency side of the signal carrier. In an attempt to determine whether one explanation provides a better description of our results than the other, we calculated the improvement in AM detection threshold for both possible mechanisms.

A. Increase in number of channels

The goal of this analysis was to determine whether an increase in the number of channels resulting from an increase in carrier level from 30 dB SL to 80 dB SPL could theoretically support the roughly 8-dB improvement in threshold that was measured in experiment 1. To do this, we first estimated the number of channels available to the listeners.

Figure 4 shows excitation patterns (Glasberg and Moore, 1990) for a 1000-Hz sinusoid presented at 30 dB SL (dashed line) and 80 dB SPL (solid line), along with a mean audibility curve (dashed-dotted line) based on normal detection thresholds. The audibility curve was used to estimate the amount of "usable" excitation. The frequencies at which the excitation patterns intersect the audibility curve (indicated by bidirectional arrows for the lower level carrier) provide an estimate of the "usable" bandwidth of excitation at each carrier level. These estimates were 518 Hz (from 762 to 1280 Hz) for the 30 dB SL carrier and 4492 Hz (from 605 to 5097 Hz) for the 80 dB SPL carrier.

These usable bandwidths were converted into equivalent rectangular bandwidths (ERBs) (e.g., Moore and Glasberg, 1983); each ERB was considered as one channel. According to this, there were four channels available at the lower level and about 17 at the higher level. If we assume that the channels are independent and that the sensitivity to AM in any given channel (d'_i) is the same, then the overall sensitivity (d'_{ov}), should improve as a function of the square root of the number of channels ($\sqrt{N}$). From our results, we cannot estimate d'_i, because we do not know the sensitivity in a single channel. Instead, we have threshold for a single group of four channels, and thus can estimate sensitivity for that group. To predict the improvement in AM detection threshold with increasing carrier level (i.e., to predict the improvement when going from four to 17 channels), we assumed that each group of four channels is independent and that the sensitivity to AM in any group of four channels (d'_g) is the same. Thus the overall sensitivity (d'_{ov}) should improve as a function of the square root of the number of groups of four channels. This suggests that d'_{ov} should be about twice d'_g (i.e., $d'_{ov} = [(\sqrt{17}/4)d'_g]$).

To predict how the AM detection threshold should change with a doubling of d'_g, we used the slope of previously published psychometric functions for the detectability of low modulation frequencies (Eddins, 1993). Based on those slopes, we predict an improvement in AM detection threshold of about 3.5 dB, which is considerably less than the 8-dB improvement measured in experiment 1. This suggests that the decrease in AM detection threshold with increasing level is not simply due to increasing the number of listening channels. A caveat, however, is that the assumption of equal sensitivity to AM in each channel (or group of channels) may not be valid, as sensitivity may be greater in higher frequency channels where there is less compression

B. Listening on the high-frequency side of the excitation pattern

The goal of this analysis was to determine whether the 8-dB improvement in AM detection threshold might be understood by listening to a relatively restricted place on the high-frequency side of the signal's excitation pattern, where the response growth is less compressive than it is at the peak. To do this, we assumed that at AM detection threshold, the internal change in level (ΔL_i) is equal to 1 dB. We then used input/output (I/O) functions, like those shown in Fig. 5, to determine the external change in level (ΔL_e) needed to yield threshold ($\Delta L_i = 1$ dB). In Fig. 5, the function is either linear (solid line) or compressive with a slope of 0.3 (dashed line). As can be seen, the compressive function requires a much greater change in input level to yield a 1-dB change in output. Because psychophysical (e.g., Oxenham and Plack, 1997) and physiological (e.g., Ruggero *et al.*, 1997) studies indicate that compression in the normal auditory system is about 0.2–0.3, we calculated the change in input needed for a 1-dB change in output for compression exponents (slopes) of 0.3, 0.5, 0.7, and 1.0. These are shown in the second column of Table I.

The external change in level was converted into AM modulation depth using the following relationship:

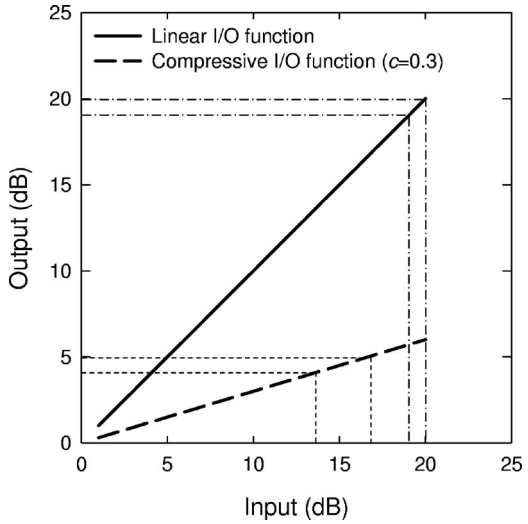


FIG. 5. Linear (heavy solid line) and compressive (heavy dashed line; slope or compression value of 0.3) I/O functions. Reference lines for the linear (light dash-dot lines) and compressive (light dashed lines) I/O functions indicate how much of a change in the input sound level is required for a 1-dB change in output.

$$\Delta L = 20 \log[(1+m)/(1-m)].$$

By solving for m, this yields predicted AM detection thresholds ($20 \log m$), which are shown in the third column of the table. These threshold values are within a few dB of those observed behaviorally. The improvement in AM detection threshold when going from a compressive I/O function with a slope value of 0.3 to a linear function (slope of 1.0) was 9.4 dB, compared to the 8-dB improvement observed in experiment 1 when the level of the carrier was increased from 30 dB SL to 80 dB SPL. This analysis indicates that the improvement in AM detection threshold could be explained by listening to a relatively restricted region on the high-frequency side of the excitation pattern, where the response growth is more linear than it is at the peak.

VI. SUMMARY AND CONCLUSIONS

The general purpose of the present study was to evaluate the role of spread of excitation in the improvement in AM detection thresholds with increasing level of a sinusoidal carrier. Previous studies have suggested that the detection of AM imposed on both narrowband noise carriers and sinusoidal carriers is largely dependent on the use of excitation above the nominal frequency of the carrier (e.g., Zwicker, 1970; Strickland and Viemeister, 1997; Kohlrausch *et al.*, 2000). The data in the present study are consistent with this possibility, as the restrictors on the high-frequency side of the carrier had a larger effect on AM detection thresholds than those on the low-frequency side. Indeed, restrictors on the low-frequency side alone had no effect on AM detection thresholds.

TABLE I. For compression or I/O slope values ranging from 0.3 to 1.0, the second column shows the change in input or external level (ΔL_e) needed for a 1-dB change in output or internal level. The third column shows the modulation depth corresponding to the change in input level.

I/O slope	(ΔL_e)	Threshold for ΔL_e $20 \log(m)$
0.3	3	−15.3
0.5	2	−18.8
0.7	1.5	−21.3
1	1	−24.7

Although listening on the high-frequency side of the signal carrier appears to be important and what we normally do, at least at high presentation levels in quiet, the results presented here suggest that listeners *can* use spread of excitation on the low-frequency side to improve AM detection when excitation on the high-frequency side is restricted. This may explain why the high-frequency restrictors alone did not completely eliminate the improvement in performance with increasing carrier level: Listening on the low-frequency side of the carrier may have offset—at least to some extent—the restricted use of excitation on the high-frequency side.

The behavioral results do not clearly indicate whether listeners are using a restricted region on the high-frequency side or are integrating across a wider region of the excitation pattern. The modeling included in this study attempted to address this issue. The results suggested that the improvement in performance with increasing carrier level could not be fully explained by integrating across multiple channels, but could be explained by listening in a restricted frequency region on the high-frequency side of the signal carrier's excitation pattern where the response growth is more linear than it is at its peak.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD Grant No. DC01376). We are grateful to John Grose and two anonymous reviewers for helpful comments on previous versions of this manuscript.

ANSI (**1996**). *Specifications for audiometers, ANSI S3.6-1996* (American National Standards Institute, New York).
Bacon, S. P., and Viemeister, N. F. (**1985**). "Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects," Audiology **24**, 117–134.
Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**). "Decision rules in detection of simple and complex tones," J. Acoust. Soc. Am. **80**, 1646–1657.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modelling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modelling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.
Dau, T., Verhey, J. L., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation detection thresholds for narrowband noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.
Eddins, D. A. (**1993**). "Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region," J. Acoust. Soc. Am. **93**, 470–479.
Fleischer, H. (**1982**). "Calculating psychoacoustic parameters of amplitude-modulated narrow noise bands," Biol. Cybern. **44**, 177–184.
Florentine, M., and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. **70**, 1646–1654.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Hawkins, J. E., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Kohlrausch, A. (**1993**). "Comment on 'Temporal modulation transfer functions in patients with cochlear implants' [J. Acoust. Soc. Am. 91, 2156–2164 (1992)]," J. Acoust. Soc. Am. **93**, 1649–1650.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.
Maiwald, D. (**1967**). "Die berechnung von modulationsschwellen mit hilfe eines funktionsschemas," Acustica **18**, 193–207.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.
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., and Sek, A. (**1994**). "Effects of carrier frequency and background noise on the detection of mixed modulation," J. Acoust. Soc. Am. **96**, 741–751.
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.
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.
Shannon, R. V., and Houtgast, T. (**1980**). "Psychophysical measurements relating suppression and combination tones," J. Acoust. Soc. Am. **68**, 825–829.
Strickland, E. A. (**2000**). "The effects of frequency region and level on the temporal modulation transfer function," J. Acoust. Soc. Am. **107**, 942–952.
Strickland, E. A., and Viemeister, N. F. (**1997**). "The effects of frequency region and bandwidth on the temporal modulation transfer function," J. Acoust. Soc. Am. **102**, 1799–1810.
Viemeister, N. F. (**1976**). "Modulation thresholds and temporal modulation transfer functions," J. Acoust. Soc. Am. **60**, 5117.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Yost, W. A., and Sheft, S. (**1997**). "Temporal modulation transfer functions for tonal *stimuli:* Gated versus *continuous conditions*," Aud. Neurosci. **3**, 401–414.
Zwicker, E. (**1952**). "Die grenzen der hörbarkeit der amplitudenmodulation und der frequenzmodulation eines tones," Acustica **2**, 125–133.
Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).
Zwicker, E. (**1981**). "Dependence of level and phase of the $(2f_1\text{-}f_2)$-cancellation tone on frequency range, frequency difference, level of primaries, and subject," J. Acoust. Soc. Am. **70**, 1277–1288.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models* (Springer-Verlag, Berlin).

Binaural processing of modulated interaural level differences

Eric R. Thompson[a] and Torsten Dau[b]
Centre for Applied Hearing Research, Acoustic Technology, Ørsted.DTU, Technical University of Denmark, Building 352, Ørsteds Plads, 2800 Kgs. Lyngby, Denmark

(Received 13 September 2006; accepted 9 November 2007)

Two experiments are presented that measure the acuity of binaural processing of modulated interaural level differences (ILDs) using psychoacoustic methods. In both experiments, dynamic ILDs were created by imposing an interaurally antiphasic sinusoidal amplitude modulation (AM) signal on high-frequency carriers, which were presented over headphones. In the first experiment, the sensitivity to dynamic ILDs was measured as a function of the modulation frequency using puretone, and interaurally correlated and uncorrelated narrow-band noise carriers. The intrinsic interaural level fluctuations of the uncorrelated noise carriers raised the ILD modulation detection thresholds with respect to the pure-tone carriers. The diotic fluctuations of the correlated noise carriers also caused a small increase in the thresholds over the pure-tone carriers, particularly with low ILD modulation frequencies. The second experiment investigated the modulation frequency selectivity in dynamic ILD processing by imposing an interaurally uncorrelated bandpass noise AM masker in series with the interaurally antiphasic AM signal on a pure-tone carrier. By varying the masker center frequencies relative to the signal modulation frequency, broadly tuned, bandpass-shaped patterns were obtained. Simulations with an existing binaural model show that a low-pass filter to limit the binaural temporal resolution is not sufficient to predict the results of the experiments. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821800]

PACS number(s): 43.66.Pn, 43.66.Mk [RLF] Pages: 1017–1029

I. INTRODUCTION

Information in sound signals is carried not only by the fine structure of the sound, but also by the intensity fluctuations of its envelope. In a reverberant environment, reflections can reduce the depth of those envelope fluctuations and can change their phase. The effective amount of envelope modulation and modulation phase transmitted to a receiver can be derived from the source–receiver impulse response as a function of the modulation frequency (Schroeder, 1981). This complex modulation transfer function (MTF) shows the modulation attenuation and phase shift as a function of modulation frequency for the particular source–receiver transmission path. A normal human auditory system has two working ears, thereby receiving information from a given source via two transmission paths and through two MTFs. Interaural differences in the modulation phase and/or depth can create fluctuating interaural level differences (ILDs) and interaural time differences (ITDs). In order to understand how ILD fluctuations are perceived and to begin to understand the binaural processing of envelopes in reverberation, artificial stimuli were generated in the present study with sinusoidal amplitude modulation and a controlled interaural modulation phase difference. The stimuli were presented over headphones to listeners in psychoacoustic tests.

An ILD is usually perceived as a lateralization of the sound source toward the ear with the higher intensity sound. When an ILD changes slowly, the sound is perceived to move, while more rapid ILD fluctuations are usually perceived as a stationary sound source with a broad or diffuse sound image (e.g., Blauert, 1972; Grantham, 1984; Griesinger, 1997). This is analogous to the ability of the auditory system to follow slow intensity fluctuations in monaural or diotic stimuli, and the perception of roughness with more rapid fluctuations (e.g., Terhardt, 1968).

Dynamic ILDs can be created by imposing amplitude modulation with an interaural modulation phase difference. However, a static interaural modulation phase difference can also be interpreted as a static envelope ITD, corresponding to the phase difference divided by the angular modulation frequency. High-frequency sounds, which cannot be lateralized based on the ITD of their fine structure (e.g.,Klumpp and Eady, 1956; Mills, 1960), can be lateralized based on the ITD of their envelopes (e.g., Klumpp and Eady, 1956; Henning, 1974; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1994). A static modulation phase difference could then create a percept of a sound lateralized toward the leading ear instead of creating a moving or diffuse sound image, depending on the envelope ITD. However, with a phase difference of π, as was used in the present study, it is unclear which ear should be leading because of the temporal symmetry of the sinusoid. For complex sounds with random interaural level fluctuations, those fluctuations may actually be encoded internally as a combination of time-varying ITDs and ILDs. In situations with ambiguous localization cues, such as with a π interaural phase difference, onset cues may dominate localization of the ongoing signal (Buell *et al.*, 1991).

The temporal acuity of the auditory system is often measured by determining the threshold of detection of the sinusoidal modulation of a physical parameter as a function of the modulation frequency, referred to as a "temporal modu-

[a]Electronic mail: et@oersted.dtu.dk.
[b]Electronic mail: tda@oersted.dtu.dk.

lation transfer function" (TMTF). For example, the TMTF with diotic amplitude modulation (AM) was measured by Viemeister (1979) with broadband noise carriers, by Fleischer (1982) and Dau *et al.* (1997a) with narrow-band noise carriers, and by Kohlrausch *et al.* (2000) with pure-tone carriers. Other studies have investigated the temporal acuity to dynamic interaural parameters, such as interaural time or phase differences (e.g., Grantham and Wightman, 1978; Witton *et al.*, 2000) and interaural correlation (e.g., Grantham, 1982). Grantham (1984), Grantham and Bacon (1991), and Stellmack *et al.* (2005) measured the acuity of the binaural system in the detection of modulated ILDs, generated with interaurally antiphasic sinusoidal AM signals.

The TMTFs measured with pure-tone and diotic broadband noise carriers show a high sensitivity to slow AM, with a minimum modulation depth, m, required for detection of around 0.04 (often discussed on a decibel scale as $20\log_{10}m$, here −28 dB) for low frequency modulations. As the modulation rate increases, larger modulation depths are required for detection, thereby exhibiting an overall low-pass characteristic (e.g., Viemeister, 1979; Kohlrausch *et al.*, 2000). However, the thresholds measured with narrow-band noise carriers can exhibit high-pass as well as low-pass characteristics, depending on the bandwidth of the noise (Fleischer, 1982; Dau *et al.*, 1997a). This led Dau and colleagues to propose a modulation filterbank model with bandpass filters acting on the envelope of a stimulus (Dau *et al.*, 1997a, b), which can simulate AM detection performance with narrow-band as well as broadband noise. Bacon and Grantham (1989), Houtgast (1989), and Ewert *et al.* (2002) made more direct measurements of modulation frequency selectivity by measuring AM detection thresholds in the presence of a noise AM masker. These measurements also showed a bandpass characteristic with approximately constant filter bandwidth relative to the filter center frequency (constant Q value).

Grantham (1984) also reported a low-pass shape in his ILD modulation detection thresholds. Those data were obtained through measurements of the threshold of *discriminability* of interaurally antiphasic AM from homophasic AM imposed on interaurally uncorrelated, bandpass noise carriers. In contrast to the diotic TMTFs described earlier, the modulation depths required to discriminate the ILD modulation with low-frequency AM were quite high, around $m=0.15$ (−16 dB). In another study, Grantham and Bacon (1991) measured monaural and ILD modulation *detection* thresholds with broadband noise carriers and unmodulated reference intervals. Their monaural AM detection thresholds were very similar to those from Viemeister (1979) with thresholds of around −28 dB for low modulation frequencies. The ILD modulation detection thresholds were almost identical to the monaural thresholds, thereby showing 12 dB greater sensitivity to the modulation than reported by Grantham (1984). This increase in performance can be attributed to the difference in paradigm (AM detection versus discrimination). Since relatively small AM depths can be detected monaurally, characterization of binaural processing of modulated ILDs should only be done with the elimination of the monaural AM cues through an AM discrimination paradigm [as done by Grantham (1984)].

Grantham and Bacon (1991) also measured monaural and binaural frequency tuning in the envelope domain by measuring the TMTF in the presence of an AM masker. One diotic broadband-noise carrier, with a diotic tonal or narrow-band-noise amplitude modulator (the masker), was added to a second diotic broadband-noise carrier with an interaurally antiphasic sinusoidal amplitude modulator (the signal). They reported a bandpass tuning in the masked detection thresholds with the tonal modulator, but could not conclude whether that tuning was due to monaural or binaural processing. However, with a noise modulator masker, they did not see an effect of masker bandwidth and this led them to argue against a binaural modulation frequency tuning. Grantham (1984) described diotic AM as creating an "up-and-down flutter" with perceived changes in level or roughness, and antiphasic AM as creating a "side-to-side flutter" with a perception of motion or broadening between the ears. Assuming that the detection of the signal interval in their 1991 study was based on a comparison of the perceived motion or width of the two presentation intervals, a *diotic* masker, with no interaural fluctuations itself, would be perceived as motionless and narrow, and should have had little effect on the detectability of the modulated ILD signal. A *dichotic* masker, which does generate interaural fluctuations, would be better to measure the masked sensitivity to ILD modulations, along with a task of discriminating between interaurally antiphasic and homophasic AM ($AM_\pi - AM_0$), where the monaural cues have been made ambiguous. In this way, the results and any modulation frequency tuning could be attributed to purely binaural processing.

Stellmack *et al.* (2005) measured the sensitivity to ILD modulations using high-frequency (5 kHz) pure-tone and narrow-band noise carriers (30 and 300 Hz wide, interaurally correlated and uncorrelated), and an $AM_\pi - AM_0$ discrimination task. The thresholds measured with the pure-tone carrier were approximately constant at about −20 dB up to about $f_m=100$ Hz, where the sensitivity worsened with increasing f_m until the threshold could no longer be determined above $f_m=500$ Hz. There was a small increase in thresholds (up to 7 dB with the 30-Hz-wide carrier) when using the correlated noise carriers, relative to the pure-tone carrier data, particularly with low modulation rates ($f_m<20$ Hz). This increase was much smaller than the increase in thresholds seen with uncorrelated noise carriers, and was described as independent of the intrinsic carrier fluctuations. Therefore, the main focus of their paper was on the thresholds measured with uncorrelated narrow-band noise carriers.

Narrow-band Gaussian noises fluctuate randomly with envelope frequencies up to the bandwidth of the noise (see, e.g., Lawson and Uhlenbeck, 1950; Price, 1955). Monaurally, those inherent fluctuations can make it more difficult to detect an imposed AM, as compared to AM imposed on a pure-tone (i.e., flat envelope) carrier, especially when the AM frequency is less than the bandwidth of the noise carrier. This increase in threshold can be viewed as the result of masking of the signal AM by the intrinsic envelope fluctuations of the carrier. Binaurally, presenting interaurally uncor-

related narrow-band noises to each ear creates a dynamic ILD and the perception of a randomly moving or broad sound, depending on the bandwidth. The modulation spectrum of the ILD fluctuations is governed by the frequency content of the envelopes of the stimuli. The difference between the ILD modulation thresholds measured with uncorrelated and correlated noise carriers by Stellmack *et al.* (2005) also showed that the intrinsic ILD fluctuations from the uncorrelated carriers had the largest effect on thresholds for AM frequencies within the bandwidth of the noise. This suggests that there might be a frequency-selective mechanism in the processing of ILD fluctuations, similar to the monaural modulation AM processing from Dau *et al.* (1997a), but with broader frequency tuning.

The goal of the present study was to further investigate the modulation frequency tuning of the processing of ILD fluctuations. In the first experiment, detailed in Sec. III, the measurements of sensitivity to modulated ILDs from Stellmack *et al.* (2005) with narrow-band noise carriers were repeated with an additional carrier bandwidth (3 Hz, added to the 30- and 300-Hz-wide carriers) and a lower modulation frequency range (2–128 Hz instead of 4–600 Hz). A 3-Hz-wide Gaussian noise carrier has intrinsic modulations that can easily be followed (as loudness fluctuations monaurally, or as motion interaurally), where the 30- and 300-Hz-wide carriers are perceived with more roughness or width from the higher intrinsic modulation frequencies. The addition of the 3-Hz-wide carrier also enabled a comparison with all three carrier bandwidths (3, 31, and 314 Hz wide) used by Dau *et al.* (1997a). With the additional data from the present study, a different interpretation of the results than that of Stellmack *et al.* is proposed, which includes more emphasis on the threshold differences with diotic carriers.

Section IV details a second experiment for directly measuring the modulation frequency tuning for ILD fluctuations, using experimental design elements from Bacon and Grantham (1989), Houtgast (1989), Ewert and Dau (2000), and Ewert *et al.* (2002). In addition, simulations were made with an existing binaural computational model (from Breebaart *et al.*, 2001a), which was designed mainly for static interaural conditions, but includes a sliding integrator window (low-pass filter) to limit the temporal resolution. This enables it to predict some signal detection thresholds with dynamic interaural conditions (see Breebaart *et al.*, 2001c). The simulations should show whether an existing binaural model can predict similar thresholds to those of a human listener when used as an artificial observer.

II. GENERAL METHODS

Two psychoacoustic experiments were performed in order to investigate the sensitivity of the binaural system to modulated ILDs. In both experiments, the listener's task was to discriminate between a stimulus with interaurally antiphasic AM (AM_π; subscript indicating the interaural modulation phase) and a stimulus with homophasic (diotic) AM (AM_0). The AM frequency and depth was the same in all stimulus intervals of a three-interval, three-alternative forced-choice (3-AFC) trial, with only an interaural difference in modulation phase in the signal interval. The stimuli were defined as in Eq. (1) with carriers x_L and x_R (subscripts L and R for left and right ears, respectively):

$$\text{Left: } [1 + m\sin(2\pi f_m t + \phi_L)]x_L(t),$$

$$\text{Right: } [1 + m\sin(2\pi f_m t + \phi_R)]x_R(t), \tag{1}$$

where m is the modulation depth, f_m is the modulation frequency, and $\phi_{L/R}$ is the initial modulation phase for the respective ear's stimulus. The reference intervals were defined by Eq. (1) with $\phi_R = \phi_L$ (AM_0) and the signal interval was defined with $\phi_R = \phi_L + \pi$ (AM_π). The instantaneous ILD of a stimulus is defined as the ratio of the envelopes (as in Stellmack *et al.*, 2005):

$$\mathrm{ILD}(t) = 20\log_{10}\left(\frac{E_L(t)}{E_R(t)}\right). \tag{2}$$

A diotic sound does not create any ILDs itself. Therefore, the instantaneous ILD with an AM_π signal imposed on a diotic carrier is simply the ratio of the modulators:

$$\mathrm{ILD}(t) = 20\log_{10}\left(\frac{1 + m\sin(2\pi f_m t + \phi_L)}{1 - m\sin(2\pi f_m t + \phi_L)}\right) \tag{3}$$

and the maximum ILD is a function of the modulation depth only:

$$\mathrm{ILD}_{\max} = 20\log_{10}\left(\frac{1+m}{1-m}\right). \tag{4}$$

Interaurally uncorrelated noises produce stochastic ILD fluctuations, which add linearly (on a decibel scale) to the deterministic signal ILD modulation. These random ILD fluctuations will change the distribution of ILDs and the maximum ILD of the stimulus.

The two experiments differed in specifics of the stimuli (e.g., modulation phase and carrier), which will be presented in Secs. III and IV, but the general methods were the same.

A. Test subjects

Four test subjects were used for all experiments. They were not paid directly for their participation, but were all involved at the research center, and included both authors of this paper. All had pure-tone audiometric thresholds of 15 dB HL or better for octave frequencies between 250 Hz and 8 kHz. They were all experienced in psychoacoustic measurements and particularly in AM detection experiments. That experience was mostly with monaural or diotic stimuli, so their experience with the detection of interaural fluctuations was limited. All listeners were encouraged to listen to example stimuli and performed a limited set of training runs of approximately 1 h duration.

B. Equipment

All signals were generated and presented using the AFC-Toolbox for MATLAB (Math-Works), developed at the University of Oldenburg, Germany and the Technical University of Denmark, at a sample rate of 44.1 kHz through a sound card (RME DIGI 96/8 PAD) and headphones (Sennheiser

HD-580). The test subjects sat in a sound insulated booth with a computer monitor, which displayed instructions and visual feedback, and a keyboard for response input.

C. Procedure

A 3-AFC paradigm was used with an adaptive one-up, two-down tracking rule, which should converge at the 70.7% correct point on the psychometric function (Levitt, 1971). In a given track, the modulation frequency of the AM signal was fixed and the modulation depth was varied to find the AM depth required for identification of the signal. During each trial, a computer monitor displayed a window with three buttons, representing the three stimuli. Each button was highlighted when the corresponding interval was played. The signal interval was randomly selected with equal probability of occurrence from the three presentation intervals. The test subject responded via the computer keyboard and received immediate feedback on whether the response was correct or incorrect. All tracks were assembled in one long experiment that the test subject could start and stop at will after the completion of any track. The typical duration of a session was about 30 min, but could be longer or shorter depending on the circumstances.

Each track started with a modulation depth of −2 dB ($20\log_{10} m$). The step size started at 4 dB, and was halved after every second reversal until the final step size of 1 dB was reached after the fourth reversal. The track continued for six further reversals with this step size, and the threshold was determined as the mean of those last six reversals. Each test subject completed four repetitions for each modulation frequency and set of experimental parameters, and the results shown are the mean and standard deviation of all test subjects and repetitions. If the test subject could not identify the correct interval with the maximum modulation depth ($m=0$ dB) twice in a row, the track was skipped and the experiment continued to the next track.

D. Common stimulus parameters

All stimuli were centered at 5 kHz, so that all frequency components would lie well above the range of frequencies in which interaural timing differences in the fine structure of the carriers would affect the lateralization of the stimuli (e.g., Klumpp and Eady, 1956; Mills, 1960). The stimuli were gated simultaneously in the two ears, and presented at a level of 65 dB SPL, with 300 ms of silence between intervals. In tracks where noise carriers were used, a new noise sample was generated for each presentation.

III. EXPERIMENT I: MODULATION DISCRIMINATION WITH NARROW-BAND NOISE CARRIERS

The first experiment was designed to measure the sensitivity of binaural processing to modulated ILDs, using an experimental design based on the diotic AM detection measurements from Dau *et al.* (1997a). Parts of this experiment are a repetition of similar experiments performed by Stellmack *et al.* (2005).

A. Specific stimulus details

ILD modulations were created by applying an interaurally antiphasic sinusoidal AM to pure-tone and narrow-band noise (3-, 30-, and 300-Hz-wide) carriers, centered at 5 kHz. In order to eliminate monaural modulation cues, an $AM_\pi - AM_0$ discrimination paradigm was used. The noise carriers were generated by creating a 1-s-long independent Gaussian noise sample for each interval in the time domain and setting the frequency components outside of the passband to zero in the spectral domain. Measurements were made with interaurally correlated (symbol N_0) and uncorrelated (N_u) noise carriers. Sinusoidal AM was applied to the carrier as given in Eq. (1) with $\phi_L=\phi_R=0$ (AM_0) in the reference intervals and $\phi_L=0$ and $\phi_R=\pi$ (AM_π) in the signal interval. With this choice of AM phase parameters, the change in modulation phase in the right ear could have been used as a monaural cue for signal detection. Therefore, a control experiment was performed to measure the modulation depth required for discrimination of monaural AM phase change (referred to as AM_m disc) using a pure-tone carrier and the right ear only. For these tracks, the signal interval had an initial modulation phase of π (negative-going zero crossing) and the reference intervals started with a modulation phase of zero (positive-going zero crossing).

Thresholds were measured with AM frequencies (f_m) in octave steps from 2 to 32 Hz and 128 Hz. At the highest modulation frequency used (128 Hz), an interaural modulation phase difference of π is equivalent to an ITD of ± 3.9 ms. Since this is well above the ecologically relevant range of ITDs [approximately 650 μs for humans, Feddersen *et al.* (1957)], and because the test subjects reported informally hearing a diffuse sound image, and not a static lateralization of the sound, it is assumed that listeners did not use the static envelope ITD to localize the sound source. Therefore, the experiments will be discussed in terms of dynamic ILDs. Tracks with $f_m \geqslant 8$ Hz had a stimulus duration of 500 ms, including 50 ms $\cos^2$ onset and offset ramps, while tracks with $f_m=2$ or 4 Hz had a duration of 1000 ms in order to reduce the interference of the windowing function on the desired envelope frequency components [the stimulus windows of Stellmack *et al.* (2005) were 1 s long with 150 ms ramps]. The intervals were separated by 300 ms of silence, where the intervals of Stellmack *et al.* (2005) were only demarcated by the ramps with no additional separating silence. Previous measurements from Dau (1996) showed that listeners could not reliably discriminate (defined there as $P_c > 33\%$) between monaural AM phase with full modulation ($m=1$) for $f_m > 12$ Hz, using a 5 kHz pure-tone carrier. Therefore, the AM_m phase discrimination threshold was only measured in the present study with $f_m \leqslant 8$ Hz. More recent results from Sheft and Yost (2007) showed that some listeners could only discriminate modulation starting phase with broadband noise carriers up to about $f_m=12.5$ Hz, while others were still able to perform the task up to around 50 Hz.

Two additional control measurements were made with an AM *detection* paradigm, where only the signal interval in a three-interval trial had an applied AM and the two reference intervals were unmodulated. Monaural AM_m and inter-

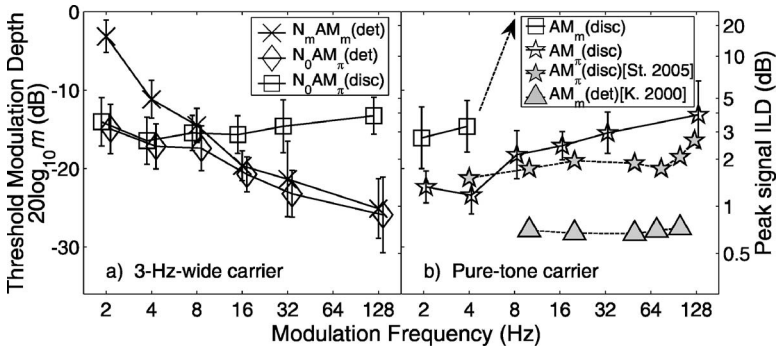


FIG. 1. Amplitude modulation detection and discrimination thresholds in decibels for various monaural and binaural conditions. The left ordinate labels (modulation depth) apply to all curves in both panels, and the right ordinate labels (peak ILD) applies to the binaural (AM_π) thresholds in both panels. (a) Thresholds measured with a 3-Hz-wide noise carrier. Monaural AM detection (N_mAM_m, X's), ILD modulation detection (N_0AM_π, diamonds), and ILD modulation discrimination (N_0AM_π–N_0AM_0, squares). (b) Thresholds measured with a pure-tone carrier for monaural AM phase discrimination (AM_m, squares) and ILD modulation discrimination (AM_π–AM_0, open stars). The data shown with shaded symbols are ILD modulation discrimination [shaded stars, adapted from Stellmack *et al.* (2005), Fig. 3], and monaural AM detection thresholds [shaded triangles, adapted from Kohlrausch *et al.*, (2000), Fig. 2, 5 kHz carrier]. Note that the data in (a) are offset around the AM frequency for visual clarity of the error bars.

aural AM_π detection thresholds were measured with a 3-Hz-wide carrier in order to demonstrate the importance of eliminating monaural cues in the binaural experiment. Diotic noise carriers were used for this AM_π detection measurement.

B. Results

The results from the four test subjects were similar in shape and value, so the plots shown in Figs. 1–3 display the mean and standard deviation over all test subjects and runs. The modulation depths required for detection or discrimination of monaural AM and modulated ILDs are plotted in decibels ($20\log_{10}m$) as a function of the signal modulation frequency for the pure-tone [Fig. 1(b)] and narrow-band noise (Fig. 2) carriers. Note that the ordinates are shown with larger modulation depths and therefore poorer sensitivity toward the top of the axis. In Figs. 1 and 2, there is also a second ordinate on the right of each plot showing the peak *signal* ILD in decibels, which applies to all of the AM_π data. This peak ILD was calculated from the modulation depth at threshold according to Eq. (4). As discussed earlier, the actual peak ILD seen with the uncorrelated noise carriers varied around this value. In the following, the results are presented in terms of the modulation depth at threshold in decibels, unless otherwise noted. The data points are offset slightly from the AM frequency in Figs. 1(a), 1(b), 2(b), and 2(d) for visual clarity of the error bars. A two-way analysis of variance (ANOVA) with repeated measures was used to

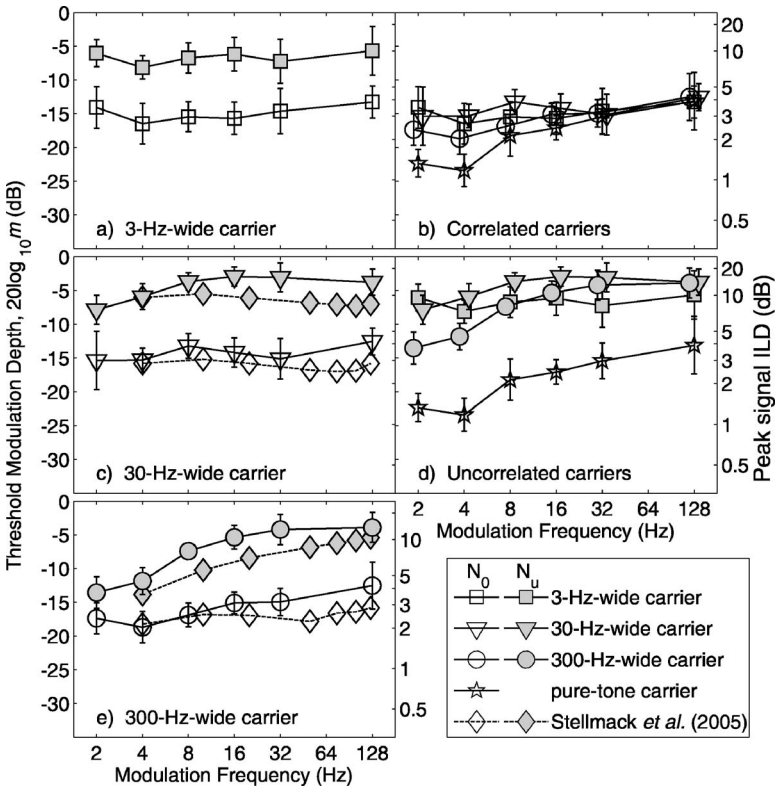


FIG. 2. ILD modulation discrimination thresholds measured with narrow-band noise carriers. The left and right ordinate labels apply to all curves in all panels, with the AM depth (left ordinate) converted to peak ILD in the right ordinate according to Eq. (4). In all panels, the interaural correlation of the carriers is indicated by the shading. Open symbols indicate correlated carriers and shaded symbols indicate uncorrelated carriers. The symbols indicate the carrier bandwidth: Squares for 3 Hz wide, triangles for 30 Hz wide, and circles for 300 Hz wide. The pure-tone thresholds from Fig. 1(b) are replotted with open star symbols in (b) and (d). External data from Stellmack *et al.* (2005) are indicated by diamonds for both 30- and 300-Hz-wide data. The thresholds for the 3-, 30-, and 300-Hz-wide carriers are grouped by band-width in (a), (c), and (e), respectively. The data measured with correlated carriers are replotted in (b), and with uncorrelated carriers in (d). Note that the data in (b) and (d) are offset around the AM frequency for visual clarity of the error bars.

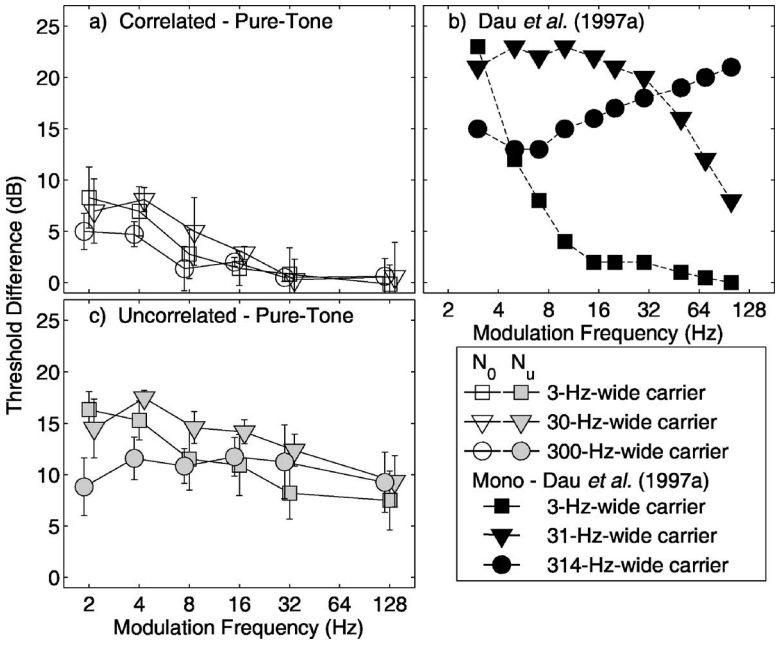


FIG. 3. The difference in discrimination thresholds measured with narrow-band noise carriers and pure-tone carriers [AM_π, stars from Fig. 1(b)]. (a) The difference for thresholds measured with correlated noise carriers [N_0AM_π, from Fig. 2(b)]. (b) The difference between monaural AM detection with narrow-band noise and pure-tone carriers [data adapted from Dau *et al.* (1997a)]. (c) The difference for thresholds measured with uncorrelated noise carriers [N_uAM_π, from Fig. 2(d)]. The error bars in (a) and (c) show the standard deviation of the mean threshold difference across listeners.

compare the data curves with a threshold of $p<0.05$ required for significance.

Figure 1(a) shows the thresholds obtained for the monaural N_mAM_m detection (X's), N_0AM_π detection (diamonds), and N_0AM_π discrimination (squares) with a 3-Hz-wide carrier. The monaural curve shows the AM frequency selectivity described by Dau *et al.* (1997a) in that the highest thresholds (−3 dB at $f_m=2$ Hz) are at a frequency within the bandwidth of the noise (i.e., $f_m<3$ Hz), and the thresholds for AM detection decrease with increasing f_m (down to −25 dB at $f_m=128$ Hz). Discrimination of AM_π from AM_0 with the 3-Hz-wide carrier is approximately constant with thresholds between −16 and −13 dB for all measured f_m. This shows that while the low-frequency modulations ($f_m<8$ Hz) are obscured by the fluctuations of the carrier in each ear (up-and-down), the ILD modulation (side-to-side) is still easily detectable with harmonic ILD oscillations with amplitudes of 2.8–4 dB [see the right ordinate in Fig. 1(a)]. The N_0AM_π *detection* threshold demonstrates how the auditory system switches from monaural to binaural cues, depending on cue salience. For low AM frequencies ($f_m<8$ Hz), where the monaural cues are obscured by the envelope fluctuations of the 3-Hz-wide carrier, there is no significant difference between the N_0AM_π detection and discrimination thresholds. For $f_m>8$ Hz, where the carrier fluctuations have a relatively small influence on the monaural detectability, the N_0AM_π detection and N_mAM_m thresholds have no significant differences.

In the control experiment to ensure that the test subjects could not perform the binaural discrimination tasks based solely on a monaural AM phase cue, thresholds for monaural AM phase discrimination could only be measured for $f_m\leqslant4$ Hz [squares in Fig. 1(b)]. At $f_m=8$ Hz, the discrimination task could not reliably be performed by any of the test subjects, even at full modulation depth ($m=0$ dB) and the arrow indicates that no threshold was measurable. These data correspond well to those from Dau (1996) and with some of the listeners from Sheft and Yost (2007), but cannot help to explain how some listeners in the latter study were still able to discriminate modulation starting phase at rates up to about 50 Hz. The thresholds for monaural modulation phase discrimination in the present study for the 2- and 4-Hz AM signals showed which of the results from the other measurements presented in this experiment (Figs. 1 and 2) could have been influenced by a monaural modulation phase cue. Those measurements were repeated informally with only the right ear's signal to verify that the tasks could not be performed monaurally at the measured threshold levels, and none of the listeners tested were able to do so.

The ILD modulation discrimination threshold curve measured with a pure-tone carrier [open stars in Fig. 1(b)] shows an overall low-pass tendency with thresholds around −23 dB (1.2 dB peak ILD) for $f_m=2$ and 4 Hz and increasing to about −13 dB (4 dB peak ILD) for $f_m=128$ Hz. Stellmack *et al.* (2005) reported a flatter threshold shape with thresholds around −20 dB (1.7 dB peak ILD) from $f_m=4$ to almost 100 Hz [shaded stars in Fig. 1(b)], above which the threshold increased. The monaural TMTF with a pure-tone carrier [shaded triangles in Fig. 1(b)], reported by Kohlrausch *et al.* (2000), shows that the auditory system is much more sensitive to envelope fluctuations (thresholds around −28 dB up to $f_m>100$ Hz) than to the ILD fluctuations caused by an interaural modulation phase inversion.

The data measured with narrow-band noise carriers (3, 30, and 300 Hz wide) are plotted twice in Fig. 2. The left panels show the data grouped by carrier bandwidth, and the right panels by carrier interaural correlation. In all panels, the squares represent data for the 3-Hz-wide carrier, the triangles for the 30-Hz-wide carrier, and the circles for the 300-Hz-wide carrier data. Open symbols indicate interaurally correlated carriers (N_0), and shaded symbols are for interaurally uncorrelated carriers (N_u). In addition, the corresponding data from Stellmack *et al.* (2005) for the 30- and 300-Hz-wide carriers are shown in Figs. 2(c) and 2(e), re-

spectively, with diamonds and dashed lines, and the pure-tone (AM_π) thresholds are replotted from Fig. 1(b) in Figs. 2(b) and 2(d) with stars. The 30- and 300-Hz-wide carrier data show a good agreement with the data from Stellmack *et al.* (2005), with only small differences that could be the result of the procedural differences discussed earlier (e.g., stimulus length, windowing).

By grouping the thresholds by carrier bandwidth [Figs. 2(a), 2(c), and 2(e)], a strong effect of the interaural carrier correlation can be seen. The N_uAM_π thresholds are much higher than the N_0AM_π thresholds. With the 3-Hz-wide carriers [Fig. 2(a)], the N_uAM_π thresholds show an almost constant offset of about 8 dB from the N_0AM_π curve. The differences between the thresholds measured with wider bandwidth carriers [Figs. 2(c) and 2(e)] increase with increasing modulation frequency from 7 to 12 dB with the 30-Hz-wide carriers and from 4 to 11 dB with the 300-Hz-wide carriers.

A two-way ANOVA with repeated measures was calculated on the N_0AM_π data [Fig. 2(b)] with carrier bandwidth and modulation frequency as factors. The analysis showed no significant effect of bandwidth ($p=0.13$), but a significant effect of modulation frequency ($p<0.05$) and a significant interaction between bandwidth and modulation frequency ($p<0.05$). Adding the pure-tone AM_π data to the analysis as an additional bandwidth yielded a significant effect of bandwidth ($p<0.01$). An ANOVA on the data measured with interaurally uncorrelated noise carriers [N_uAM_π; Fig. 2(d)] showed a significant effect of bandwidth, even without the pure-tone carrier data, and of modulation frequency, as well as a significant interaction between the factors ($p<0.01$ for both factors and the interaction).

C. Discussion

The modulation depths required to discriminate AM_π from AM_0 imposed on diotic noise carriers were significantly larger than those required with pure-tone carriers, particularly with low modulation rates ($f_m<16$ Hz). The discrimination thresholds measured with interaurally uncorrelated noise carriers were even higher than those measured with the correlated noise carriers. The increase in thresholds when using noise carriers instead of pure-tone carriers can be considered as masking of the modulation signal by the intrinsic envelope fluctuations of the noise carriers themselves, since a pure-tone carrier does not have these random fluctuations. Figure 3(a) shows the *difference* in thresholds measured with correlated noise carriers and the pure-tone carrier. Note that this reflects an increase in the threshold of ILD modulation discrimination as the result of *diotic* fluctuations of the noise carriers, which do not create ILD fluctuations themselves. Looking at the differences between the N_uAM_π thresholds and the pure-tone carrier thresholds [plotted in Fig. 3(c)], the 3- and 30-Hz-wide carriers show the greatest difference for f_m below the bandwidth of the carrier, beyond which the difference decreases monotonically toward an asymptotic value of about 7–9 dB. The threshold difference with the 300-Hz-wide carrier increases from about 8 dB with $f_m=2$ Hz to 12 dB with $f_m=16$ Hz, and then decreases again to 9 dB at $f_m=128$ Hz.

Stellmack *et al.* (2005) also observed an increase in $AM_\pi-AM_0$ discrimination thresholds with the diotic 30- and 300-Hz-wide carriers, but did not report an effect of the carrier-envelope frequency content on the shape of the increase, and therefore focused on the difference between the thresholds with uncorrelated and correlated noise carriers. However, there is a significant interaction between the diotic noise bandwidths and the modulation frequency in the present study, which can be seen in the threshold differences for $f_m<16$ Hz [see Fig. 2(b)]. These differences could suggest a dependence on the carrier envelope spectrum, but this needs to be investigated in further studies.

A control experiment was performed to investigate the difference between the thresholds measured in the AM_π discrimination experiments with the pure-tone and the diotic narrow-band noise carriers. The ILD modulation discrimination threshold was measured with a 30-Hz-wide "low-noise noise"[1] (Pumplin, 1985) carrier. Measuring with a low-noise noise carrier tests the hypothesis that the difference in thresholds between AM_π discrimination with a pure-tone carrier and with a narrow-band noise carrier is caused by the envelope fluctuations of the noise and not by the noise's broader bandwidth per se. A pairwise t-test of the results showed no significant difference between thresholds measured with the low-noise noise carrier and the pure-tone carrier (thresholds at −19.0 and −17.8 dB, respectively, $p=0.43$), but there was a significant difference between the low-noise noise and the Gaussian noise thresholds (−19.0 and −14.7 dB, respectively, $p<0.05$). This suggests that it is the fluctuations in level of the Gaussian noise carriers that impede the discrimination of AM_π from AM_0.

The diotic Gaussian noise carrier and the pure-tone carrier do not create any ILDs themselves. Therefore, the binaural sensitivity to the AM_π signal is only limited by the internal variability of the auditory system, or by "internal noise." Since there is a significant increase in thresholds when using diotic Gaussian noise carriers, this suggests that the internal noise increases when the envelopes of the carriers fluctuate, or that the encoding of fluctuating envelopes is noisier than the encoding of steady envelopes. The range of modulation frequencies over which there is an increased threshold with the correlated noise carriers, and the differences between the thresholds from the three carrier bandwidths should provide some insight into how the interaural processing differences can be modeled.

Interaurally uncorrelated Gaussian noise carriers cause large stochastic fluctuations in ILD. This "external" ILD variability is in addition to the internal noise described earlier. Therefore, it is not surprising that the ILD modulation discrimination thresholds are higher with N_uAM_π than with N_0AM_π. It is unknown how the effects of the external and internal variances with Gaussian noise carriers combine in the auditory system. Therefore, the data obtained with *pure-tone* carriers were used as the reference threshold in the present study. This is in contrast to Stellmack *et al.* (2005), who compared the thresholds with uncorrelated and correlated noise carriers, leaving out the extra effect of the diotic

level fluctuations on the measurements. The thresholds measured with the uncorrelated carriers are up to 18 dB higher than those measured with the pure-tone carrier, particularly at AM frequencies below the bandwidth of the carrier. By comparing with the pure-tone carrier thresholds instead of with the diotic noise carriers, the shapes of the threshold difference curves with uncorrelated noise carriers from Fig. 3(c) have similar aspects to those from monaural experiments [Fig. 3(b); adapted from Dau *et al.* (1997a)], but also large differences. The monaural curves [Fig. 3(b)] with 3 and 31-Hz-wide carriers drop off quickly toward zero with f_m greater than the carrier bandwidth, indicating relatively sharp modulation frequency tuning. The binaural curves [Fig. 3(c)] also roll off with f_m greater than the carrier bandwidth, but do not roll off as quickly as the monaural curves, and seem to reach a plateau at about 8 dB, even with f_m much greater than the carrier bandwidth. This indicates much broader tuning in the binaural domain than in the monaural domain, as also suggested by Stellmack *et al.* (2005). In contrast, Grantham and Bacon (1991) argued against a bandpass ILD modulation tuning after measuring detection thresholds with a 16-Hz AM_π signal in the presence of a diotic noise AM masker. At that AM frequency, the data in the present study show no significant differences between the AM_π discrimination thresholds with correlated noise or pure-tone carriers, even though the 30- and 300-Hz-wide carriers have envelope frequency components around 16 Hz. This suggests that the diotic masker in their study probably did not have a significant effect on the AM_π detection threshold. This analysis suggests that the data from Grantham and Bacon (1991) is equivocal on the presence or absence of bandpass ILD modulation tuning.

The qualitative similarities between the binaural and monaural masking curves suggest that an element could be introduced in a binaural model that is similar to the monaural modulation filterbank from Dau *et al.* (1997a). However, it appears that the tuning of the ILD modulation filters in such a model must be broader than those of the monaural filterbank.

IV. EXPERIMENT II: MASKED MODULATION DETECTION

The results of the first experiment suggested that there may be modulation frequency selectivity in the processing of ILD fluctuations. Therefore, further experiments were performed to directly measure the shape of this tuning. These experiments were based on similar experiments performed with diotic signals by Ewert *et al.* (2002), where a sinusoidal signal AM was masked by a narrow-band noise modulator applied to a common pure-tone carrier.

A. Specific stimulus details

An interaurally uncorrelated, bandpass Gaussian-noise masker modulation was applied to the envelope of pure-tone carriers in a discrimination task, according to

$$x_L(t) = a\,\sin(2\pi f_c t)[1 + N_L(t)][1 + m\,\sin(2\pi f_m t + \phi_L)]\ ,$$

$$x_R(t) = a\,\sin(2\pi f_c t)[1 + N_R(t)][1 + m\,\sin(2\pi f_m t + \phi_R)]\ , \tag{5}$$

where a controls the presentation level, f_c is the carrier frequency (in this case, 5 kHz), the subscripts L and R indicate left or right ear, and $N_{L/R}$ is the masking noise modulator (power set in this study to −10 dB re 1), spectrally centered at f_N, for the respective ears. The signal modulation was applied with AM frequency f_m, modulation depth m, and starting phases ϕ_L and ϕ_R. When two amplitude modulators are to be applied to a carrier (e.g., a masker, N, and a signal modulator, S), they can be added together and applied as a common modulator $(1+S+N)$ or applied in series as separate modulators $(1+S)(1+N)$. The additive approach can result in overmodulation if either the signal or the masker has a large negative amplitude (i.e., $S+N<-1$). The multiplicative approach, used in this study, avoids overmodulation as long as $S>-1$ and $N>-1$, which allows for signal modulation depths (m) close to 0 dB (see also Houtgast, 1989). However, by multiplying the two modulators, additional spectral sidebands are created, which can complicate analysis of the data, as discussed in Sec. IV C.

The design of the stimuli was based on Ewert *et al.* (2002). Each stimulus had an overall duration of 600 ms, windowed with 50 ms $\cos^2$ onset and offset ramps. The AM signal was applied to the middle 500 ms of the carrier, gated with 50 ms $\cos^2$ onset and offset ramps, leaving 400 ms with the desired signal AM depth. Measurements were made with f_m=4, 8, and 32 Hz. In order to avoid monaural cues, the experiment was designed as a discrimination task, so all three intervals in a trial (signal and two references) had an applied signal modulation with the same modulation depth in each interval. The start phase of the signal modulation in the left ear ϕ_L was chosen randomly for each trial interval over the range $[0, 2\pi]$ with a uniform probability distribution. In the two reference intervals, the modulation start phase in the right ear was set equal to ϕ_L (AM_0), while in the signal interval, ϕ_R was set equal to $\phi_L+\pi$ (AM_π). With the randomized modulation phase and equal modulation depth on all intervals in a trial, successful discrimination could only be performed by combining information from the two ears, not based on one ear's analysis alone.

The masker modulations had a fixed bandwidth of 1.4, 2.8, and 11.1 Hz for the f_m=4, 8, and 32 Hz, respectively, corresponding to one half-octave centered at f_m. The masker center frequencies, f_N, were at octave steps from f_m over a range from −4 to +4 octaves, but with the additional limitation that f_N could not be below 2 Hz or above 128 Hz. This hard frequency limit was put in place because the envelope of the window function itself could interfere with detection below 2 Hz, and the modulation sidebands could be resolvable above 128 Hz. In Ewert *et al.* (2002), the masker modulation was placed over a range from −2 to +2 octaves with a 2/3 octave step size. The larger range and step size were chosen here in expectation of broader tuning after the results from Experiment 1 (see Sec. III). A new masker modulator $(N_{L/R})$ was created for each presentation interval by generat-

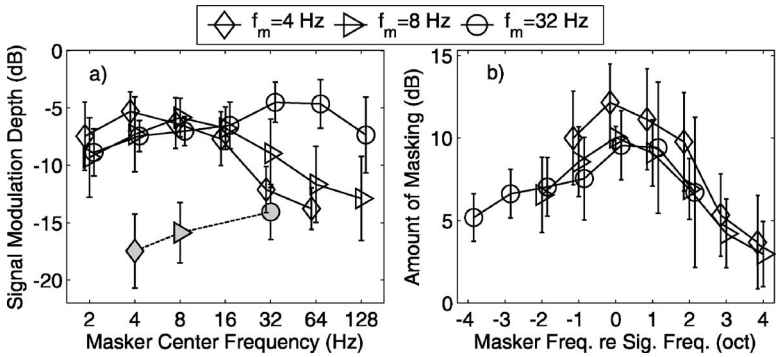


FIG. 4. (a) Modulation depths required for discrimination of interaurally antiphasic AM from homophasic AM imposed on a pure-tone carrier in unmasked (shaded symbols, dashed line) and masked (open symbols, solid lines) conditions. Measurements were made with a fixed signal AM frequency of 4 Hz (diamonds), 8 Hz (triangles), and 32 Hz (squares) with interaurally uncorrelated narrow-band noise maskers with a fixed power and bandwidth for a range of masker center frequencies. The data points for each curve are offset to more clearly show the error bars. (b) The same data from the left panel, but normalized for the unmasked threshold and signal frequency. The error bars in (b) show the standard deviation of the mean threshold difference across listeners.

ing a 10 s Gaussian white noise in the time domain, setting all frequency components outside the passband to zero, and then scaling the variance to 0.1 (−10 dB re 1). The resulting noise was then added to a dc component $(1+N_{L/R})$ and applied to the carrier as in Eq. (5). At this masker level, there was a small probability (less than 0.08% of samples, or less than 0.5 ms per presentation, on average) of overmodulation (i.e., $1+N_{L/R}<0$). This small occurrence was assumed to not have a significant effect on the results.

B. Results

Figure 4(a) shows the mean and standard deviation of the masked threshold patterns measured with a pure-tone carrier. The signal modulation depth in decibels $(20 \log m)$ is plotted as a function of the masker center frequency, with the signal modulation frequency as the parameter. In addition, the modulation depth required for discrimination without a masker present is plotted as a function of the signal modulation frequency (dashed line, shaded symbols). Note that the three masked curves and their respective unmasked points have been offset slightly around the exact frequencies so that the error bars are more visible. In Fig. 4(b), the same curves are replotted as an amount of masking, defined as the difference between the masked and unmasked thresholds for each signal modulation frequency at each masker center frequency, normalized by the signal modulation frequency in octaves. The error bars in Fig. 4(b) show the standard deviation of the mean threshold difference across listeners. In both panels, the symbols (diamond, triangle, and circle) represent the 4, 8, and 32 Hz signal modulation frequencies, respectively.

The masking patterns [Fig. 4(b)] for the three signal frequencies are very similar in shape and amount of masking. All three curves show the highest amount of masking (approximately 10 dB) for the on-frequency condition and a monotonic decrease in masking as the spectral distance between signal and masker center frequency increases. The decrease in masking is greater when the masker center frequency is above the signal frequency than with lower masker frequencies. When the masker is 4 oct. above the signal, there is only about 3 dB of masking, but there is still about 6 dB of masking with the masker 4 oct. below the signal.

C. Discussion

The masking patterns obtained in the $AM_{\pi}-AM_0$ discrimination task with a narrow-band noise modulator masker imposed in series with a sinusoidal signal modulator on a pure-tone carrier showed consistency in shape and amount of masking for the three measured signal AM frequencies. The mean values of the three masking curves at each relative masker frequency are replotted in Fig. 5 (circles) along with a typical masking curve (squares) from the monaural masked AM detection experiments of Ewert *et al.* (2002) (adapted from their Fig. 2; 5.5 kHz carrier, 64 Hz AM signal, $Q=1.25$). The two curves show a maximum amount of masking when the masker is centered at the signal frequency, although the monaural curve shows a clearly higher masking value (about 17 dB) than the binaural curve. For masker frequencies above the signal frequency, the monaural curve rolls off more rapidly than the binaural curve, so that the monaural curve already shows less masking than the binaural curve for maskers centered 1 oct. above the signal modula-

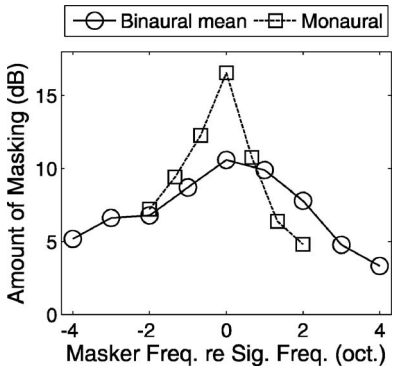


FIG. 5. The mean of the three masking curves from panel Fig. 4(b) (circles, solid line) is shown with a typical monaural AM masking curve [dashed line, squares, adapted from Ewert *et al.* (2002), from their Fig. 2, 5.5 kHz carrier, 64 Hz AM signal].

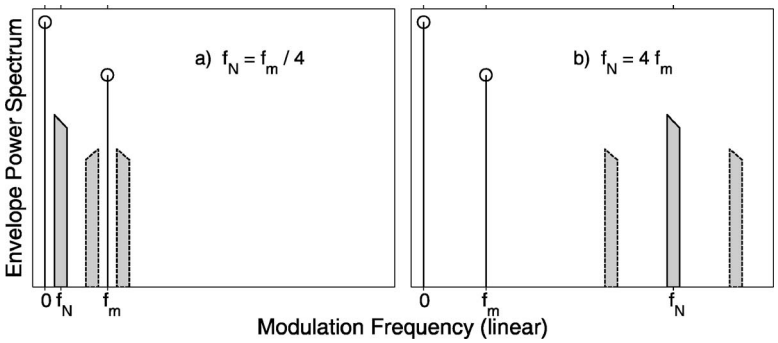


FIG. 6. Theoretical envelope power spectra resulting from the application of a band-pass noise masker modulator and a tonal signal modulator. The dc component ($f=0$) and tonal component ($f=f_m$) are plotted with circles and the spectrum of the noise masker is shown with a bar with a solid line centered at f_N. The bars shown with the dashed lines show the result of applying the masker and signal modulators in series with the multiplicative approach described in the text. The left panel shows a case where $f_N<f_m$ and the right panel shows a case where $f_N>f_m$. Note that only the dc component and positive frequencies are shown.

tion frequency. This is consistent with the idea that monaural envelope processing has a sharper tuning than binaural processing of dynamic ILDs.

Multiplication of the signal and masker modulators creates additional sidebands in the stimulus through spectral convolution. This is represented in a sketch of the envelope spectra of two idealized stimuli in Fig. 6. A stimulus with only an applied noise masker AM would show an envelope spectrum with a dc component ($f=0$) and a band of noise centered at f_N. Multiplication of the masking modulator with the signal modulator (tonal component at f_m) results in the two sidebands shown with dashed lines in Fig. 6 centered at $f_m \pm f_N$ (note that only positive frequencies are shown in the sketch). In an AM *detection* experiment, as in the monaural experiments from Houtgast (1989) and Ewert *et al.* (2002), where the listener's task is to distinguish between presentation intervals with only a masker modulator and a target interval with masker and signal modulators, the sidebands are only present in the target interval. Therefore, they can serve to enhance the detectability of the signal AM. However, with an AM *discrimination* experiment, like the one here, all stimuli have the same modulation and the same sidebands. In this case, the sidebands do not provide any cues for signal detection, and may actually hamper signal detection.

The amplitude of the sidebands is determined by the amplitudes of the masker and signal AM components. In this experiment, with a fixed masker energy, the sidebands' energy scales with the signal energy at a fixed ratio (−10 dB). The effect of these sidebands should be considered when designing a model to account for the measured masking patterns. For example, a model could be designed with a symmetric bandpass modulation filter centered at the signal's modulation frequency, and a certain signal-to-noise ratio (SNR) required after the filter for detection of the signal modulation. With this model, the sidebands' energy would create a noise floor at a SNR that depends on the signal and masker frequencies. When $f_N \ll f_m$, the sidebands will be very close, spectrally, to the signal [Fig. 6(a)], and the sidebands' energy will be passed through the filter with little attenuation, creating a relatively high noise floor. As f_N increases, the sidebands move away from the signal in frequency, becoming more attenuated by the filter and reducing the noise floor. The sidebands are only centered at frequencies larger than f_m if $f_N > 2f_m$. The result could be an asymmetric masking pattern, even though the filter was assumed to be symmetric around the signal modulation frequency.

V. IMPLICATIONS FOR BINAURAL MODELS

The above-presented experimental data suggest that a binaural model should include an array of ILD modulation bandpass filters to simulate human performance. Some preliminary simulations were made using the binaural model from Breebaart *et al.* (2001a) as an artificial observer in the experiments described in Sec. IV. These simulations were performed with the original model, which uses a sliding integrator (low-pass filter) to limit its temporal resolution. The Breebaart model was designed for static binaural conditions, such as for predicting binaural masking level differences (BMLD), and is quite successful at predicting human performance under many experimental conditions (see also Breebaart *et al.*, 2001b,c, for more details). Breebaart *et al.* (2001c) focused on temporal parameters, including a simulation based on an experiment from Grantham (1984), where the listener's task was to discriminate between interaurally antiphasic and homophasic AM imposed on uncorrelated broadband noise carriers. Grantham's data showed a large variance between test subjects, but Breebaart's model was able to capture the general trend of the results. Since their model was able to simulate experimental results similar to those described earlier in Sec. III, it was chosen as a basis for testing with the new experiments and for possible future development.

The model starts with two parallel peripheral processing stages (one for each ear, see schematic in Fig. 7), based on the monaural processing model from Dau *et al.* (1996), which did not include a modulation filterbank. The first stages are an outer- and middle-ear transfer function, basilar-membrane filtering, consisting of an array of gammatone filters, inner hair cell transduction, modeled with half-wave rectification and a low-pass filter with a cutoff frequency of 770 Hz, and finally a series of five adaptation loops, which enable the simulation of forward masking. The output from each pair (right/left) of peripheral channels is then passed to an array of excitation-inhibition (EI) elements, which calculate the difference in the corresponding channels for a range of characteristic interaural gains and delays. This is similar in concept to the equalization-cancellation (EC) model from Durlach (1963), which finds the optimal gain and delay before calculating the channel difference. The EI concept is based on neurons that receive excitatory input from the ipsilateral side and inhibitory input from the contralateral side, effectively calculating a difference between the two auditory signals. The output from each EI element is then smoothed

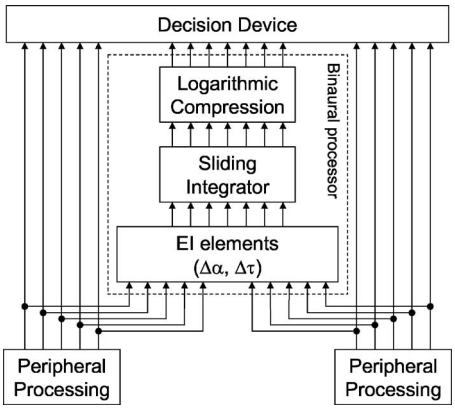


FIG. 7. Schematic of the binaural model from Breebaart *et al.* (2001a). The model consists of two parallel monaural peripheral channels, including a gammatone filterbank, a half-wave rectification and low-pass filter inner hair cell model, and adaptation loops. The two monaural signals are combined in the binaural processor through an array of excitation-inhibition (EI) elements, which calculate the difference of the two signals for a range of applied interaural gains and delays. The resulting signals are smoothed with a sliding integrator and compressed with a logarithmic compression. Finally, an optimal detector tries to find a signal based on all monaural and binaural inputs.

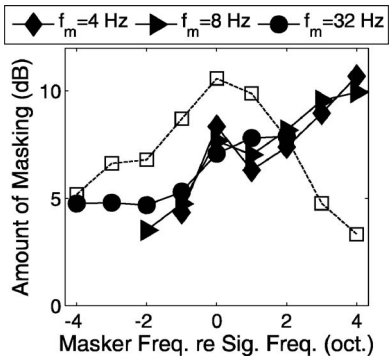


FIG. 8. Masked tuning curves predicted by the model from Breebaart *et al.* (2001a) when used as an artificial listener in the experiments described in Sec. IV. The mean tuning curves from the human listeners are shown with the dashed line and open squares [mean of the three curves from Fig. 4(b)]. The simulation was made with f_m=4 Hz (diamonds), 8 Hz (triangles), and 32 Hz (circles).

with a sliding integrator, consisting of a symmetric double-sided exponential window with time constants of 30 ms. This sliding integrator acts as a low-pass filter with a cutoff frequency of about 5.3 Hz. A compressive (logarithmic) nonlinearity is applied to the smoothed signal. The resolution of the system is limited by the addition of an internal noise. Finally, an optimal detector, with inputs from all monaural and binaural channels, is used as the decision device.

The model does not track perceived motion or predict spatial perception of the sound source, but rather looks at the energy in each EI channel (i.e., for a fixed ILD and ITD combination) in order to detect a signal. Diotic signals have no energy in the ILD=0, ITD=0 channel (perfect cancellation, internal noise is added later) and any interaural decorrelation will result in an increase in energy in this channel. Therefore, the addition of an antiphasic tone to a diotic noise (N_0S_π) will result in a much larger increase in energy than the addition of a homophasic tone (N_0S_0), demonstrating the classic BMLD (see e.g., Licklider, 1948; Hirsh, 1948).

The experimental conditions described in Sec. IV were simulated using the model. The simulation results are summarized as masking curves in Fig. 8, like those shown in Fig. 4(b). It is clear from a comparison of the human listeners' (open symbols) and the model's results (closed symbols) that the tuning described in Sec. IV is not captured by the model. The model does show a small peak at the signal frequency, but then the masking level increases with higher relative masker center frequencies, while the human listeners show a decrease in masking level with higher masker frequencies. The reason for the increase in masking in the model with high masker center frequencies is that the sliding integrator smooths out the interaural fluctuations from the masker, thereby removing any locations with good cancellation and increasing the energy at the output of the EI channels. Adding the diotic AM to the reference intervals further increases the energy so that there is less of a difference between the signal and reference intervals, making the discrimination task harder to perform. Part of this effect stems from the fact that the model was not designed to look at temporal differences and only compares the total energy at the output of the EI channels.

In order for this binaural model to be able to predict thresholds with fluctuating stimuli, it requires frequency selectivity in the processing of monaural and interaural level fluctuations. The monaural modulation filterbank (MFB) from Dau *et al.* (1997a) could be added to the peripheral channels, but then a question arises as to the sequence of model stages: Should the taps for the EI array come from before or after this filterbank? Two possible design concepts are shown in Fig. 9. The sequence of the stages would not be important except for the nonlinearities in both the monaural MFB and the EI process. The monaural MFB has a nonlinear reduction of modulation phase information for frequencies above 10 Hz. Without the modulation phase information, there would be no interaural differences with an AM_π signal, and the model would not be able to discriminate between AM_π and AM_0. If the EI inputs were to come from after the monaural modulation filters, but before this nonlinearity (left panel of Fig. 9), then the sharpness of tuning would be preserved through the output of the EI elements. That sharpness might be reduced to fit the measured data by adding additional noise and/or interaural differences in the processing, but the effect of this additional noise on other experiments would have to be investigated. Another option would be to take the EI inputs from before the monaural modulation filterbank (right panel of Fig. 9). In this manner, the interaural modulation phase differences would be preserved going into the binaural processor. However, a new model stage would then be required, namely a ILD modulation filterbank at the output of the EI system in addition to the two monaural amplitude modulation filterbanks. This filterbank would replace the sliding integrator from the original Breebaart model. The optimal sequence for the linear and nonlinear model stages should be investigated in further simulations.

VI. SUMMARY AND CONCLUSIONS

The first experiment showed that interaurally correlated and uncorrelated narrow-band noise carriers have a signifi-

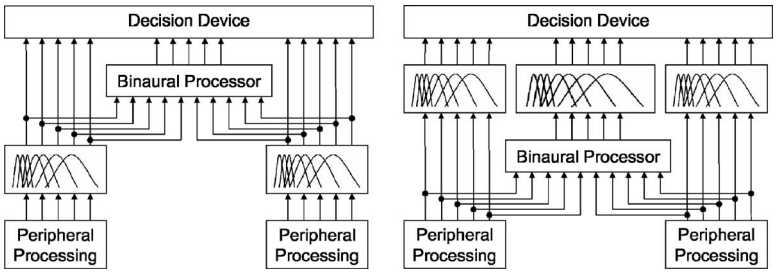


FIG. 9. Possible concepts for the inclusion of modulation frequency selectivity in binaural processing of ILD fluctuations. In the left panel, the modulation filterbank (MFB) from Dau *et al.* (1997a) is applied to each monaural peripheral channel before the input to the binaural processor. The binaural system would thereby inherit its ILD modulation frequency tuning from the monaural AM processing. The second concept, in the right panel, takes the inputs to the binaural processor from before the monaural MFBs. This then requires an additional binaural ILD modulation filterbank to add frequency selectivity in the processing of interaural level fluctuations.

cant effect on the discriminability of modulated ILDs (AM_π) from diotic AM (AM_0), particularly for modulation frequencies below the bandwidth of the carrier. This suggested that the binaural system shows broad bandpass modulation frequency tuning in processing of ILD fluctuations. A comparison of the results obtained with diotically modulated and unmodulated references underscored the importance of eliminating monaural cues in the design of binaural detection tasks because the signal detection will be based on monaural detection if the monaural cues are more salient than the binaural cues.

This modulation frequency tuning was further explored in the second experiment with AM_π discrimination in the presence of masking narrow-band noise modulators. The masking patterns also showed bandpass tuning, but with a broader tuning than that shown in similar monaural experiments (e.g., Ewert *et al.*, 2002).

An analysis with an existing binaural model (from Breebaart *et al.*, 2001a) showed that the model, which uses a low-pass filter to limit its temporal resolution in the processing of fluctuating interaural differences instead of a modulation filterbank, cannot predict the thresholds or the masking patterns measured with human listeners.

Further experiments should be performed to investigate the effect of diotic level fluctuations on the perception of ILD fluctuations through additional psychoacoustic tests as well as modeling. In addition, a binaural model should be developed that can predict the frequency selectivity shown here in the processing of interaural level fluctuations.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Richard Freyman, the associate editor, Dr. Wes Grantham, and three anonymous reviewers for their very helpful comments and suggestions on earlier revisions of the manuscript. This research was supported by a Ph.D. scholarship from the Technical University of Denmark.

[1]Low-noise is a bandpass noise for which the phase angles of the individual frequency components have been optimally selected to minimize the fourth moment of the signal, thereby minimizing the envelope fluctuations. Kohlrausch *et al.* (1997) produced low-noise noise using an iterative process of bandpass filtering the noise and normalizing the noise with its envelope. Each filtering step produces envelope fluctuations, and each normalization step produces a flat envelope, but a broader frequency spectrum. After many iterations, a noise is produced which has both a flat envelope and the desired bandwidth. The control experiment was performed with one listener, a 30-Hz-wide carrier, and a 4 Hz AM_π signal.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.
Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise," J. Acoust. Soc. Am. **95**, 3561–3567.
Blauert, J. (**1972**). "On the lag of lateralization caused by interaural time and intensity differences," Audiology **11**, 265–270.
Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.
Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1104.
Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001c**). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.
Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1991**). "Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays," J. Acoust. Soc. Am. **90**, 3077–3085.
Dau, T. (**1996**). "Modeling auditory processing of amplitude modulation," Ph.D. thesis, Universität Oldenburg, Oldenburg, Germany.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.
Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.
Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.
Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). "Spectro-temporal processing in the envelope-frequency domain," J. Acoust. Soc. Am. **112**, 2921–2931.
Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988–991.
Fleischer, H. (**1982**). "Modulationsschwellen von Schmalbandrauschen [Modulation thresholds of narrow-band noise]," Acustica **51**, 154–161.
Grantham, D. W. (**1982**). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," J. Acoust. Soc. Am. **72**, 1178–1184.
Grantham, D. W. (**1984**). "Discrimination of dynamic interaural intensity differences," J. Acoust. Soc. Am. **76**, 71–76.
Grantham, D. W., and Bacon, S. P. (**1991**). "Binaural modulation masking," J. Acoust. Soc. Am. **89**, 1340–1349.
Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.
Griesinger, D. (**1997**). "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," Acust. Acta Acust.

83, 721–731.
Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex wave-forms," J. Acoust. Soc. Am. **55**, 84–90.
Hirsh, I. J. (**1948**). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.
Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.
Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.
Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acust. Acta Acust. **83**, 659–669.
Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals*, Radiation Laboratory Series Vol. **24** (McGraw-Hill, New York).
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Licklider, J. C. R. (**1948**). "The influence of interaural phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. **20**, 150–159.
Mills, A. W. (**1960**). "Lateralization of high-frequency tones," J. Acoust. Soc. Am. **32**, 132–134.
Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.
Price, R. (**1955**). "A note on the envelope and phase-modulated components of narrow-band gaussian noise," IRE Trans. Inf. Theory **1**, 9–13.
Pumplin, J. (**1985**). "Low-noise noise," J. Acoust. Soc. Am. **78**, 100–104.
Schroeder, M. R. (**1981**). "Modulation transfer-functions: Definition and measurement," Acustica **49**, 179–182.
Sheft, S., and Yost, W. A. (**2007**). "Discrimination of starting phase with sinusoidal envelope modulation," J. Acoust. Soc. Am. **121**, EL84–EL89.
Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2005**). "Monaural and interaural temporal modulation transfer functions measured with 5-kHz carriers," J. Acoust. Soc. Am. **118**, 2507–2518.
Terhardt, E. (**1968**). "Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung. [The auditory sensation produced by amplitude modulated tones]," Acustica **20**, 210–214.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Witton, C., Green, G. G., Rees, A., and Henning, G. B. (**2000**). "Monaural and binaural detection of sinusoidal phase modulation of a 500-Hz tone," J. Acoust. Soc. Am. **108**, 1826–1833.

Localization cues with bilateral cochlear implants

Bernhard U. Seeber[a] and Hugo Fastl
Institute for Human-Machine-Communication, Technische Universität München Arcisstr. 21, 80333 München, Germany

(Received 10 April 2007; accepted 11 November 2007)

Selected subjects with bilateral cochlear implants (CIs) showed excellent horizontal localization of wide-band sounds in previous studies. The current study investigated localization cues used by two bilateral CI subjects with outstanding localization ability. The first experiment studied localization for sounds of different spectral and temporal composition in the free field. Localization of wide-band noise was unaffected by envelope pulsation, suggesting that envelope-interaural time difference (ITD) cues contributed little. Low-pass noise was not localizable for one subject and localization depended on the cutoff frequency for the other which suggests that ITDs played only a limited role. High-pass noise with slow envelope changes could be localized, in line with contribution of interaural level differences (ILDs). In experiment 2, processors of one subject were raised above the head to void the head shadow. If they were spaced at ear distance, ITDs allowed discrimination of left from right for a pulsed wide-band noise. Good localization was observed with a head-sized cardboard inserted between processors, showing the reliance on ILDs. Experiment 3 investigated localization in virtual space with manipulated ILDs and ITDs. Localization shifted predominantly for offsets in ILDs, even for pulsed high-pass noise. This confirms that envelope ITDs contributed little and that localization with bilateral CIs was dominated by ILDs.
 [DOI: 10.1121/1.2821965]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.66.Sr, 43.66.Ts [RYL] Pages: 1030–1042

I. INTRODUCTION

Localization of sounds is one of the most important tasks for the auditory system as it not only helps orientation in space but it is also crucial for the segregation of multiple sounds in the auditory scene. Normal hearing subjects show an outstanding ability to localize sounds and directional changes of 1° can be detected in the front (Mills, 1958; Blauert, 1997). Several recent studies demonstrated that cochlear implant (CI) subjects can regain the ability to localize sounds after bilateral implantation (Tyler *et al.*, 2002; van Hoesel *et al.*, 2002; Grantham *et al.*, 2003; van Hoesel and Tyler, 2003; Laszig *et al.*, 2004; Nopp *et al.*, 2004; Schoen *et al.*, 2005). One subject in a study by van Hoesel (2004) made no errors discriminating between the two frontal loudspeakers at ±13° in an array spanning 180° using a clinical stimulation strategy and a pink noise stimulus. However, root mean square-error increased for non-frontal directions to 10–35°. Seeber *et al.* (2004) studied localization of bilateral CI subjects as well as of subjects with hearing aid and cochlear implant using a high-resolution pointing technique. All four bilateral CI subjects were able to localize but one subject showed excellent localization ability with quartiles of only 4.4° and a regression slope of 1.15. This localization ability is close to one of normal hearing subjects who showed quartiles of 1.7° and a slope of 0.95 in the same task (Seeber, 2002). The purpose of the present article is thus to explain this surprisingly good localization ability by studying the localization cues used by this and another subject.

Several recent studies investigated the sensitivity of CI subjects to binaural cues. All studies correspondingly showed high sensitivity to interaural level differences (ILDs). Lawson *et al.* (1998), (2000), and van Hoesel and Tyler (2003) found sensitivities as small as one current step. Sensitivity rarely exceeded a few current steps across several subjects and electrodes. Despite the dynamic range compression in CI processors this translates into high sensitivity to ILDs at the acoustical input. Van Hoesel (2004) measured just-noticeable differences (JNDs) for ILDs at the input of a research processor. Both of his subjects showed ILD-JNDs smaller than 1 dB. Likewise, Laback *et al.* (2004) measured ILD-JNDs of two subjects at the direct input of clinical processors with disabled automatic gain controls (AGCs). They found ILD-JNDs of 1.4–2.7 dB (S1) and 1–5.2 dB (S2) for various stimuli which were about 1 dB larger than for normal hearing subjects.

A larger range of sensitivities was found for interaural time differences (ITDs) in several recent studies. Several subjects demonstrated JNDs beyond 1 ms for ITDs in the carrier, which is outside the naturally occurring range (van Hoesel *et al.*, 1993; van Hoesel and Clark, 1997; Lawson *et al.*, 2000; van Hoesel, 2004). Selected subjects achieved higher sensitivities at certain electrode combinations. Lawson *et al.* (1998) found ITD-JNDs of 150 μs in one subject with synchronized processors for pulses at a rate of 480 pulses per second (pps). Van Hoesel and Tayler (2003) showed JNDs of 90, 150, and 250 μs each for two subjects for carrier ITDs in pulses at 50 pps. The subjects studied by Lawson *et al.* (2000) showed mixed results, but several subjects demonstrated ITD-JNDs of 50–150 μs on at least one electrode combination. The bilateral CI-subject BW, who

[a]Author to whom correspondence should be addressed. Present address: MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, United Kingdom; electronic mail: seeber@ihr.mrc.ac.uk.

demonstrated excellent localization ability in our previous study (Seeber *et al.*, 2004), participated in their study (ME5) and showed ITD-JNDs of 50 μs on one and 150 μs on two electrode combinations. One subject in the study by Lawson *et al.* (2000) (ME8) showed ITD-JNDs of 50 μs on 5 electrode combinations and JNDs of 150 μs on 16 other tested combinations which were by far the best and most consistent results of any subject in the test. Because of his high sensitivity to ITDs this subject was selected to participate in the current study (DF).

While previous studies investigated localization with bilateral CIs, the sensitivity to binaural cues, or lateralization with isolated binaural cues the present article focuses on the combination of binaural cues for *localization*. If available, the auditory system utilizes information from multiple cues, but cue weighting depends on the spectral content and the temporal structure of the sound. ITDs are present in the envelope and the carrier of the sound while ILDs are physically larger only at higher frequencies. Normal hearing subjects can evaluate ITDs in the carrier up to 1500 Hz and the threshold for ITDs can be as low as 10 μs (Klumpp and Eady, 1956; Buell and Hafter, 1991). At higher frequencies ITDs can be well detected in the envelope and thresholds of a few 10 μs were reported. (Henning, 1974; van de Par and Kohlrausch, 1997). ILD detection is relatively independent of frequency and thresholds can be as low as 0.5 dB in normal hearing (Yost, 1981). However, the contribution of ILDs to localization is limited to their range of natural occurrence. Physical ILDs do not exceed 5 dB below 1 kHz, but they can be as large as 30 dB at high frequencies (Shaw, 1974). Strutt (Lord Rayleigh) (1907) postulated in the "Duplex Theory" that ITDs dominate ILDs below about 1500 Hz while ILDs provide the information for localization at higher frequencies. This general trend has been confirmed, e.g., in trading experiments for tones (Hafter and Jeffress, 1968). The perceived direction of wide-band sounds is mostly determined from ITDs at low frequencies, i.e., carrier ITDs (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002), while the contribution of envelope ITDs seems to be limited to sound onsets (Hafter, 1996; Freyman and Zurek, 1997; Hartmann *et al.*, 2005). Envelope ITDs might also help resolve ambiguities (Buell and Hafter, 1991). However, the dominance of carrier ITDs for localization can be broken if ILDs are consistent with envelope ITDs (Smith *et al.*, 2002; Zeng *et al.*, 2004). This is important for localization with CIs since current processors do not encode ITDs in the carrier, but they do transmit ILDs and envelope ITDs. The role of envelope ITDs in CI listening might therefore be much higher than in normal hearing. Some CI subjects were shown to be highly sensitive to envelope ITDs and thresholds down to 25 μs have been reported (Lawson *et al.*, 2001), while other reports showed thresholds of about 290 μs (van Hoesel and Tyler, 2003) or 260–380 μs (Laback *et al.*, 2004). Complete lateralization with envelope ITDs of 600–700 μs has been reported for selected subjects (Schoen *et al.*, 2005). It is therefore conceivable that localization of CI subjects is governed by ILDs and envelope ITDs.

The purpose of the present study was to investigate localization cues of two selected CI subjects with three different approaches. Unlike previous studies, this study evaluated the contribution of binaural cues to *localization* rather than pure sensitivity to binaural cues. Both subjects were selected for their outstanding localization ability and subject DF was chosen for his high sensitivity to ITDs. In experiment 1, localization cues of both subjects were studied through a localization test in which several sounds of different spectral content and temporal envelope structure were presented. Experiment 2 assessed how subject DF utilized cues for localization by modifying the placement of the CI processors away from their normal position on the ears. By placing the processors above the head, ITDs and ILDs could be physically altered as no head shadow was effective. In experiment 3, ITDs and ILDs were manipulated in virtual acoustical space. Head-related transfer functions (HRTFs) of the CI processors were altered to slightly offset ITDs or ILDs from their natural combination. A shift in localization correlated to the offset would indicate sensitivity to the altered cue (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). Obviously, this test can be administered only to subjects with good localization ability, as previously demonstrated by subject BW (Seeber *et al.*, 2004). Localization was tested with an accurate, continuous method which requires subjects to point to the perceived azimuth of the sound with an adjustable light pointer (Seeber, 2002).

II. METHODS

A. Subjects

Two bilateral cochlear implant subjects participated in the experiments. Subject BW was singled out because of his extraordinary localization ability in a previous study (Seeber *et al.*, 2004). Subject BW was 50 years old and he received his first implant on the right ear 4.2 years and his second implant 3.2 years before the study. Both implants were Med-El Combi 40+ (C40+) controlled by Med-El Tempo+ speech processors. Both processors were set up with a logarithmic map in 300–5547 Hz for 12 channels. BW deafened on Otosclerosis 1.1 years before he received his first implant. Travel costs were reimbursed at government mileage rates.

Subject DF was 54 years old and he received his first implant 7.8 years and his second implant 3.1 years before the tests. He deafened on Meningitis and Endolymphatic Hydropress. The first implant on the left side was a Med-El Combi 40 (C40) while the right side received a Med-El C40+ later, both controlled by Med-EL Tempo+ speech processors. All 8 channels of the C40 (left) were mapped logarithmically to cover an extended range of 200–8500 Hz. Channel 12 was not assigned on the C40+ (right). The remaining 11 channels were set to cover the same frequency range as the implant on the left ear according to logarithmic spacing. This implies that individual channels analyzed different frequency ranges on both ears. The subject preferred this map as it allowed him better music perception. DF was reimbursed for his travel, and hotel accommodation was arranged for him. BW and DF did not receive payment for the experiments.

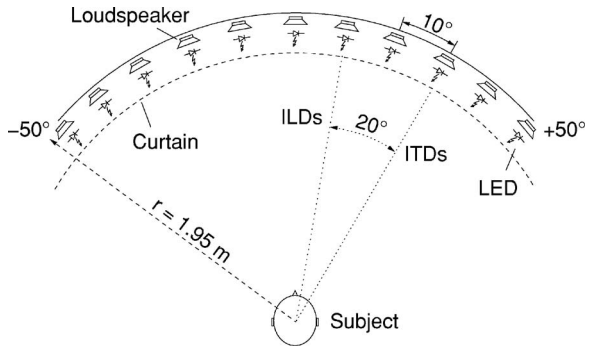


FIG. 1. Sketch of the apparatus for localization studies and exemplification of binaural cue manipulation in experiment 3. In experiments 1 and 2 sounds were played via loudspeakers while in experiment 3 spatialized stimuli were fed directly into the CI processors. The figure shows an example for binaural cue trading used in experiment 3: While the prerecorded stimulus from, e.g., +10°, was played to the subject it was run through an all-pass filter that incorporated the ITD changes necessary to represent ITDs from +30°. A shift in localization to the ITD direction would suggest a dominance of ITDs whereas a localized direction halfway between the ITD and ILD direction shows equal weighting of both cues. Likewise, conditions with manipulated ILDs were run.

B. Cochlear implants

The Combi 40+ implant by MedEl GmbH consists of an intracochlear array of 12 active electrodes in 2.4 mm spacing while the Combi 40 implant uses 8 active electrodes in 2.8 mm spacing. The stimulation occurred in monopolar mode against an extracochlear reference electrode placed on the skull beneath M. Temporalis. The Tempo+ speech processor delivered a continuous interleaved sampling strategy at a rate of 1515 pps on each of the 12 channels for subject BW. DF used a rate of 1456 pps on the C40 (left) and 1583 pps on the C40+ (right). Behind-the-ear (BTE) type speech processors were used in the standard everyday configuration for the subjects.

C. Localization test

The experiments were done in a darkened anechoic chamber which hosted an apparatus of 11 identical closed-cabinet loudspeakers (Fig. 1). The loudspeakers were placed at ear level of the subject on a horizontal arc with a radius of 1.95 m. They spanned an angle of −50° left to +50° right with a spacing of 10°. Loudspeakers were individually equalized to be frequency independent within 125 Hz–20 kHz to ±2.5 dB at the subject's head position. A detailed description of the apparatus can be found in Seeber (2002).

Localization was tested with a light pointer method according to which the subject adjusts a movable light spot to the perceived direction of the sound with a trackball (Seeber, 2002). The light spot was projected on a cylindrical, acoustically transparent curtain that covered the loudspeakers from view. The projection was done by deflecting a laser beam with a computer-controlled two-mirror system. The subject could move the light spot on a horizontal arc by turning on a trackball within −70° left to +70° right with an accuracy of 0.2°. Unlike source identification methods, this technique allows for a continuous display of the localized direction. The method shows very small variance (quartiles 1.7° for normal hearing subjects), which permits an accuracy closed to the minimum audible angle obtained in detection tasks (Mills, 1958; Seeber, 2002). The high accuracy comes in part from the fact that no part of the body directly displays the perceived sound direction like in head or hand pointing. The method is thus called the Proprioception Decoupled Pointer.

III. EXPERIMENT 1

A. Stimuli and procedures

The aim of experiment 1 was to investigate the relative contribution of binaural cues by studying localization of specific sounds played via loudspeakers in the free field. Each sound could be localized on the basis of a particular set of cues. Sound characteristics and localization cues are listed in Table I. Low-pass noise (LPN) restricts the availability of ILDs which are physically small at low frequencies. It could be localized predominantly on the basis of carrier or envelope ITDs. High pass noise (HPN) could be localized using ILDs which are large at high frequencies (Fig. 9). Envelope ITDs introduced either by envelope modulation of the signal or by the bandpass filtering in the CI processor could also contribute to localization. For HPN of larger bandwidth spectral cues might be utilized. Access to spectral cues can be limited by scrambling the spectrum of the signal such that the level in each frequency channel is randomly chosen and thus stands in no relation to sound direction (Wightman and Kistler, 1997). Spectral cues could be particularly useful by comparing the direction dependent levels in a high-frequency channel with the relatively constant level in a low-frequency channel. This low-frequency anchor is restrained with high-pass sounds. Pulsation of the noise with steep slopes emphasizes envelope ITDs, and slow envelope changes in the order of 200 ms evoke the opposite. Test sounds were individualized according to the channel mapping of the subject.

Prior to all experiments the sensitivities of the speech processors were adjusted to give a centered image for speech coming from the front. After a short training session for accommodation with the method, localization was tested in separate runs for each test sound. Localization with bilateral CIs was studied for all sounds and each single CI was tested with pulsed wide-band noise (WBN). A localization experiment for a single sound consisted of ten trials for each of the 11 sound directions (110 trials total) which took approximately 11 min to complete. Presentation was divided into ten blocks, with the eleven directions presented in random order within each block. Sound level was roved in random order in 2 dB steps in 61–69 dB sound pressure (SPL) with each of the five level steps administered twice in the ten trials per sound direction. Ten separate frozen noise tokens were generated and used for the ten trials. No feedback was provided during the experiment, but the training session consisting of 1–3 presentations of each test direction included feedback.

In a single localization trial the test sound was played followed by a pause of 500 ms after which the light spot appeared at 0° in front of the subject. The subject then moved the light spot to the remembered direction of the sound and confirmed this direction by pressing a button at the trackball. The light spot ceased and after a pause of

TABLE I. Overview of stimuli.

Stimulus[a]	Bandwidth[b] [Hz]	Envelope[b]	Spectrum	Duration[b] [ms]	Available cues[c] ILDs	Carrier-ITDs	Envelope-ITDs	Spectrum
WBN, pulsed	125-20000	5 pulses of 30 ms, 70 ms pause, 3 ms slopes,	Gaussian noise	500	++	++	++	++
WBN-CI, 200 ms env	BW: 300–5547; DF: 200–8500	200 ms slopes	Gaussian noise	500	++	++	o	++
LPN, 200 ms env.	BW: 300–486; DF: 200–525	200 ms slopes	Gaussian noise	500	o	++	–	–
LPN, pulsed	BW: 300–486; DF: 200–525	10 pulses of 10 ms, 40 ms pause, 1 ms slopes	Scrambled by up to 40 dB in: BW: [300, 381, 486], DF: [200, 328, 525] Hz	500	o	++	++	––
HPN, scrambled	BW: 2644–5547; DF: 3103–8500	200 ms slopes	Scrambled by up to 40 dB in: BW: [2644, 3377, 4360, 5547], DF: [3103, 4362, 6134, 8500] Hz	500	++	––	o	––
LPN 400 Hz	DF only: 200–400	200 ms slopes	Gaussian noise	500	–	++	–	––
Additionally in experiment 3 (BW only):								
HPN, 100 ms env.	2644–5547	100 ms slopes	Scrambled by up to 40 dB in [2644, 3377, 4360, 5547] Hz	164	++	––	–	––
HPN, pulsed	2644–5547	10 pulses of 10 ms, 40 ms pause, 0.5 ms slopes	Scrambled by up to 40 dB in [2644, 3377, 4360, 5547] Hz	500	++	––	++	––

[a]WBN: Wide-band noise; WBN-CI: Wide-band noise limited to frequencies processed by the CI; LPN: Low-pass noise; HPN: High-pass noise.
[b]Corner frequencies of noises in Hz. All slopes had Gaussian shape. Signal durations computed at 67.5% points.
[c]Estimate of availability of physical localization cues, coded in five steps as –– (unavailable), o (medium), ++ (strong presence). Carrier ITDs are reported to be available as in the duplex theory.

500 ms the next trial started with sound being played from another direction. At the beginning of the experiment and between each of the 10 blocks a light spot appeared for 5 s in front of the subject followed by a pause of 500 ms. This allowed the subject to take a short break and to align the head to the front. The head was fixed by leaning it against a headrest. The experimenter verified through an infrared camera system that the head was not moved during sound presentation parts of the experiment.

B. Results

Localization results for all stimuli of experiment 1 are given in Figs. 2 and 3 for subjects BW and DF, respectively. Median results are plotted along with error bars which depict quartile ranges. Summary statistics of the results are listed in Tables II and III.

BW was able to discriminate the side of sound origin using either the first or the second implanted ear (panel A of Fig. 2) while subject DF could do so only with his first CI (Fig. 3(A)). Subject BW localized sounds coming from the contralateral side of the CI towards that side whereas sounds from the ipsilateral side were localized in the front, independent of their true direction. Despite the strong variance in the results, discrimination of left from right is significant in a test which compared the pooled responses for −50° and −40° to responses for +50° and +40° (Wilcoxon rank sum test, last column in Table II). As BW has no residual hearing on either implanted ear the ability to discriminate left from right has to rely on the evaluation of monaural spectral cues. Despite the overall level rove, timbral cues can be picked up and used for localization as the spectrum of the white noise pulses is predictable. It is interesting to note that BW was not able to discriminate left from right with the second CI at the time of our first test 1.3 years before the current study (Seeber *et al.*, 2004). Although he wore both implants continuously throughout that time he learned the information needed for the task with a single CI.

Localization results of wide-band noise (WBN) with pulsed and with slowly changing envelope using both CIs are shown in panel B of Figs. 2 and 3. These data confirm what we previously showed for pulsed WBN that subject BW has excellent localization ability (Seeber *et al.*, 2004). Compared to his previous data the slope of the regression line is slightly shallower (0.82 vs 1.15) in the current data set, but average quartiles are also smaller (3.3° vs 4.4°). The ratio of quartiles to slope could be seen as a measure of internal sensitivity to localization cues. Since it did not change (4.0 vs 3.8) it can be assumed that localization ability did not change in the 1.3 years between both tests. Based on a similar consideration it is interesting to see that localization ability did not change if envelopes were changed slowly instead of being pulsed. Localization results for WBN with slow envelope changes show a perfect regression slope of 0.97 paired with an offset of 3.5° and very small quartiles of 3.9°; however, with the increase in slope quartiles increased similarly compared to the pulsed WBN. Apparently, envelope ITDs seem not to contribute to localization of wide-band stimuli beyond that due

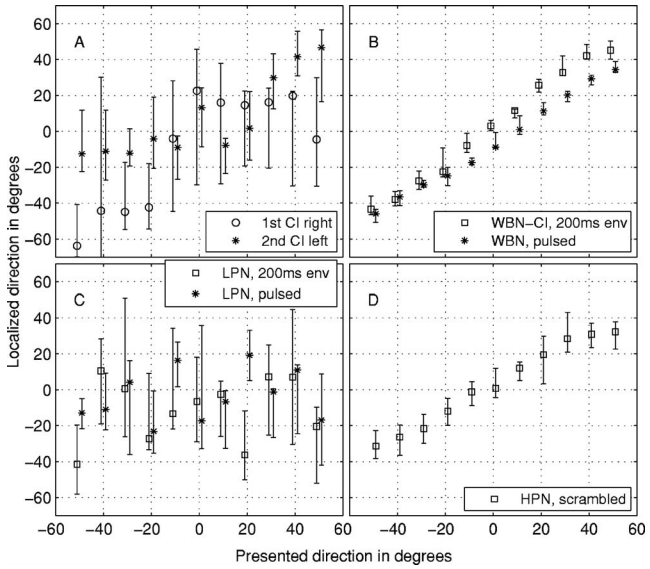


FIG. 2. Free-field localization results of subject BW in experiment 1. Results are presented as medians and quartiles. Panel A shows results for each single CI for pulsed WBN, whereas panels B-D show results for various stimuli with bilateral CIs (see inlets and Table I).

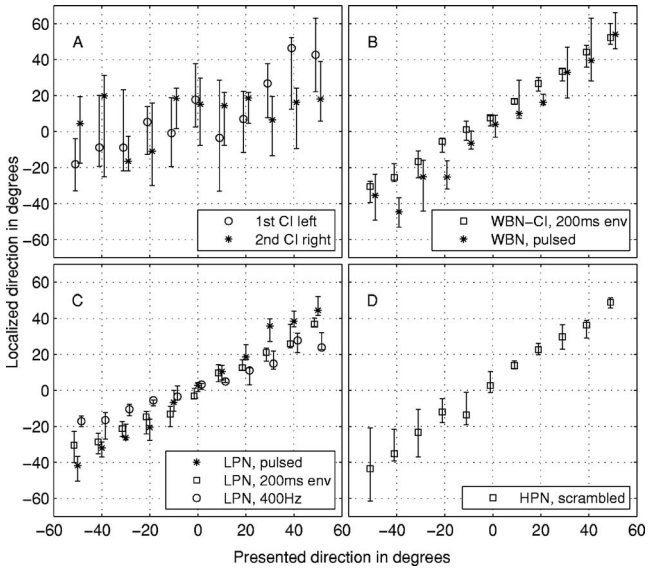


FIG. 3. Free-field localization results of subject DF in experiment 1. Panel C shows additional results for LPN limited to 400 Hz. Results are plotted as in Fig. 2, stimuli are described in Table I.

to other binaural cues. The reduced slope with pulsed WBN might stem from the compression in the CI processor which is assumed to be larger with the pulsed stimulus because of its larger maximum amplitude.

Subject DF shows nearly a similar level of localization performance as both WBN-stimuli can be localized well (Fig. 3(B)). Localization was best for WBN with slow envelope changes. Despite an offset of the regression line (8.5°) the good localization ability is supported by the near perfect slope (0.86) combined with small quartiles (4.5°), which yields high, significant correlation of presented direction to localized direction (0.97). Interestingly, localization ability seems slightly poorer for the pulsed WBN as quartiles increase to 8.0°, although this stimulus provides more binaural information. The increase in variance might be attributable to temporal effects of compression in the CI processor as ILDs might change during the first pulses.

Localization tests with low-pass noise (LPN) were done to investigate the contribution of ITDs at low frequencies at which ILDs are small. Despite his excellent localization ability of wide-band sounds, subject BW showed no localization ability for LPN (Fig. 2(C)). Neither the LPN with slow envelope changes nor the pulsed noise could be localized. The slope of the regression line is zero and quartiles are large (22.4° and 14.3°, respectively). This confirms that ITDs at low frequencies could not be evaluated by BW. Moreover, ITDs in the envelope of this low-frequency carrier also did not contribute to localization, indicating that neither ITDs in the phase at low frequencies nor in the envelope contributed to localization. The results thus suggest that localization of subject BW is exclusively based on the evaluation of ILDs which will be tested in experiment 3.

Localization results of DF for low-pass sounds differed from the results of BW. DF was able to localize the LPN

TABLE II. Summary statistics to free-field localization results of subject BW in experiment 1 for different stimulus conditions, cf. Fig. 2.

	Error (°)[c]		Quartile	Correlation	Regression line		Side
Condition	Absolute[a]	Arithmetic[b]	(°)	coefficient[d]	Slope	Offset	diff[e]
1st CI right	30.0	−17.3	25.7	0.36**	0.48	−8.6	+
2nd CI left	17.7	5.2	15.2	0.50**	0.46	6.0	**
Both CIs							
WBN, pulsed	8.5	−7.8	3.2	0.97**	0.82	−6.5	**
WBN-CI 200 ms env	4.5	2.1	3.9	0.97**	0.97	3.5	**
LPN, 200 ms env	31.1	−9.7	22.4	0.03	0.03	−7.0	
LPN, pulsed	30.8	−3.6	14.3	0.05	0.05	−3.6	
HPN, scrambled	9.9	1.4	7.6	0.91**	0.71	2.3	**

[a]Absolute error: Mean absolute deviation of single localization results from presented direction.
[b]Relative, arithmetical error: Mean deviation of single localization results from presented direction.
[c]Average value of single quartiles.
[d]Significance of correlation coefficient: **0.001, *0.01, +0.05.
[e]Differentiation of side of sound origin: Wilcoxon rank sum test on identity of the results at −50° and −40° (pooled) vs +40° and +50° (pooled). Results for both sides are significantly different at **0.001, *0.01, +0.05.

TABLE III. Summary statistics for localization results of subject DF in free-field experiments 1 and 2, cf. Figs. 3 and 4. Statistics as in Table II.

	Error (°)		Quartile	Correlation	Regression line		Side
Condition	Absolute	Arithmetic	(°)	coefficient	Slope	Offset	diff
1st CI left	23.7	9.9	16.3	0.46**	0.45	10.4	**
2nd CI right	23.1	4.8	13.8	0.16	0.13	7.5	
Both CIs							
WBN, pulsed	8.1	1.8	8.0	0.91**	0.98	3.2	**
WBN-CI 200 ms env	8.0	7.8	4.5	0.97**	0.86	8.5	**
LPN, 200 ms env	8.1	−1.8	4.2	0.95**	0.71	−0.3	**
LPN, pulsed	5.1	2.4	4.6	0.97**	0.92	2.2	**
HPN, scrambled	6.1	2.4	4.8	0.93**	0.90	3.0	**
LPN, 400 Hz	14.5	2.0	3.1	0.92**	0.48	2.5	**
Elevated CIs, Experiment 2							
Together	24.2	0.9	9.8	0.16	0.08	3.2	
Distanced	19.2	3.6	8.2	0.55**	0.39	6.2	**
Plate	10.2	−4.3	7.6	0.87**	0.83	−2.5	**

with slow envelope changes relatively well (Fig. 3, panel C). His responses were compressed towards the front which is indicated by a reduced regression slope of 0.71. Variance was small (quartiles 4.2°). Localization improved considerably if the LPN was pulsed. This is indicated by the ideal regression slope of 0.97 combined with small quartiles of 4.6° and a small offset of 2.2°. The results seem to confirm that localization with CIs can be based on the evaluation of ITDs through the phase or the envelope. The emphasis of envelope ITDs also leads to an improvement of localization, which again suggests a contribution of ITDs to localization, even though this is opposite to that with WBN above. However, an alternate explanation would be that ILDs still contributed enough information at 500 Hz and that the pulsation improved localization because of increased spectral splatter. The combination of pulsation and the relatively shallow upper slope of the noise might result in sufficient energy at high frequencies to be useful for evaluating ILDs.

To address this question DF was tested again using noise with a lower cut-off frequency and slow envelope changes. The lower cut off at 400 Hz still provides considerable energy in the second implant channel while the information in phase and envelope ITDs remains nearly unchanged. Results are given in Fig. 3(C). Localization ability appears considerably reduced compared to the similar stimulus with slightly wider spectrum. The slope of the regression line for "LPN 400 Hz" is only 0.48, but quartiles remain small (3.1°). The fact that localization ability was reduced for this stimulus suggests that ILDs clearly contributed to localization of the other two LPNs. However, it cannot be ruled out that ITDs also contributed, which will be tested further in experiment 2.

Both subjects were able to localize high-pass noise (HPN) with scrambled spectrum (Figs. 2 and 3, panel D). Localization ability deteriorated only slightly compared to WBN (BW: correlation for HPN 0.92 cf. 0.97 for WBN; DF: 0.93 cf. 0.97). This is surprising since the HPN was limited to spectrally cover only the three highest channels in the CI. Further, monaural spectral cues were not available due to spectral scrambling between channels and the absence of energy in low-frequency channels which could otherwise facilitate level comparison to high-frequency channels. As carrier ITDs are not encoded in the CIs and the results with LPN suggest that they were not used, localization of the HPN must have relied on ILDs or envelope ITDs. A discrimination between both is not possible with the standard CI configuration in the free field. Experiments 2 and 3 were designed to yield a more definitive answer by going beyond the limitations of the free-field approach of experiment 1.

IV. EXPERIMENT 2

A. Stimuli and procedures

In experiment 1 ILDs and ITDs co-varied in natural fashion which limits the possible testing to manipulations of spectral content and temporal shape of the stimuli. In experiment 2 CI processors were raised 18 cm above the ears in order to minimize the acoustical effects of the head. Processors were mounted on thin rods which were fixed on a stiff plastic head band taken from a protective helmet. There were three possible configurations. In session "Together," processors were brought together in the center of the head, but 18 cm above ear level. This arrangement minimizes ILDs and ITDs. For the setting "Distanced," processors were placed exactly above the ears. In this setting processors were about one head diameter apart which roughly preserved natural ITDs but gave nearly no ILDs. In the "Plate" setting, processors were placed next to each other but separated by a carton board. The board had about the size of the cross section of the head (10*13 cm). The purpose of the board was to evoke ILDs, while ITDs were nearly absent due to the small distance between the processors. Localization procedures were identical to experiment 1 and a pulsed wide-band noise was used in all three sessions with subject DF (WBN, pulsed, Table I).

B. Results

Localization results of subject DF can be seen in Fig. 4 for three different processor placement conditions. Summary

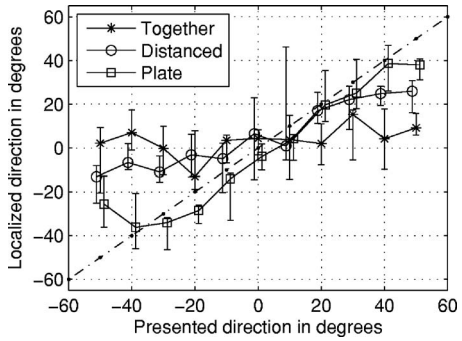


FIG. 4. Free-field localization results of experiment 2 of subject DF with elevated CI processors for pulsed wide-band noise (Table I). In the condition "Together" (*), processors were placed above the head next to each other, whereas in condition "Plate" (□) a cardboard was placed between the two processors that were still next to each other. The elevated processors were spaced at ear distance in the condition "Distanced" (o), but there was no plate between them. Medians are connected by lines and error bars show quartiles.

statistics are given in the lower part of Table III. In the first condition, "Together," processors were placed next to each other above the head. This placement minimizes ITDs and ILDs and, as expected, the pulsed wide-band noise stimulus could not be localized. The slope of the regression line shows independence of the presented direction (0.08) and quartiles are large (9.8°).

In condition "Distanced," processors were placed at ear distance above the ears which mainly introduces ITD cues. Differentiation of left from right became possible. The regression slope is shallow (0.39) and variance is high (quartiles 8.2°) which does not indicate that multiple directions could be localized. This suggests that ITDs contributed to localization but they provide only limited information.

A board separated both processors in condition "Plate", which introduced mainly ILD cues while ITDs were kept small. Localization improved considerably in this condition. While quartiles remained large (7.6°) localization covered a much broader azimuthal range and the regression slope reached 0.83. Although localization was nearly perfect within ±40°, responses for ±50° were compressed to the center. This indicates that ILD cues did not increase beyond ±40°. The pronounced increase of localization performance in the plate condition suggests that ILDs are the main localization cue, while the ability to discriminate left from right in the distanced condition indicates a small contribution of ITD cues. Since ITDs and ILDs were very inconsistent, for example, with ILDs being present while ITDs were near zero, it is also conceivable that the weighting of both localization cues differed strongly from the natural state. Experiment 3 was designed to investigate cue weighting particularly for HPN for which envelope ITDs and ILDs could be cues.

V. EXPERIMENT 3

A. Overview

Experiment 3 took a different approach than the previous experiments in two ways: (1) Localization of subject BW was investigated in virtual auditory space with direct input to the CI processors instead of playing sounds in the free field, and (2) Binaural cue weighting was studied by slightly offsetting binaural cues from their natural combination as this technique is thought to provide a better look at the normal cue weighting in the auditory system. In the present study, binaural stimuli prerecorded at one location were played directly into the processors while one binaural cue was filtered to originate from a different nearby location. A shift in localization towards the new location of the manipulated cue would indicate its contribution to directional perception.

B. Virtual acoustics with CIs

Head-related transfer functions (HRTFs) of Tempo+ processors, the type used by subject BW, were measured at the standard BTE positions on a human subject. The processor provides an output of the signal of the built-in microphone after the analog amplification and compression stage. The signal from the processor was routed via isolation transformers and measurement amplifiers to a measurement system or a digital recording system. Two measurements were done prior to the experiments:

(1) The device HRTFs were measured with maximum-length sequences (MLS) using audio measurement equipment (Audio Precision System One Dual Domain) inside an anechoic chamber. While the head was fixed by a headrest the subject was turned on a swivel chair and measurements were taken in 10° steps relative to the front. MLS sequences were played from a high-quality studio monitor (Klein & Hummel O98). Measured impulse responses were shortened to 512 taps at 44.1 kHz. Both processors were set to minimum amplification which is the least compressive setting. The remaining level offset between left and right HRTF was normalized according to the offset at frontal sound incidence.
(2) The test stimuli for experiment 3 (see below) were recorded directly off the processors with the amplification and compression settings commonly used by subject BW. The stimuli were played sequentially from the loudspeakers of the localization apparatus and the subject was seated at the standard position. The recorded stimuli contain all the directional information, ILDs and ITDs, that is forwarded to the later stages in the processor, and they include the alterations that stem from the compression stage of the processor. Since attack and release times of the compressors affect ILDs differently for each stimulus all test stimuli were recorded off the processor.

In experiment 3 the relative salience of ITD and ILD cues was tested by bringing both cues into opposition. However, according to the plausibility hypothesis the relative weighting of both cues will change if they are inconsistent (Hartmann, 1996; Wightman and Kistler, 1996). Thus, the natural weighting can only be studied for small cue discrepancies. Figure 1 outlines the approach taken here. ITDs and ILDs were brought into a small disparity of 20° or 40° by filtering the prerecorded stimulus with an ILD- or ITD-only filter (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). The filter contained only the difference in ITDs or ILDs between the new target location and the origi-

nal location while the recorded stimulus contained ITDs and ILDs belonging to the original location. The filtered signal can be thought of containing ILDs of one location and ITDs of the other. For linear, time-invariant systems this would be perfectly true, but here the situation was complicated by compression. The processing was done differently for ITD and ILD filters.

The ITD filter was computed from the fast Fourier transform (FFT)-phase spectra of the measured CI-HRTFs. The difference between the phase spectrum of the new and the old location was computed separately for the left and the right ear HRTF which preserves the difference in ITDs between both locations; 512-taps all-pass finite impulse response (FIR) filters were generated for both ears by the inverse FFT of the difference phase spectrum. If this pair of filters is applied to a signal that is already filtered with HRTFs of the original location, ITDs will exactly point to the new location while ILDs and the monaural spectrum still correspond to the original location. The procedure was verified by informal listening. It should be noted that the ITDs had to be computed from the measured CI-HRTFs since exact phase relationships could not be maintained between recordings of the stimuli. The computation could be done from HRTFs since the effect of compression on ITDs is small.

The situation is different for ILDs since they are severely affected by compression. The difference in ILDs between two directions was computed from the prerecorded test stimuli as a temporal average over the total duration of the recorded stimulus. This average ILD difference was incorporated into 512-point FIR filters with linear phase and equal delay for both ears. If a binaural stimulus is filtered this way ITDs remain at the original location while ILDs shift to the average ILD of the new location. It is worth pointing out again that compression affects the ILDs as a function of the stimulus and of time. As the computation was done separately for each stimulus, part of the variance between stimuli was captured by the procedure, however, the temporal dependence of ILDs on the stimulus could not be reproduced exactly. This might be especially important for transient stimuli as the compression takes time to react.

Prerecorded stimuli were played into the direct input of the CI processor that bypasses the compression stage. This way compression is applied only once to the signal. The stimuli were played from a PC-type computer via a digital soundcard, an external digital-to-analog converter and an amplification stage. The amplified signal was routed via analog attenuators with 0.1 dB resolution and isolation transformers to the direct input of the CI.

C. Stimuli and procedures

The weighting of ILDs and ITDs was studied with a special focus on envelope ITDs at high frequencies. Besides the pulsed wide-band noise of experiment 1 high-pass noise with pulsed or slowly changing envelope was used. The pulsation enhances the availability of envelope ITDs. Stimuli are described in Table I.

The procedures in experiment 3 were similar to experiment 1 but subject BW localized virtual stimuli that were fed directly into the speech processor instead of stimuli played via loudspeakers. At the beginning of the session the loudness of the directly fed prerecorded stimulus (WBN, pulsed) was adjusted to the loudness that the same stimulus evoked when played from the frontal loudspeaker at 69 dB SPL. The attenuation of the direct signal was changed recursively by the experimenter to minimize the loudness difference between free field and virtual presentation. The attenuation was also adjusted to yield a centered image for the stimulus played from the virtual front.

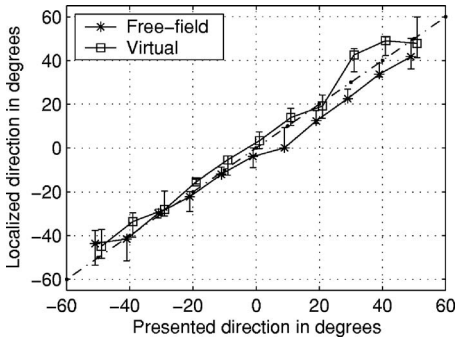


FIG. 5. Comparison of free-field localization (*) and virtual localization (□) of subject BW (experiment 3). The stimulus was pulsed wide-band noise (Table I).

Localization was tested for the three stimuli each for ILD and ITD shifts. The following conditions with disparity were run (original/cue shifted location): −50° paired with −30°, −10°; −30° paired with −50°, −10°, 10°; −10° paired with −50°, −30°, +10°, +30°, as well as their symmetrical counterparts on the right hand side. The conditions with disparity were interleaved in random order with conditions without disparity for the azimuths $-50°, -40°, \ldots, +50°$. Level was roved in 2 dB steps in 61–69 dB SPL. Ten trials were taken per condition which gave a total of 290 trials for each test sound. Similar to experiment 1 a light emitting diode lit up in the front for 5 s between blocks of 11 trials, but here trials were presented in completely randomized order. A break was introduced after 145 trials (13 min). Eleven trials without directional disparity were presented with feedback before the data collection began to help the subject get used to the virtual sound presentation.

Experiment 3 was run with subject BW 8 months after experiment 1. On the same day a single localization session identical to experiment 1 was administered to gather comparison data to free-field localization for the pulsed wide-band noise stimulus.

D. Results

Figure 5 shows localization results of subject BW taken in the free field and in virtual space without cue discrepancies for comparison and for verification of the technique. Table IV lists summary statistics. The free-field results, taken anew on the day of testing with virtual stimuli, confirm the outstanding localization performance for pulsed WBN (cf. Fig. 2). Surprisingly, the virtual stimulus presentation leads to a similar level of excellent localization performance, indicated by the ideal slope of the regression line of 0.97 and

TABLE IV. Summary statistics for localization results of subject BW in experiment 3, cf. Figs. 5 and 6: localization in the free field of pulsed wide-band noise and of virtual stimuli with consistent binaural cues. Virtual localization results were combined from trials without disparities from a session with ILD and one with ITD disparities. Statistics as in Table II.

Condition	Error (°)		Quartile (°)	Correlation coefficient	Regression line		Side diff
	Absolute	Arithmetic			Slope	Offset	
Free field							
WBN, pulsed[a]	5.2	−2.7	3.8	0.98**	0.90	−3.9	**
WBN, pulsed[b]	8.5	−7.8	3.2	0.97**	0.82	−6.5	**
Virtual presentation							
WBN, pulsed	6.3	4.3	4.4	0.99**	0.97	4.7	**
HPN, 100 ms env.	7.6	−3.6	5.4	0.99**	0.86	−3.7	**
HPN, pulsed	6.8	3.2	6.0	0.99**	0.95	4.2	**

[a]Taken on day of testing with virtual stimuli.
[b]From experiment 1 eight months earlier (cf. Table II).

very small quartiles of 4.4°. Quartiles increase with the transition from free field to virtual presentation by only 0.6°, while the regression slope increases as well and approaches 1. This suggests that virtual presentation captures all binaural information that is needed for horizontal localization by subject BW. The good level of performance with the prerecorded virtual stimuli is surprising given the fact that compression settings in the recording are likely to differ somewhat from the settings used by the subject on the day of testing, that recoding processors were only of the same type, but not identical to the processors used by the subject, and that a long signal transduction chain was used for recording and playback of virtual stimuli that is likely to alter the signals somewhat, e.g., in the isolation transducers. This is to our knowledge the first report of a successful application of virtual auditory space techniques with CI subjects.

Figure 6 shows base line localization results without cue manipulation for all stimuli in virtual space and Table IV lists associated statistics. Results for pulsed WBN were replotted from Fig. 5 for reference. Both high-pass noises (HPNs) could be localized very well; however, at first sight localization of the HPN with slow envelope changes seems slightly worse than localization of the pulsed HPN. This is evidenced in a slightly lower regression slope (0.86 for slow vs 0.95 for pulsed envelope) which is based on deviating responses for +20°, +30°, and +50°. However, the increase in quartiles with pulsation (5.4° to 6.0°) and the equally high correlation coefficient of 0.99 for both HPNs rather suggest that localization ability was stable, independent of envelope modulation. Thus, these results extend in virtual space the previous results of the free-field experiment 1 by adding that envelope ITDs seem not to contribute to localization of high-pass sounds while experiment 1 showed this for wide-band sounds only. This distinction is important since ITDs in the envelope of high-frequency stimuli can be detected very well in normal hearing as well as with CIs.

Even though envelope-ITDs might not be the dominant localization cue, their redundancy to other cues might cover their contribution to localization. If ITDs contribute, small offsets in ITDs should affect localization responses. An example of localization results for stimuli with cue discrepancies is in Fig. 7. The WBN from +10° was localized at about +10° if binaural cues were not manipulated, i.e., no offset was introduced. If ILDs were set to correspond to a position with an offset of +20° to the right, i.e., to +30°, localization responses shifted by about 12° rightwards. The introduced offset was not followed completely, but a strong sensitivity to ILDs was apparent. On the contrary, localization was unaffected by offsets in ITDs, which shows that ITDs did not contribute to localization in this condition.

Figure 8 presents summary results across all directions and disparity conditions for all stimuli. Localization of pulsed WBN follows an ILD offset by about 50% while offsets in ITDs are not followed by more than 10% at 20° offset

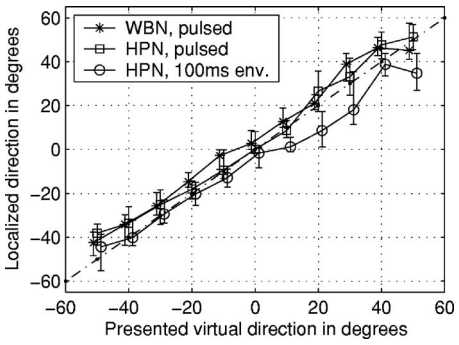


FIG. 6. Localization results of subject BW with virtual sound presentation without binaural cue disparities for different stimuli (Inlet and Table I).

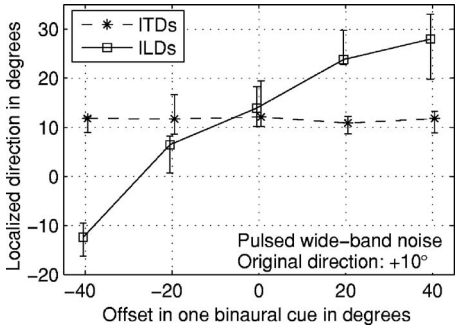


FIG. 7. Localization results of experiment 3 with binaural cue disparity (subject BW). A binaural recording of pulsed wide-band noise taken at +10° was manipulated so that one binaural cue (ILDs or ITDs) stemmed from a different direction ±20/40° away from +10°. Localization followed this offset only for ILDs.

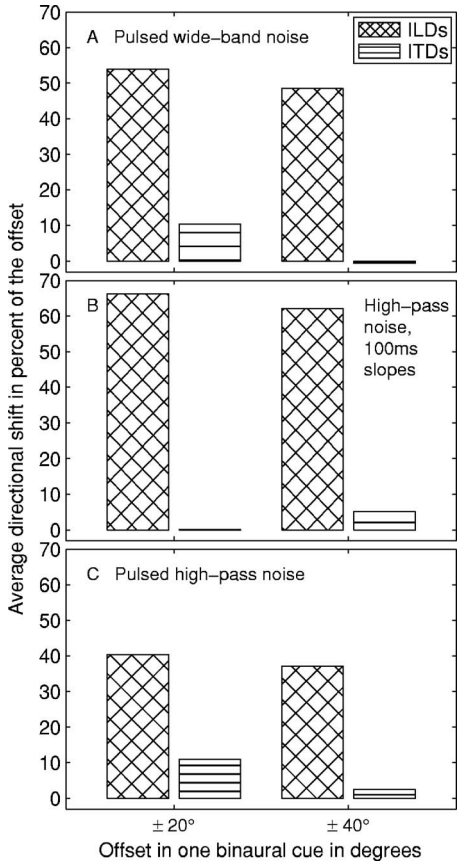


FIG. 8. Average directional shift observed in experiment 3 with subject BW for different disparities between ILD and ITD cues. Panels A–C give results for pulsed wide-band noise, high-pass noise with slow slopes and pulsed high-pass noise, respectively (cf. Table I). Data are averages over several originating directions.

(panel A). This confirms that ILDs provided the predominant information for localization of wide-band sounds for subject BW while the relative contribution of ITD cues is small despite the pulsation. For HPN with slow envelope changes ILD-weight seems to increase and ITD-influence appears absent (Panel B). The dominance of ILDs did not change even for pulsed HPN (panel C). The pulsation emphasized envelope ITDs and their weight increased, but it still remained low. For small discrepancies ILDs evoked directional shifts of about 40% while ITDs shifted directions by 10%. It should be noted that the directional shift of 2° corresponding to 10% ITD influence is well below the localization variance (quartiles of 6.0°). Interestingly, there might be a small trend for pulsed stimuli to yield a slightly higher ITD weighting at 20° compared to 40° offsets (panels A+C). This trend is consistent with the idea that cue weighting changes with cue discrepancy.

It is apparent from Fig. 8 that ILD-based shifts did not exceed 67% of the introduced offset despite the low contribution of ITDs. One reason for the apparently limited contribution of ILDs is that the cue weighting is computed relative to the introduced physical cue discordance, not relative to the change in localized direction. This has the advantage that localization errors in the base line without cue discordance have no effect on computed cue weighting, however, a base line localization slope smaller than one results in cue weighting below 100%. Given that most slopes are close to one (Table IV) we found that the resulting reduction in weight would be acceptable as the maximal error should be in the order of the observed variance. Another reason for the reduced contribution of ILDs can be found in the way ILD filters were generated. ILDs were computed by integration over the total duration of the recorded stimulus which included compression. While integration over pauses in pulsed stimuli should not offset the ILD-estimation per se, the effect of compression could. ILDs are larger at the beginning of the pulses before compression starts. The averaged ILDs incorporated in the filters were thus smaller than ILDs at sound onsets. This will lead to compression of localization responses towards the front if the subject would evaluate ILDs mostly from sound onsets.

It can be concluded from experiment 3 that ILDs were the dominant cue for localization of subject BW for wide-band and for high-pass stimuli and that ITDs played only a minor role even if stimuli were envelope modulated.

VI. OVERALL DISCUSSION

The relative weighting of binaural cues with bilateral CIs was tested in a localization task with three different methods, in both free field and virtual space. In experiment 1, two subjects localized sounds differing in spectral and temporal composition in the free field. Both subjects showed the ability to discriminate the side of sound origin for wide-band noise (WBN) with a single CI, indicating the evaluation of monaural spectral cues. Using two CIs, both subjects were able to localize WBN with pulsed or slowly changing envelope with similar, high accuracy, suggesting that envelope ITDs played only a minor role. Low-pass noise (LPN) could not be localized by subject BW, while performance for subject DF was good for a cut-off frequency at 525 Hz, but deteriorated clearly when the cut off was reduced to 400 Hz. This indicates a strong contribution of remaining ILDs as ITD evaluation should not be affected by the reduction in bandwidth. Both subjects localized high-pass noise (HPN) with scrambled spectrum nearly as well as WBN which indicates the use of interaural cues. Experiments 2 and 3 were designed to clarify the contribution of ITDs by artificially dissociating ITDs from ILDs.

In experiment 2, localization of subject DF was studied with physically manipulated binaural cues in the free field. It was meant to elucidate the contribution of ITDs in light of his inconclusive results with LPN. CI processors were placed above the head and localization of WBN was best if a cardboard was placed between processors, while mere processor spacing by ear distance that gave mainly rise to ITDs elicited only discrimination of left from right. Experiment 2 shows that ILDs were the more efficient localization cue for subject DF while ITDs contributed somewhat, consistent with the results of experiment 1.

Experiment 3 was designed to clarify the contribution of envelope ITDs to localization of HPN by subject BW which could not be answered by experiment 1. The introduction of a discordance between ITD and ILD cues in virtual acoustical space provided a clean way to assess their relative importance. Virtual localization of subject BW was shown to be

similar to the free field. When binaural cues were brought in conflict to each other, localization of WBN or HPN followed the introduced offset more for ILD shifts while ITDs contributed slightly only when the envelope was pulsed. In agreement with the results of experiment 1 the data from virtual space show a clear ILD dominance for subject BW. Experiment 3 adds that ILDs dominate even when the evaluation of ITDs was emphasized by pulsation.

A. Interaural time differences

The preprocessing and the pulsatile stimulation in CIs alter monaural and binaural cues available to the subject. Signal processing for CIS-type processors, as used by the subjects of this study, consists of compression, bandpass filtering, envelope extraction and pulse forming parts. The log-transformed value of the envelope is used to determine the current of the stimulation pulses while their temporal position is independent of the signal and pulses follow a fixed rate (Zierhofer *et al.*, 1995). This leads to a dissociation of fine-structure and envelope information. The fine structure does not carry information about the signal and hence the best strategy for the auditory system would be to ignore it. In support of this view CIs use stimulation rates too high to be followed directly by the auditory nerve. As pulses are not placed temporally with respect to the signal phase ITDs are not encoded directly and hence ITDs cannot be detected through the phase at low frequencies. The problem is enhanced by the fact that processors are not synchronized between the ears. The lack of synchronization introduces arbitrary ITDs that change from day to day or drift during the day. Subject DF used different stimulation rates on both ears which lead to varying carrier ITDs. A good strategy for the auditory system would be to ignore unreliable, changing carrier ITDs.

Even though the envelope is sampled at a rate that is too low for a good direct representation of ITDs it seems that the auditory system can nevertheless extract ITDs fairly well from the *envelope*. Envelope-ITD thresholds with CIs can be close to thresholds seen in normal hearing and values of 25 μs have been reported in two subjects (Lawson *et al.*, 2001). In comparison, only a few subjects show thresholds for carrier ITDs down to 50 μs on selected electrodes; the same subjects show higher thresholds on most other electrode pairs. Thresholds often range from 100 to 300 μs while some subjects cannot detect carrier ITDs of 1 ms (van Hoesel and Clark, 1997; Lawson *et al.*, 1998; Lawson *et al.*, 2000; van Hoesel *et al.*, 2002; Long *et al.*, 2003; van Hoesel and Tyler, 2003). An exception is subject DF who showed very low ITD thresholds across many electrode combinations, the reason for which he was chosen for the current study. Lawson *et al.* (2000) identified five electrode combinations for him (coded as ME8) with ITD thresholds lower than 50 μs and 16 other combinations with thresholds lower than 150 μs. In the same study, subject BW (ME5) showed an ITD threshold below 50 μs on one electrode combination, below 150 μs on another two electrode combinations, while four combinations showed thresholds of 1–2 ms. Despite the good sensitivity of DF to ITDs in the carrier the results of the current study indicate that his localization was not dominated by ITDs. This finding can be explained at least partly by the fact that the CIS strategy does not encode ITDs in the carrier pulses. But the question remains why both subjects, DF as well as BW, weighted ITDs low which *are* transmitted in the envelope and which *are* detectable, but relied on ILDs for localization instead.

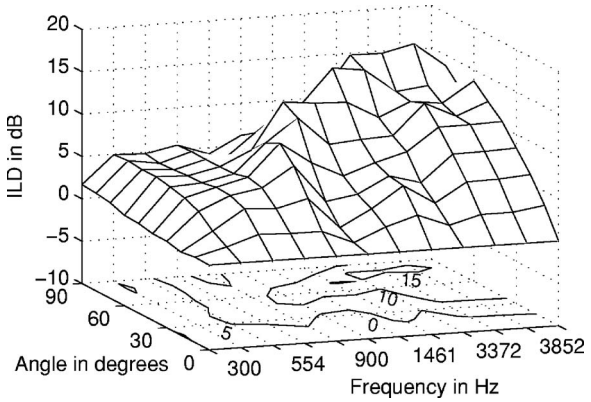


FIG. 9. ILDs computed from HRTFs measured on Med-El Combi 40+ processors. ILDs were derived from the level of HRTF-filtered narrow-band noise signals. The frequencies and the bandwidth of the noise corresponded to the log-spaced channel mapping in 300–5547 Hz used by subject BW. Tickmarks on the frequency axis correspond to CI channels. Levels were normalized to be zero for frontal sound incidence to correct for amplification offsets in both CI processors. ILDs were averaged for sound from the right and the left side.

B. Interaural level differences

One possible reason why the contribution of ITDs is small is because sensitivity to ILD changes is far better than to ITDs for most subjects. ILD thresholds were often as low as the minimum possible current step (van Hoesel *et al.*, 1993; Lawson *et al.*, 1998; Lawson *et al.*, 2000; van Hoesel and Tyler, 2003). This could be as low as 16 μA at 1.1 mA current, which is equivalent to a change of 0.125 dB at a dynamic range of 9.5 dB (Lawson *et al.*, 1998). Given this high sensitivity, we assessed the physically occurring ILDs at the input of the speech processors. Figure 9 shows ILDs in the measured HRTFs of the speech processors computed in channel bands for the common logarithmic map in 300–5547 Hz as used by subject BW. For frontal sound incidence and for low-frequency channels ILDs are small ($<$5 dB). They are unlikely to serve as a good localization cue, but they might provide some information in $\pm$30° where a monotonic increase with angle can be observed in channels 2–4. For frequencies beyond 800 Hz (channel 5) ILDs change over a larger angular range and ILDs up to 15 dB can be observed. A strong monotonic increase of ILDs with angle can be seen for the upper 3 CI channels between 0° and 60°. Beyond 60° ILDs do not increase further, but a channel specific pattern of peaks and troughs could be evaluated by the auditory system. In general, the pattern of ILDs measured at the CI-speech processors is consistent with ILDs measured in the ear canal but some fine detail is lost due to the averaging in CI channels. The occurrence of ILDs up to 5 dB in low-frequency channels 2–4 might seem surprising at first. However, HRTFs measured in the ear canal have similar ILDs in the corresponding frequency range of 500–1000 Hz. ILDs of

5 dB may be sufficient to allow for localization of frontal directions. Subject DF showed coarse localization of 400 Hz low-pass noise and good localization of 500 Hz LPN and pulsed LPN in experiment 2. We analyzed the low-pass stimuli of experiment 2 for their spectral content. The −20 dB points of the averaged spectrum were found at 710 Hz both for the pulsed noise and for the 500 Hz LPN. One would assume that pulsation leads to spectral widening. However, due to the 40 dB spectral scrambling applied to the pulsed LPN the −20 dB bandwidth turns out to be wider in only 50% of the sequences, whereas it is smaller in the other 50%. Both sounds led to about the same localization performance, which is in support of an ILD-based localization process since the underlying bandwidth was identical. For the 400 Hz LPN the −20 dB point lies at 592 Hz. Given the electrode mapping of subject DF the 400 Hz LPN covers only 2/3 electrodes at the left and right ear, respectively, whereas the 500 Hz LPN provides energy over 3/4 electrodes. Localization performance was considerably worse with the 400 Hz LPN which is consistent with the limitation in spectral range for ILD evaluation. If the localization process was based on ITDs this slight limitation in spectral extent should not reduce the envelope-ITD information considerably.

Further support for an ILD-dominated localization process with CIs stems from lateralization studies. Van Hoesel and Tyler (2003) showed lateral position of 50 pps pulse sequences to be more dependent on ILDs than ITDs in 5 out of 5 tested subjects when electrodes were stimulated directly through a research interface. The stronger dependency of lateralized position on ILDs is also in line with results by Long *et al.* (2003) and Schoen *et al.* (2005).

C. Plausibility of localization cues

According to the plausibility hypothesis less weight is given to localization cues that are in conflict across the spectrum or against other localization cues (Hartmann, 1996; Wightman and Kistler, 1996). This applies to ITDs with CIs. ITDs in the carrier pulses are random with unsynchronized devices while envelope ITDs transmit useful information, albeit quantized. Assuming that the binaural system derives ITDs from the carrier in low-frequency regions and from the envelope at high frequencies inconsistent ITDs will be computed across the spectrum and the binaural system should thus ignore ITDs altogether according to the plausibility hypothesis. Subject DF uses different stimulation rates on both ears which potentially introduces varying ITDs. Although recent studies suggest that it is unlikely that the auditory system extracts carrier ITDs at stimulation rates above 1400 pps (Majdak *et al.*, 2006; van Hoesel, 2007), the variable carrier ITD would contrast with the envelope ITD. According to the plausibility hypothesis this could reduce DF's reliance on ITDs. Also, ITD sensitivity varies largely across electrodes, whereas ILD detection is consistently good on most electrodes. This high sensitivity for ILDs across many electrodes is more likely to provide the binaural system with a reliable, consistent localization cue.

D. Consistency with other studies

The dominant influence of ILDs on localization is utilized in the loudness balancing procedure at the beginning of the localization tests. By changing the amplification in the devices the auditory image can be centered which relies on the dominant influence of ILDs for lateralization. Localization offsets for frontal sounds simply depend on the relative amplification in both devices—a setting that is changed often throughout the day. Thus, offsets do not characterize the localization *ability* of the subject. The localization method must be able to distinguish between offsets, localization slope, and variance since only the latter two characterize the auditory system in the case of CIs (Seeber, 2002). The localization method used here provides this information.

In the present as well as in several other studies with bilateral CIs sound sources from the sides were localized more towards the center (van Hoesel *et al.*, 2002; van Hoesel and Tyler, 2003; Nopp *et al.*, 2004; Schoen *et al.*, 2005). In the discussion to Fig. 9 this was related to a compression of ILDs in the HRTFs for azimuths from 60° to 90°. Alternately, azimuth compression might also occur for overly compressive settings of the mapping law. A too compressive map will reduce the stimulation current at high levels more than necessary and thus compress effective ILDs. Likewise, the automatic gain control (AGC) employed in most processors compresses ILDs in a similar way. As compression does not severely affect envelope ITDs the observation of a compressed azimuthal range suggests an ILD-dominated localization process.

VII. CONCLUSIONS

The relative dominance of binaural cues was studied with three different approaches in two bilateral CI subjects with excellent localization ability. Although both subjects did not participate in all experiments the results show consistently that both subjects predominantly relied on ILDs for localization of all tested types of sounds while ITDs contributed only a small amount of information. ITD influence seemed strongest when their evaluation was emphasized by modulating the sound envelope, but it still remained far below the contribution of ILDs. In contrast, in normal hearing ITDs dominate localization for most wide-band sounds and ILDs play a role only for some high-frequency sounds. CI subjects thus show reversed dominance of localization cues compared to normal hearing. As ITDs play only a minor role, CI subjects loose the redundancy to rely on either localization cue. As long as subjects face situations in which the single localization cue provides enough information, no shortcomings in localization might occur. However, this might be different for situations with multiple sounds in which one cue might prove unreliable and the redundancy of cues would be needed to resolve ambiguities. We predict that localization of CI subjects would suffer considerably in those situations similar to the strong deterioration of speech understanding in background noise (Qin and Oxenham, 2003; Firszt *et al.*, 2004).

ACKNOWLEDGMENTS

We would like to thank Dr. Uwe Baumann for providing the contact to subject BW and both subjects for their participation. Patient DF contributed to experiment 2 with ideas and by bringing the raisers for the processors. We thank Dr. Peter Nopp of Med-El GmbH, Innsbruck, for lending us two CI processors, isolation transformers, and cables for the measurement of CI-HRTFs. Med-El GmbH paid for flight and accommodation of subject DF. Dr. Sridhar Kalluri gave many valuable suggestions regarding the manuscript for which we are very grateful.

Blauert, J. (**1997**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., Ruge-Samuelson, C. L., and Wackym, P. A. (**2004**). "Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems," Ear Hear. **25**, 375–387.

Freyman, R. L., and Zurek, P. M. (**1997**). "Onset dominance in lateralization," J. Acoust. Soc. Am. **101**, 1649–1659.

Grantham, D. W., Ashmead, D. H., and Ricketts, T. A. (**2005**). "Sound localization in the frontal horizontal plane by post-lingually deafened adults fitted with bilateral cochlear implants," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. d. Cheveigne, S. McAdams, and L. Collet, Springer-Verlag, New York, pp. 390–397.

Hafter, E. (**1996**). "Binaural adaptation and the effectiveness of a stimulus beyond its onset," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, Mahwah, NJ), pp. 211–232.

Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Hartmann, W. A. (**1996**). "Listening in a room and the precedence effect," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, Mahwah, NJ), pp. 191–210.

Hartmann, W. M., Rakerd, B., and Koller, A. (**2005**). "Binaural coherence in rooms," Acta. Acust. Acust. **91**, 451–462.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.

Laback, B., Pok, S. M., Baumgartner, W. D., Deutsch, W. A., and Schmid, K. (**2004**). "Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors," Ear Hear. **25**, 488–500.

Laszig, R., Aschendorff, A., Stecker, M., Mueller-Deile, J., Maune, S., Dillier, N., Weber, B., Hey, M., Begall, K., Lenarz, T., Battmer, R.-D., Boehm, M., Steffens, T., Strutz, J., Linder, T., Probst, R., Allum, J., Westhofen, M., and Doering, W. (**2004**). "Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results," Otol. Neurotol. **25**, 958–968.

Lawson, D. T., Brill, S., Wolford, R. D., Wilson, B. S., and Schatzer, R. (**2000**). "Speech processors for auditory prostheses," in *Ninth Quarterly Progress Report* (Research Triangle Park: Research Triangle Institute).

Lawson, D. T., Wilson, B. S., Zerbi, M., Honert, C. v.-d., Finley, C. C., Farmer, J. C., McElveen, J. T., and Roush, P. A. (**1998**). "Bilateral cochlear implants controlled by a single speech processor," Am. J. Otol. **19**, 758–761.

Lawson, D. T., Wolford, R. D., Brill, S., Schatzer, R., and Wilson, B. S. (**2001**). "Speech processors for auditory prostheses," in *Twelfth Quarterly Progress Report* (Research Triangle Park: Research Triangle Institute).

Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Majdak, P., Laback, B., and Baumgartner, W. D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Nopp, P., Schleich, P., and D'Haese, P. (**2004**). "Sound localization in bilateral users of MED-EL COMBI 40/40+ cochlear implants," Ear Hear. **25**, 205–214.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Schoen, F., Mueller, J., Helms, J., and Nopp, P. (**2005**). "Sound localization and sensitivity to interaural cues in bilateral users of the Med-El Combi 40/40+cochlear implant system," Otol. Neurotol. **26**, 429–437.

Seeber, B. (**2002**). "A new method for localization studies," Acta. Acust. Acust. **88**, 446–450.

Seeber, B., Baumann, U., and Fastl, H. (**2004**). "Localization ability with bimodal hearing aids and bilateral cochlear implants," J. Acoust. Soc. Am. **116**, 1698–1709.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Strutt, J. W. L. R. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A. W., Preece, J. P., Witt, S., and Lowder, M. W. (**2002**). "Three-month results with bilateral cochlear implants," Ear Hear. **23**, 80S–89S.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

van Hoesel, R., Ramsden, R., and O'Drisoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.

van Hoesel, R. J. (**2007**). "Sensitivity to binaural timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.

van Hoesel, R. J. M. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neuro-Otol. **9**, 234–246.

van Hoesel, R. J. M., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects," J. Acoust. Soc. Am. **102**, 495–507.

van Hoesel, R. J. M., Tong, Y. C., Hollow, R. D., and Clark, G. M. (**1993**). "Psychophysical and speech perception studies: A case report on a binaural cochlear implant subject," J. Acoust. Soc. Am. **94**, 3178–3189.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Wightman, F. L., and Kistler, D. J. (**1996**). "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey, and T. R. Anderson (Lawrence Erlbaum, Mahwah, NJ), pp. 1–23.

Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.

Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.

Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., and Chen, H. (**2004**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351–1354.

Zierhofer, C. M., Hochmair-Desoyer, I., and Hochmair, E. S. (**1995**). "Electronic design of a cochlear implant for multichannel high-rate pulsatile stimulation strategies," IEEE Trans. Rehabil. Eng. **3**, 112–116.

Assessing the pitch structure associated with multiple rates and places for cochlear implant users

Joshua S. Stohl,[a] Chandra S. Throckmorton,[b] and Leslie M. Collins[c]
Department of Electrical and Computer Engineering, Duke University, P.O. Box 90291, Durham, North Carolina 27708-0291

(Received 9 May 2007; accepted 14 November 2007)

Cochlear implant subjects continue to experience difficulty understanding speech in noise and performing pitch-based musical tasks. Acoustic model studies have suggested that transmitting additional fine structure via multiple stimulation rates is a potential mechanism for addressing these issues [Nie *et al.*, IEEE Trans. Biomed. Eng. **52**, 64–73 (2005); Throckmorton *et al.*, Hear. Res. **218**, 30–42 (2006)]; however, results from preliminary cochlear implant studies have been less compelling. Multirate speech processing algorithms previously assumed a place-dependent pitch structure in that a basal electrode would always elicit a higher pitch percept than an apical electrode, independent of stimulation rate. Some subjective evidence contradicts this assumption [H. J. McDermott and C. M. McKay, J. Acoust. Soc. Am. **101**, 1622–1630 (1997); R. V. Shannon, Hear. Res. **11**, 157–189 (1983)]. The purpose of this study is to test the hypothesis that the introduction of multiple rates may invalidate the tonotopic pitch structure resulting from place-pitch alone. The SPEAR3 developmental speech processor was used to collect psychophysical data from five cochlear implant users to assess the tonotopic structure for stimuli presented at two rates on all active electrodes. Pitch ranking data indicated many cases where pitch percepts overlapped across electrodes and rates. Thus, the results from this study suggest that pitch-based tuning across rate and electrode may be necessary to optimize performance of a multirate sound processing strategy in cochlear implant subjects. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821980]

PACS number(s): 43.66.Ts [DOS] Pages: 1043–1053

I. INTRODUCTION

There is evidence that while speech recognition by cochlear implant users is relatively good, speech recognition in noise and the ability to identify musical patterns remains poor (e.g., Gfeller *et al.*, 2003; McDermott, 2004). Previous studies have shown that increasing the available spectral information may lead to an improvement in performance on speech recognition and melody identification tasks (Dorman *et al.*, 1997; Friesen *et al.*, 2001; Fu *et al.*, 1998; Nie *et al.*, 2005; Throckmorton *et al.*, 2006; Zeng *et al.*, 2005). Dorman *et al.* (1998) used acoustic models to conclude that an increase in the number of channels in a cochlear implant is directly related to the ability to recognize speech in noisy conditions, and Fu *et al.* (1998) found that the poor performance of cochlear implant users with respect to normal hearing individuals was most likely due to an inability to use spectral place cues as a result of the relatively small number of implanted electrodes. However, increasing the number of spectral channels by simply increasing the number of electrodes is not an option for those individuals who have already received a cochlear implant, and channel interaction acts as a limiting factor for electrode array design. Others have proposed increasing spectral information within the available channels via current steering or multirate strategies. Current steering relies on the stimulation of two electrodes such that they create a pitch percept that is in between the percepts elicited by either electrode alone (Donaldson *et al.*, 2005; Koch *et al.*, 2007; McDermott and McKay, 1994). Multirate stimulation strategies that use either a predetermined set of stimulation rates or constantly varying stimulation rate on some or all electrodes to increase the spectral resolution of each channel (Fearn, 2001; Nie *et al.*, 2005; Nobbe, 2004; Throckmorton *et al.*, 2006) have been proposed as well. Given that acoustic model simulations of multirate strategies show promising results for increasing spectral resolution, this study seeks to investigate the pitch structure for cochlear implants as it is affected by introducing multiple rates.

Numerous studies available in the literature demonstrate that increasing the stimulation rate on a single electrode typically results in a monotonic increase in pitch up to 300 pps, and as high as 1000 pps in a small number of subjects, at which point the rate pitch percept saturates (Fearn *et al.*, 1999; McDermott and McKay, 1997; Pijl and Schwartz, 1995; Shannon, 1983a; Tong and Clark, 1985; Townshend *et al.*, 1987; Zeng, 2002). However, in a multirate algorithm, many electrodes and rates are acting together. Some research (Collins and Throckmorton, 2000; McKay *et al.*, 1996) has suggested that in this scenario, percepts are two dimensional. In order to assess the feasibility of a multirate strategy, it is important to determine whether subjects can make single dimensional pitch judgments for combined rate and place

[a]Electronic mail: jss@ee.duke.edu.
[b]Electronic mail: cst@ee.duke.edu.
[c]Current address: Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708-0291. Electronic mail: lcollins@ee.duke.edu.

TABLE I. Demographic information for implanted subjects.

Subject ID	Gender	Age (years)	Age at onset of deafness (years)	Age at implantation (years)	Mode of simulation	Speech recognition (% correct)
S2	F	71	46	66	MP1+2	92.43
S4	M	19	4	12	MP1+2	96.76
S5	F	58	26	54	MP1+2	89.73
S6	M	60	54	57	MP1+2	90.30
S7	M	53	50	50	MP1+2	97.30

stimuli. It may also be important to understand the potentially complex structure of the elicited pitch space.

For example, Eddington *et al.* (1978) used a pitch scaling procedure to demonstrate that stimulating neighboring pairs of electrodes, the more apical at 300 pps and the more basal at 50–70 pps, resulted in a pitch reversal in four out of five pairs of electrodes when compared with the results from stimulating the two electrodes at the same rate. However, only one subject participated in this study, and only six of eight electrodes were active. Results from the pitch scaling studies done by Fearn *et al.* (1999) and Zeng (2002) also indicate the possibility of eliciting a nonmonotonic pitch structure when using multiple stimulation rates across various electrodes. Fearn *et al.* (1999) asked subjects to assign a pitch value to stimuli on three apical electrodes at rates spanning 100–500 pps and three widely spaced electrodes at rates spanning 100–1000 pps. In both cases, stimuli presented on more apical electrodes were given a higher average pitch rating than lower rate stimuli on more basal electrodes. Zeng (2002) observed the same phenomenon up to approximately 300 pps when four cochlear implant subjects were asked to assign a pitch magnitude to rates ranging from 50 to 2000 pps on only the most apical and most basal electrodes. While both of these studies support earlier findings of pitch overlap across electrodes, only a subset of electrodes were tested. None of these studies mapped the entire pitch structure associated with rate and place, nor did they utilize pitch ranking. Pitch ranking may be preferred to pitch scaling, as Collins *et al.* (1997) found a standard deviation in pitch estimates of 10% when using the scaling paradigm.

Anecdotal results from Shannon (1983a) also indicate that changing the stimulation rate on a given electrode may elicit a change in pitch greater than that elicited by a change in place/electrode. More explicitly, results from McDermott and McKay (1997) indicate that changes in rate may correspond to changes in pitch of up to two octaves. McDermott and McKay (1997) performed three tests with a single implant user who was trained as a piano tuner and was therefore able to give accurate estimates of pitch intervals as a function of rate, place, or a combination of rate and place. In that study stimulation rates varied from 100 to 200 pps in intervals of musical semitones, defined by a ratio of $2^{1/12}$. This 100 pps change in rate corresponded to a perceived change in pitch of up to 8 semitones, while a basal shift of seven electrodes was required to elicit approximately the same change in pitch. The only direct comparison of multiple rates applied to multiple electrodes was made in the form of increasing rate from 100 to 200 pps in either the apical or basal direction across 13 electrodes. When comparing increasing rates from the apex to base with decreasing rates from apex to base, a compression of the range of perceived intervals was found in the decreasing rate case with a few instances of place-pitch violating the tonotopic ordering of the cochlea.

The goal of this study was to better understand the combined influence that rate and place have on the perceptual pitch structure of implant users. This information may be useful when implementing a strategy in which multiple rates are meant to represent different channels of information. Throckmorton *et al.* (2006) proposed the multicarrier frequency algorithm (MCFA), which may be thought of as a quantized version of the variable stimulation rate algorithms proposed by Fearn (2001) and Nie *et al.* (2005). In MCFA a discrete number of frequencies are available on each channel as opposed to a continuum. Using an acoustic model, that study showed that a significant improvement in speech recognition in noise could be achieved by adding one additional frequency per channel. This finding motivates an initial assessment of a complete two-rate pitch space in cochlear implants. In order to determine that pitch space, a series of paired comparison pitch ranking procedures were implemented. Biphasic, rectangular pulse trains were presented at one rate on all active electrodes to probe the pitch structure due only to the place of stimulation. Subsequently, stimuli at two discriminable rates were used to provide an estimate of the pitch structure as a function of place and rate across the entire array.

II. METHODS

A. Subjects

Five postlingually deafened cochlear implant users participated in this study. All participants were implanted with Cochlear Corporation's Nucleus CI24 cochlear implants and had at least 2 years of experience with the device. Demographic information for each user can be found in Table I. The results from a speech recognition task that used a randomization of the CID sentences in quiet are also listed in Table I. Sentences were presented through desktop computer speakers in a soundproof booth at a level set by the subject. The maximum possible level of presentation was 80 dB SPL.

Only electrodes available in the user's clinical MAP were stimulated during testing. A clinical MAP contains information regarding stimulation parameters and is specific to

the user. For example, a MAP may include threshold and maximum comfort levels, mode of stimulation, timing parameters, and may indicate inactive electrodes. All participants use a monopolar 1+2 (MP1+2) mode of stimulation in their clinical MAP, and all subjects used a MP1+2 stimulation mode during this study with the exception of subject S7, who used a monopolar 1 (MP1) mode of stimulation. This change in the mode of stimulation is necessary when interfacing the SPEAR3 with a CI24RE, or Freedom, cochlear implant (Chris van den Honert, Cochlear Corporation and Andrew Vandali, Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation, private communication). When using the monopolar mode of stimulation with a CI24 cochlear implant, one (MP1/MP2) or both (MP1+2) extra-cochlear electrodes act as ground for all intracochlear electrodes. When electrodes are identified by a single number, that number refers to the intracochlear electrode position, where 1 is the most basal electrode and 22 is the most apical electrode. All data were collected in three to five sessions that lasted 2–4 h each. Subjects were compensated for their time with the exception of subject S7 who elected to volunteer his time, and approval by the Duke University Institutional Review Board was obtained for all experiments and subject S7's volunteer status.

B. Stimuli

The stimuli used in this study were 300 ms pulse trains containing biphasic rectangular pulses with 25 μs pulse widths and an 8 μs interphase gap. The interstimulus interval was fixed at 500 ms for all experiments, and stimulation rates of 199 and 398 pps were used in all experiments except for the rate discrimination task in which the rate of stimulation varied. These two rates were selected to be below the typical rate at which the rate-pitch percept saturates (500 pps). As in McDermott and McKay (1997), these rates also provide a range of one musical octave, or a doubling in frequency, and in contrast to that study, span a larger range of overall stimulation rate.

Threshold and maximum comfortable loudness levels were measured for all active electrodes using the SPEAR3 prior to implementing any other experiments. Subjects were then asked to adjust the amplitude of the most apical active electrode to a comfortable level at a stimulation rate of 199 pps, and all other electrodes were loudness balanced to the next closest available apical electrode in an adjacent fashion (Throckmorton and Collins, 2001). The results of the electrode loudness balancing procedure were then used to loudness balance a stimulation rate of 398 pps to a reference rate of 199 pps on the same electrode, and this procedure was repeated for all active electrodes.

In order to verify that 398 pps was discriminable from 199 pps, an adaptive two-down, one-up, two-interval, forced-choice Levitt procedure was implemented with flanking cues to measure the pulse rate just noticeable difference (JND) for 199 pps. Here the term flanking cues refers to the presentation of a reference interval both before and after the presentation of the two intervals that may contain the target stimulus. Rate discrimination difference limens were measured at three locations along the cochlea (apical, middle, and basal) with the assumption that these results would be indicative of JNDs across the entire electrode array.

C. Pitch ranking task

In the paired-comparison pitch ranking procedure, subjects were asked to select the higher of two pitches in each trial. Subjects performed two pitch ranking tasks in this study. In the first pitch ranking task, all active electrodes were compared at a single stimulation rate, 199 pps, in order to determine a pitch structure based solely on place of stimulation. Subjects ranked 18–21 electrodes based on the number of available electrodes in their clinical MAP. The set of stimuli used in the second pitch ranking task contained two rates on all active electrodes. Ranking both 199 and 398 pps on 18–21 electrodes resulted in 36–42 total stimuli for the two-rate ranking task. A block of trials consisted of a single comparison of each of the electrodes using the single-rate stimuli or a single comparison of each of the two-rate stimuli in one of three regions of the cochlea (explained in the following). Each block was repeated seven to ten times for each subject. All pairs of stimuli were randomized within each block and the order of presentation was randomized within each trial. Blocks were typically presented in the following order: single-rate, two-rate apical, two-rate middle, two-rate basal.

A single repetition requires that each stimulus be compared to all other stimuli in the set. The entire two-rate set, which included 36–42 stimuli depending on the subject's number of active electrodes, would correspond to 630–861 paired comparisons in a single repetition. At an average response time of 3 s per pair, this would require subjects to perform the pitch ranking task for an uninterrupted duration of over 30 min. In order to avoid exhaustion, the array was thus subdivided into three overlapping sections (apical, middle, and basal) for pitch ranking in the two-rate task. The apical and basal subsets of stimuli shared one common electrode, and the middle subset included electrodes from both the apical and basal sections of the array. The resulting average duration for a single block was 10 min for one block of trials.

D. Experiment platform

All experiments were implemented in a SPEAR3-based psychophysical paradigm developed at Duke University (Stohl *et al.*, 2007). The hardware equipment includes a SPEAR3 developmental speech processor (Hearworks Pty Ltd.) connected to an ACER desktop PC via a serial port. The software that controlled the stimuli used a modified assembly language file that allowed the dynamic updating of stimulus values in combination with graphical user interfaces created in VISUAL BASIC that provided visual cues and accepted user responses. User responses were provided by clicking a mouse, selecting a value on a keyboard, or turning and pressing a knob, and visual stimulation cues and feedback were provided via a LCD flat panel Sony monitor. The psychophysical procedure was also controlled by VISUAL BASIC code that updated stimulation parameters based on previous responses from the user.

TABLE II. Preference table for single rate task—Subject S7.

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}	E_{14}	E_{15}	E_{16}	E_{17}	E_{18}	E_{19}	E_{20}	E_{21}	E_{22}	Score
E_1	⋯	4	6	5	7	7	7	7	6	7	7	7	7	7	7	7	7	7	7	7	7	133
E_2	3	⋯	5	7	6	7	7	7	7	7	6	6	7	7	7	7	7	6	7	6	7	129
E_3	1	2	⋯	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	128
E_4	2	0	1	⋯	6	7	7	5	7	5	7	6	6	6	7	6	6	7	7	7	7	112
E_5	0	1	0	1	⋯	3	7	7	7	7	7	7	7	5	6	7	7	6	7	6	7	105
E_6	0	0	0	0	4	⋯	6	6	6	5	5	5	6	5	5	7	5	7	7	7	6	92
E_7	0	0	0	0	0	1	⋯	7	3	7	6	6	7	7	6	5	6	7	5	7	7	87
E_8	0	0	0	2	0	1	0	⋯	3	6	6	4	7	6	5	7	6	4	6	5	6	74
E_9	1	0	0	0	0	1	4	4	⋯	7	6	6	6	5	6	3	6	3	6	7	6	77
E_{10}	0	0	0	2	0	2	0	1	0	⋯	4	7	6	7	7	7	7	6	6	6	7	75
E_{11}	0	1	0	0	0	2	1	1	1	3	⋯	6	6	7	5	6	6	6	6	6	7	70
E_{12}	0	1	0	1	0	2	1	3	1	0	1	⋯	5	6	6	6	7	6	5	6	7	64
E_{14}	0	0	0	1	0	1	0	0	1	1	1	2	⋯	5	7	7	7	6	6	7	7	59
E_{15}	0	0	0	1	2	2	0	1	2	0	0	1	2	⋯	7	7	7	7	5	7	7	58
E_{16}	0	0	0	0	1	2	1	2	1	0	2	1	0	0	⋯	7	6	6	6	7	7	49
E_{17}	0	0	0	1	0	0	2	0	4	0	1	1	0	0	0	⋯	5	7	5	7	5	38
E_{18}	0	0	0	1	0	2	1	1	1	0	1	0	0	0	1	2	⋯	7	5	7	7	36
E_{19}	0	1	0	0	1	0	0	3	4	1	1	1	1	0	1	0	0	⋯	7	6	7	34
E_{20}	0	0	0	0	0	0	2	1	1	1	1	2	1	2	1	2	2	0	⋯	7	7	30
E_{21}	0	1	0	0	1	0	0	2	0	1	1	1	0	0	0	0	0	1	0	⋯	7	15
E_{22}	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	2	0	0	0	0	⋯	5

E. Methods of analysis

In this study intraelectrode and interelectrode comparisons were made at 199 and 398 pps with the purpose of determining single-rate and two-rate pitch structures. One method of investigating the results of the paired-comparison procedure is row sum analysis (David, 1988). When compared with d', row sum analysis can be used to obtain an estimate of the pitch structure without making the unidimensional assumption or assumptions about the perceptual variance within a given dimension (Collins *et al.*, 1997).

Row sum analysis requires that results from a paired comparison procedure be placed in a two-way preference table. Ties are not permitted in this paradigm, meaning subjects are forced to select one stimulus as higher in pitch than the other, and stimuli are never compared to themselves. For this reason, the diagonal of the table can always be set to a null value, and the lower triangle is redundant (David, 1988). The resulting score for each stimulus is determined by summing each element in the row corresponding to that stimulus. In this study, the score represents the number of times a stimulus was chosen as higher in pitch than all other stimuli in the set. A cumulative response matrix was constructed for each set of stimuli by combining the preference tables for each block of trials. The maximum value of any element in the cumulative response matrix is equal to the number of blocks, N. The maximum score, s_i, for one of t stimuli in a block, $A_i(i=1,2,\ldots,t)$, is equal to the number of stimuli in the block minus one (because stimuli are not compared to themselves), multiplied by the number of blocks, $s_{\max}=N(t-1)$. Table II is the cumulative response matrix obtained over seven blocks of the single-rate pitch ranking task for subject S7. Normalizing by the maximum possible score results in a value known as percent wins. This is the percentage of comparisons that resulted in a given stimulus being chosen as higher in pitch than all other stimuli in the set. This method of analysis is equivalent to that employed by Collins *et al.* (1997).

The first set of stimuli consists of only one rate per electrode; and therefore, the results of the row sum analysis should reflect the tonotopic ordering of the electrodes within the cochlea, ideally $s_i=N(t-i)$. Based on the knowledge that implant users' ability to identify a change saturates somewhere above 300 pps, row sum analysis results for the remaining three sets of stimuli, which contain two rates per electrode, should reflect a perceived increase in pitch on any given electrode when comparing 398 to 199 pps. It is possible that different pulse rates presented on adjacent or closely neighboring electrodes may result in an overlapping pitch percept.

Along with percent wins, it is possible to use the row sum analysis results to determine a coefficient of consistence for each preference table (David, 1988; Kendall and Smith, 1940). The coefficient of consistence, ζ, is unity if there are absolutely no inconsistencies in the configuration of the observed preferences, and ζ will approach zero as inconsistencies increase. Calculating ζ allows the researcher to gain insight into the subjects' individual abilities with respect to the paired comparison task under investigation and may also provide some indication that the stimuli are either indistinguishable under the given criteria, or that there may be multidimensional cues that are causing confusion during the task (Kendall and Smith, 1940). The coefficient of consistence, ζ, that accompanies each row sum analysis plot is the arithmetic mean of the coefficients of consistence for each repetition of the accompanying task.

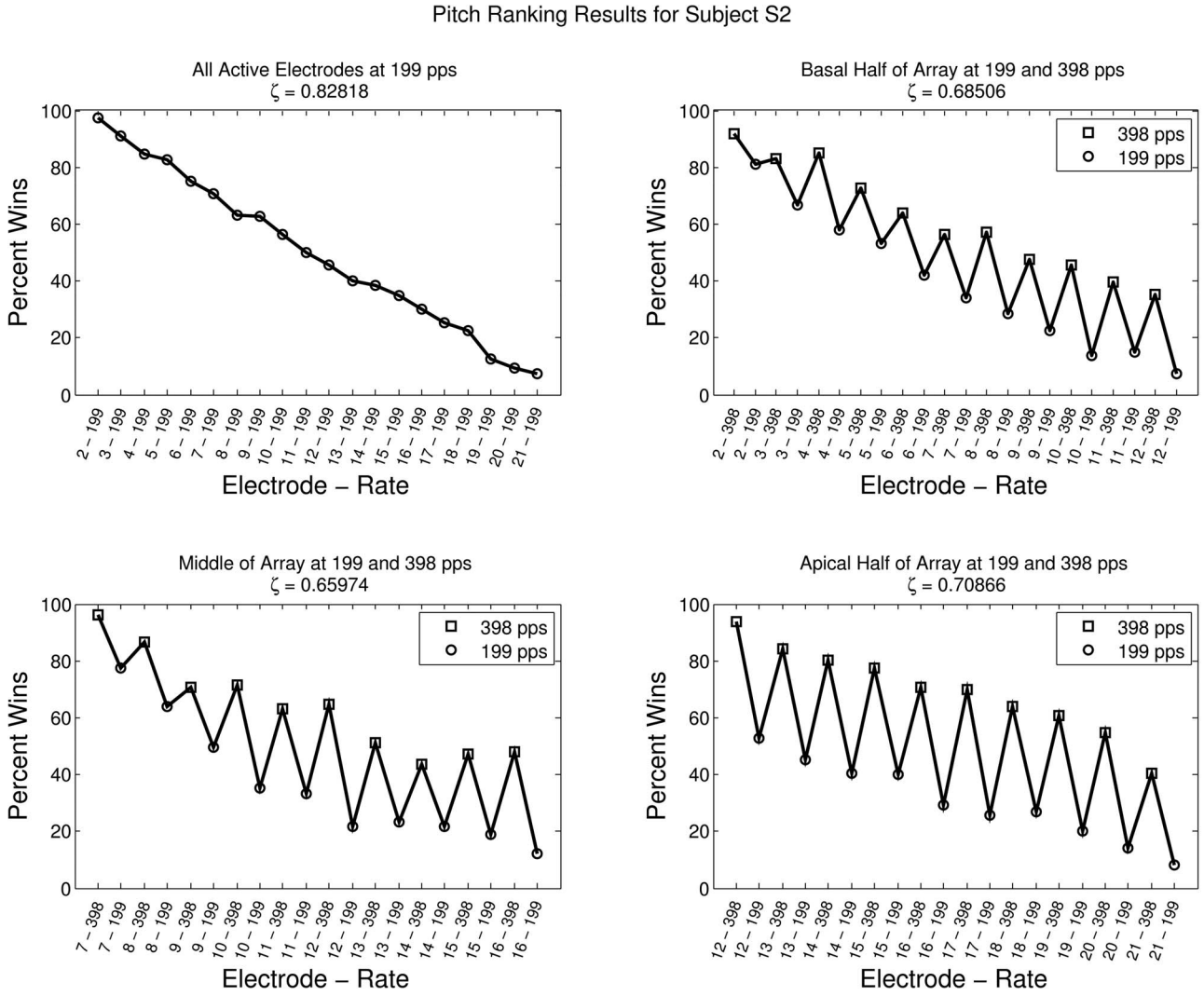


FIG. 1. Pitch ranking results for subject S2 as determined by row sum analysis. Here, percent wins are plotted vs electrode rate from most basal to most apical active electrode. Circles indicate a stimulation rate of 199 pps and square indicate a stimulation rate of 398 pps. The upper left plot contains the row sum analysis results from the single-rate pitch ranking task. The upper right, lower left, and lower right plots contain the basal, middle, and apical results of the two-rate pitch ranking task, respectively.

III. RESULTS

Figures 1–5 contain the results of the row sum analysis of the cumulative preference matrices for the single-rate experiment and the three subsets of stimuli used in the two-rate experiment, separated by subject. The electrode and rate corresponding to each stimulus are listed along the abscissa in the form "Electrode-Rate" from basal to apical electrode, and the ordinate is percent wins for each stimulus. The title of each plot indicates which of the four sets of stimuli were used to obtain the data represented in that plot as well as the average coefficient of consistence for the corresponding configuration of preferences. The upper left plot in each set of four contains the results from the single-rate pitch ranking experiment and includes all of the subject's active electrodes. The upper right, lower left, and lower right plots were derived from the set of stimuli containing two rates for the basal, middle, and apical sections of the array, respectively. In all plots, a circle represents a stimulus with a presentation rate of 199 pps and a square represents a stimulus with a presentation rate of 398 pps.

As can be seen in the single-rate row sum analysis results (upper left) of Figs. 1 and 5, subjects S2 and S7 have pitch structures that most closely resemble the tonotopic ordering of the cochlea. However, the single-rate results for subject S4 in Fig. 2 show a deviation from the ideal case, and Figs. 3 and 4 demonstrate the difficulty that subjects S5 and S6 had in identifying electrodes in an order that reflects that of the implanted electrode array. The intersubject variability of single-rate results can be seen in Fig. 6 and reflects the variability shown in previous pitch ranking studies (Collins *et al.*, 1997; Nelson *et al.*, 1995; Townshend *et al.*, 1987). The results from the two-rate pitch ranking tasks for subjects S2, S4, and S7 exhibit a clear zigzag behavior. This pattern indicates a consistent intraelectrode ranking of 398 pps over 199 pps along with the presence of some overlap across multiple electrodes. This pattern is less obvious in the two-rate results for subject S6 because interelectrode comparisons do not produce the expected downward trend when moving from base to apex. However, intraelectrode comparisons still tend to result in the subject indicating that 398 pps results in

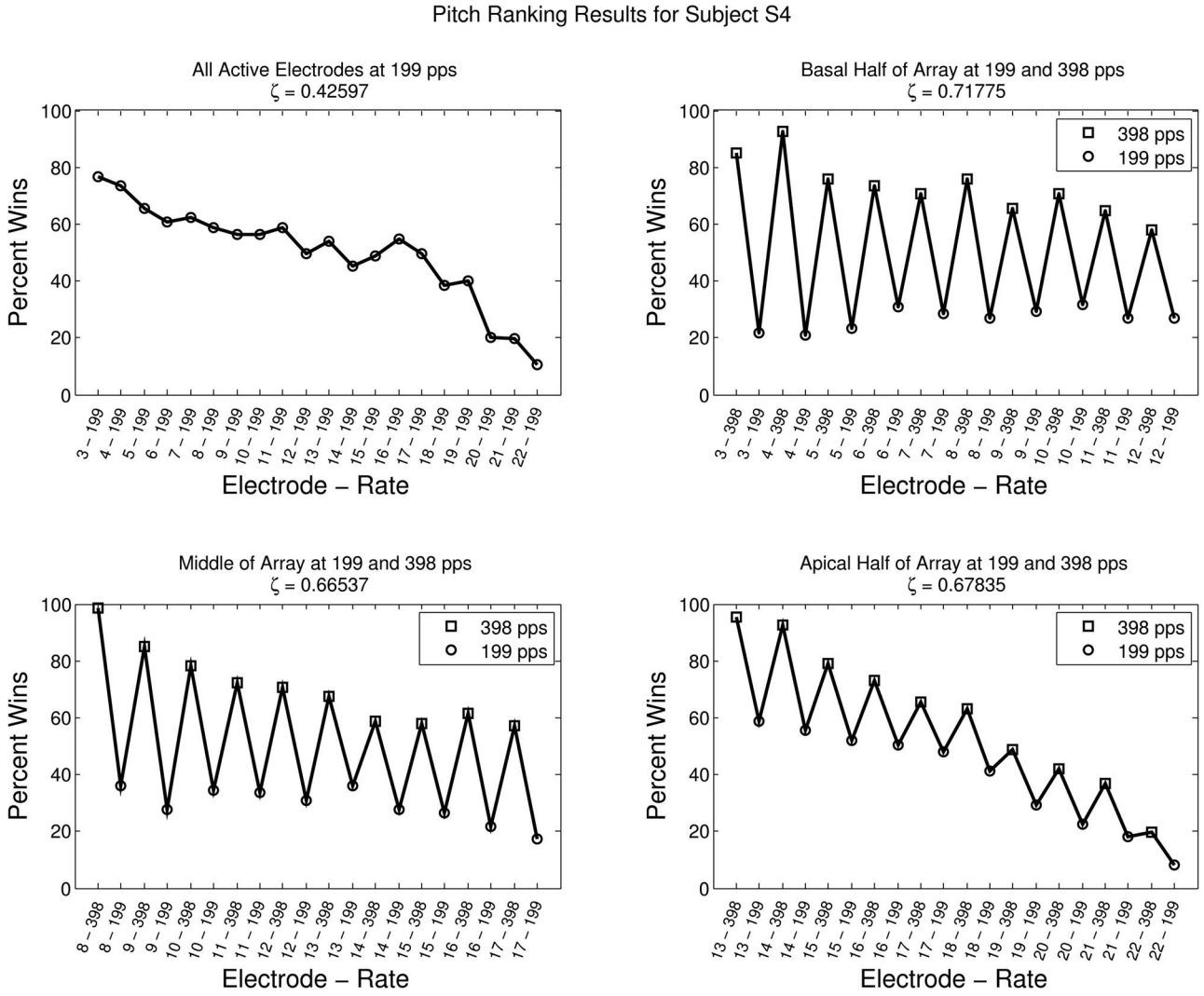


FIG. 2. Pitch ranking results for subject S4 as determined by row sum analysis. Here, percent wins are plotted vs electrode rate from most basal to most apical active electrode. Circles indicate a stimulation rate of 199 pps and squares indicate a stimulation rate of 398 pps. The upper left plot contains the row sum analysis results from the single-rate pitch ranking task. The upper right, lower left, and lower right plots contain the basal, middle, and apical results of the two-rate pitch ranking task, respectively.

a higher pitch percept that 199 pps. This does not seem to be the case with subject S5, whose two-rate pitch ranking results do not seem to indicate a clear ordering of pitch due to a change in rate or place of stimulation.

Notice that all coefficients of consistence obtained in the single-rate task range from 0.36 to 0.83, which falls within the range of values reported in Collins *et al.* (1997), 0.10–0.87. Those subjects whose judgments were most consistent also had pitch structures that most closely match the ideal case. Spearman's rank correlation coefficient was used to compare the rank of speech recognition scores to the rank of the coefficients of consistence, and there was no statistically significant relationship. In some subjects, there was a discrepancy between the single-rate and two-rate coefficients of consistence. When higher values of ζ were obtained for the two-rate task than the value of ζ obtained for the single-rate task (i.e., subject S4), it indicates that the introduction of rate-pitch resulted in a more consistent ranking of the stimuli, implying that the rate-pitch percept may be more dominant than the place-pitch percept. Conversely, obtaining a higher coefficient of consistence for the single-rate task with respect to the two-rate tasks may imply that the introduction of rate-pitch makes the pitch ranking task more difficult for some subjects.

IV. DISCUSSION

While multirate stimulation has been proposed as a possible method of improving speech perception, implementations by Fearn (2001) and Nobbe (2004) in Med-El and Nucleus cochlear implant speech processors, respectively, did not result in an improvement. Both Fearn (2001) and Nobbe (2004) did report some preference of multirate strategies when listening to music, implying that there may still be some benefit to varying the presentation rates on individual electrodes. As suggested by Throckmorton *et al.* (2006), it may be necessary to perform a battery of psychophysical experiments that allow a multirate algorithm to be optimized in order to achieve the gain in speech recognition predicted via acoustic models (Nie *et al.*, 2005).

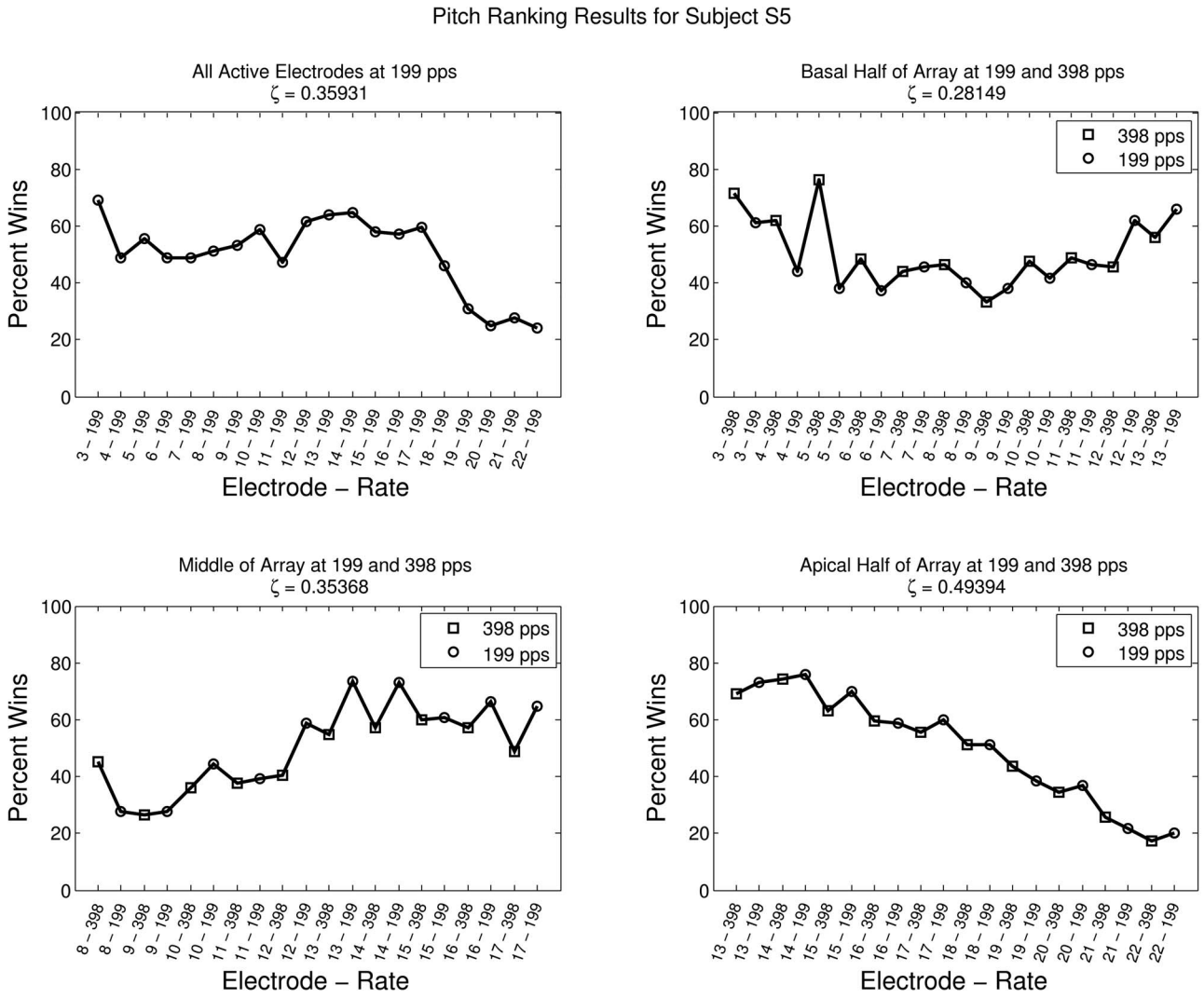


FIG. 3. Pitch ranking results for subject S5 as determined by row sum analysis. Here, percent wins are plotted vs electrode rate from most basal to most apical active electrode. Circles indicate a stimulation rate of 199 pps and squares indicate a stimulation rate of 398 pps. The upper left plot contains the row sum analysis results from the single-rate pitch ranking task. The upper right, lower left, and lower right plots contain the basal, middle, and apical results of the two-rate pitch ranking task, respectively.

The single-rate results found for the subjects in this study follow previous trends in the literature in terms of both pitch structures and variability across subjects (Collins *et al.*, 1997; Nelson *et al.*, 1995; Townshend *et al.*, 1987). The results from the two-rate task in this study indicate that using stimulation rates below the rate-pitch percept saturation point (typically 300–500 pps) often causes overlapping pitch percepts between electrodes. The results presented in this study are similar to the results of the pitch scaling task done by Eddington *et al.* (1978), and this may have contributed to the lack of improvement in speech recognition seen with previous multirate strategies. Fearn (2001) used overlapping rates on adjacent channels in an attempt to preserve more meaningful frequency information, and Nobbe (2004) used stimulation rates of 252 and 1515 pps. In both cases, the possibility of pitch reversals and anomalies was neglected, and as demonstrated here, overlapping percepts were likely present.

Examining the results from this study in combination with those found by McDermott and McKay (1997) seems to indicate that the rate-pitch percept may offer a broader range of pitch percepts than the place of stimulation alone. While McDermott and McKay (1997) found only a few, small overlapping pitch percepts between neighboring electrodes when rates were a semitone apart, this study demonstrates the substantial impact that rates with a separation of one octave can have on the overall pitch structure. A future study attempting to find the relationship between two rates on neighboring electrodes that results in a pitch reversal may provide more insight into the optimal range of rates for use in a multirate strategy.

The variability across subjects seen in both single-rate and two-rate results supports the hypothesis put forth by Throckmorton *et al.* (2006), which states that tuning algorithms to each user may be required to obtain the maximum benefit from a multirate strategy. The results obtained in this study imply that when using a multirate strategy with fixed stimulation rates, it may be beneficial to include a patient-specific mapping from filter outputs to rate-electrode combinations in order to preserve the desired monotonic pitch structure. While beginning with an ordered pitch structure may have

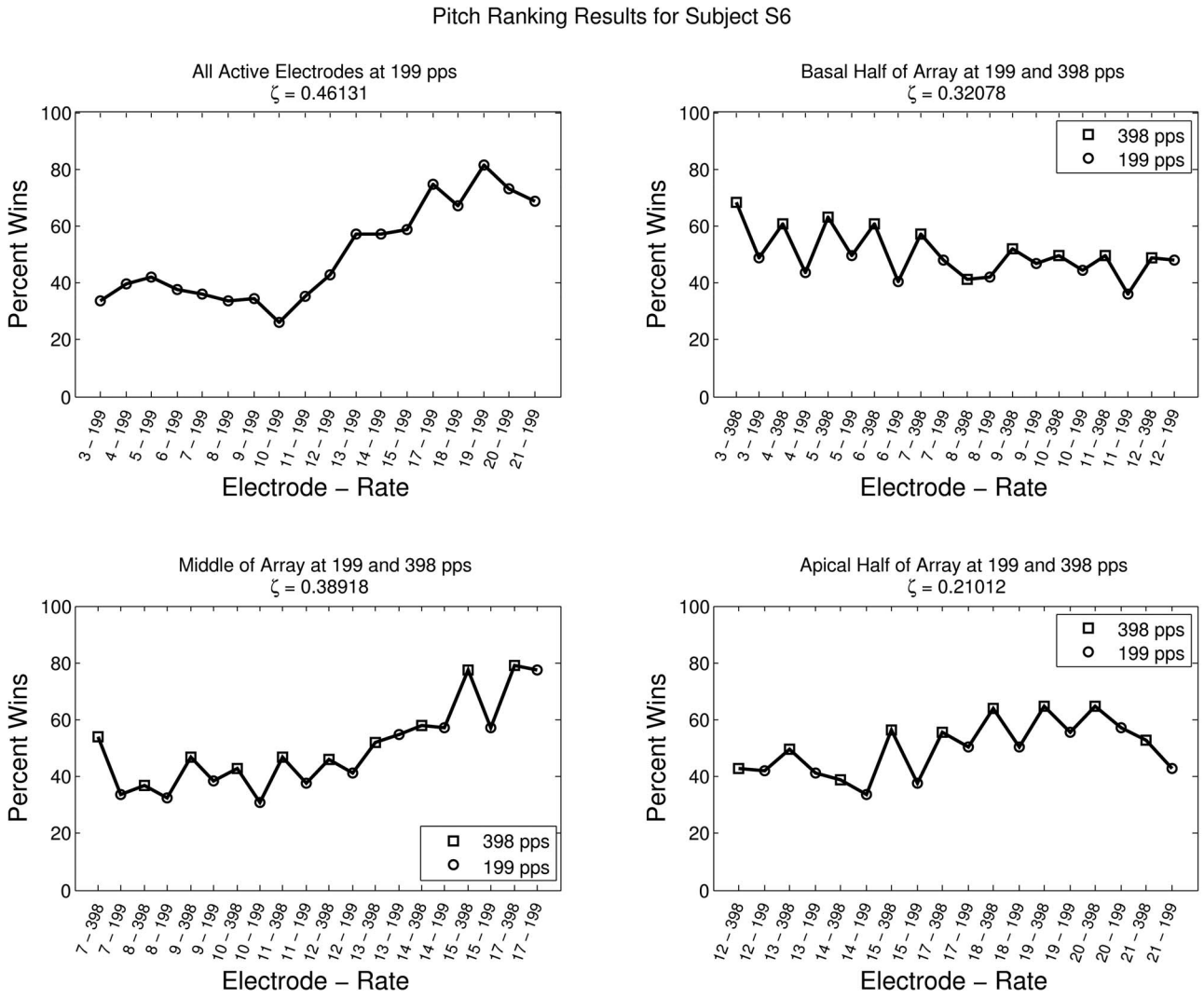


FIG. 4. Pitch ranking results for subject S6 as determined by row sum analysis. Here, percent wins are plotted vs electrode rate from most basal to most apical active electrode. Circles indicate a stimulation rate of 199 pps and squares indicate a stimulation rate of 398 pps. The upper left plot contains the row sum analysis results from the single-rate pitch ranking task. The upper right, lower left, and lower right plots contain the basal, middle, and apical results of the two-rate pitch ranking task, respectively.

had some impact on those results obtained by Fearn (2001) and Nobbe (2004), this type of reordering did not prove to be consistently beneficial in the study done by Collins *et al.* (1997). The hypothesis for this is that subjects were not able to instantly adapt to the corrected order, and that training may be required to see the desired improvement from a perceptual reordering of familiar stimuli.

In order to take advantage of the rate-pitch percept, rates less than 500 pps may be most appropriate for stimulation. There may be some concern about the loss of temporal information due to the use of relatively low pulse rates in a multirate strategy, and a number of studies have investigated the effect of different stimulation rates on speech recognition (Brill *et al.*, 1997; Fu and Shannon, 2000; Loizou *et al.*, 2000; Vandali *et al.*, 2000; Wilson *et al.*, 1993). Testing six cochlear implant users, Loizou *et al.* (2000) found that, on average, word and consonant recognition scores obtained in quiet at a stimulation rate of 2100 pps were significantly higher than those obtained at rates of 400 or 800 pps. Wilson *et al.* (1993) and Brill *et al.* (1997) also found the potential for improvement at higher stimulation rates. However, Fu and Shannon (2000) found no significant difference in vowel and consonant recognition performance when using rates from 150 to 500 pps, and Vandali *et al.* (2000) also found that an increase in stimulation rate from 250 to 1615 pps provided no significant improvement in speech recognition in quiet or in noisy conditions. To date, no definitive study exists that demonstrates a global improvement in performance due to relatively high rates of stimulation, and so it remains possible that the loss of temporal resolution required to implement a multirate strategy may not have a significant negative impact on user performance.

In examining the results from this study, the coefficients of consistence (ζ) appear to be rather low with respect to unity. Reasons for lower values of ζ may include an inability to make a reasonable distinction between one or more pair of stimuli, or it may be a result of confusion due to the fact that cues may be present in multiple dimensions, only one of which is pitch. In some cases, subjects S5 and S6 for example, the low values of ζ may simply be due to difficulty in

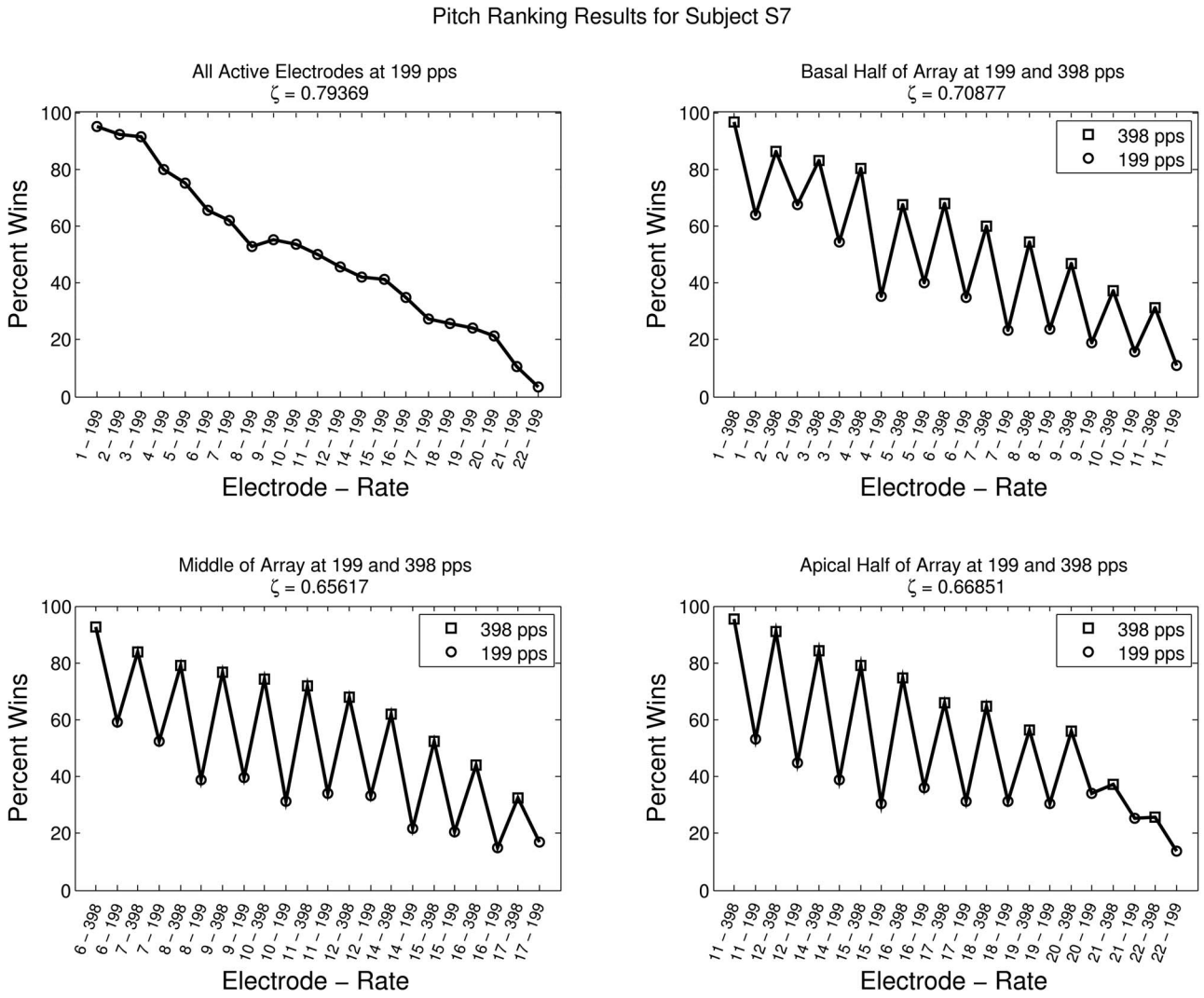


FIG. 5. Pitch ranking results for subject S7 as determined by row sum analysis. Here, percent wins are plotted vs electrode rate from most basal to most apical active electrode. Circles indicate a stimulation rate of 199 pps and squares indicate a stimulation rate of 398 pps. The upper left plot contains the row sum analysis results from the single-rate pitch ranking task. The upper right, lower left, and lower right plots contain the basal, middle, and apical results of the two-rate pitch ranking task, respectively.

discriminating between different electrodes and/or pulse rates. Incidentally, subject S5's length of profound deafness was longer than the other four subjects in this study, and it has been shown that duration of deafness may be a predictor for speech recognition abilities (Blamey *et al.*, 1992; van Dijk *et al.*, 1999). Subject S5 also suffers from tinnitus, and at times that condition caused confounding pitch cues and difficulty with pitch tasks. Upon examining the pitch ranking results for subject S6, it appears that there may have been some confusion about the pitch ranking task. That is, it is possible that subject S6 selected the lower pitch when making a paired comparison, and it is possible that the low coefficient of consistence indicates an alternation between selecting the higher and lower pitch between trials. It should be noted that subject S2 experienced a gradual onset of deafness beginning at the age listed in Table I, and the age at which profound deafness occurred was not clear. The fact that ζ was not always greater for either the single-rate or two-rate task may offer some insight into the dominance of the rate or place-pitch percept in each user. For example, subject S4 had

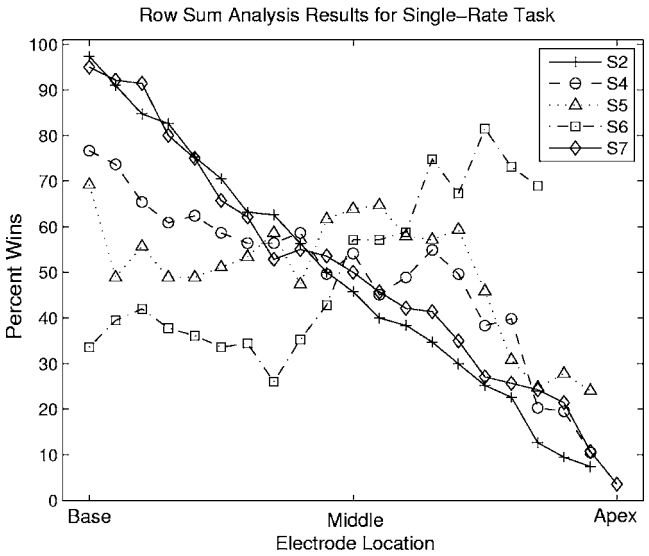


FIG. 6. Single-rate pitch ranking results for all subjects as determined by row sum analysis. Here, percent wins are plotted vs electrode location from base to apex. The large intersubject variability in the single-rate results is shown here.

consistently higher values of ζ for the two-rate task, implying that it may be easier to compare different rates within and across electrodes than to compare the same rate (199 pps) across electrodes. Fatigue may also play some role in decreasing the value of ζ, but the most relevant cause may be the possibility that subjects were confronted with a multidimensional percept when asked to select the higher pitch from a number of rate and electrode combinations.

The concept of multidimensional analysis was applied to multiple electrode configurations by Tong *et al.* (1983) and McKay *et al.* (1996), and while McKay *et al.* (1996) included some single electrode stimuli, Collins and Throckmorton (2000) extended that analysis and investigated the possible influence of multiple dimensions on single electrode judgments across each user's entire array. In that study it was found that there may be two significant perceptual dimensions that arise from a change in the location of stimulation, and it was hypothesized that the rate-pitch percept and its interaction with the place of stimulation may play a part in this multidimensionality. This potentially nonlinear interaction between the rate and place of stimulation may need a great deal of attention when designing a multirate sound processing strategy. It is possible that multidimensional percepts could be used to increase the spectral palette available to implant users; however, if the second dimension and its relationship to the first dimension (thought to be pitch) are not well understood, this could prove to be a confounding factor in multirate stimulation. A multidimensional scaling procedure applied to the same or a similar set of multirate stimuli could offer more insight into the behavior of the second dimension, both independently and with respect to the first, as well as offer further understanding into the results of the two-rate pitch ranking results of this study.

V. CONCLUSIONS

The spectral information currently available via cochlear implants is insufficient to allow users to accomplish more difficult tasks such as speech recognition in noise and melody recognition, particularly in challenging listening environments. While using variable stimulation rates on each electrode to increase the number of available stimuli may increase the number of possible percepts, these percepts do not necessarily follow an orderly, predictable pattern. For this reason, it may be necessary to perform rate-based psychophysics in order to better understand the pitch structure for a given user before attempting to implement a complex sound processing strategy. This study confirmed that an increase in rate on a single electrode from 199 to 398 pps generally causes an increase in the perceived pitch, but that when comparing these two rates across different electrodes, pitch anomalies occur frequently. Based on this finding, it may be beneficial to determine a user's individual pitch structure and tailor a multirate strategy to that pitch structure prior to implementation of the algorithm.

ACKNOWLEDGMENTS

The authors would like to thank the subjects that participated in this study for their time and patience. This work was supported by the National Institutes of Health 1-R01-DC007994-01.

Blamey, P., Pyman, B., Gordon, M., Clark, G., Brown, A., Dowell, R., and Hollow, R. (**1992**). "Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients," Ann. Otol. Rhinol. Laryngol. **101**, 342–348.

Brill, S., Gstottner, W., Helms, J., von Ilberg, C., Baumgartner, W., Muller, J., and Kiefer, J. (**1997**). "Optimization of channel number and stimulation rate for the fast continuous interleave sampling strategy in the combi 40 +," Am. J. Otol. **18**, S104–S106.

Collins, L. M., and Throckmorton, C. S. (**2000**). "Investigating perceptual features of electrode stimulation via a multidimensional scaling paradigm," J. Acoust. Soc. Am. **108**, 2353–2365.

Collins, L. M., Zwolan, T. A., and Wakefiled, G. H. (**1997**). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **101**, 440–455.

David, H. (**1988**). *The Method of Paired Comparisons* (Oxford University Press, New York).

Donaldson, G. S., Kreft, H. A., and Litvak, L. (**2005**). "Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users (1)," J. Acoust. Soc. Am. **118**, 623–626.

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.

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**, 2403–2410.

Eddington, D., Dobelle, W., Brackmann, D., Mladejovsky, M., and Parkin, J. (**1978**). "Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers," Trans. Am. Soc. Artif. Intern. Organs **24**, 1–5.

Fearn, R. (**2001**). "Music and pitch perception of cochlear implant recipients," Ph.D. thesis, University of New South Wales, New South Wales, Australia.

Fearn, R., Carter, P., and Wolfe, J. (**1999**). "The dependence of pitch perception on the rate and place of stimulation of the cochlea: A study using cochlear implants," Acoust. Aust. **27**, 41–43.

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: Comparisons of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589–597.

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.

Gfeller, K., Christ, A., Knutson, J., Witt, S., and Mehr, M. (**2003**). "The effects of familiarity and complexity on appraisal of complex songs by cochlear implant recipients and normal-hearing adults," J. Music Ther. **40**, 78–112.

Kendall, M., and Smith, B. B. (**1940**). "On the method of paired comparisons," Biometrika **31**, 324–345.

Koch, D. B., Downing, M., Osberger, M. J., and Litvak, L. (**2007**). "Using current steering to increase spectral resolution in cii and hires 90 k users," Ear Hear. **28**, 38S–41S.

Loizou, P. C., Poroy, O., and Dorman, M. (**2000**). "The effect of parametric variations of cochlear implant processors on speech understanding," J. Acoust. Soc. Am. **108**, 790–802.

McDermott, H. J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.

McDermott, H. J., and McKay, C. M. (**1994**). "Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea," J. Acoust. Soc. Am. **96**, 155–162.

McDermott, H. J., and McKay, C. M. (**1997**). "Musical pitch perception with electrical stimulation of the cochlea," J. Acoust. Soc. Am. **101**, 1622–1630.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1996**). "The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees," J. Acoust. Soc. Am. **99**, 1079–1090.

Nelson, D. A., Tasell, D. J. V., Schroder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.

Nie, K., Stickney, G., and Zeng, F.-G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Nobbe, A. (**2004**). "Pitch perception and signal processing in electric hearing," Ph.D. thesis, Munich University, Munich, Germany.

Pijl, S., and Schwartz, D. W. F. (**1995**). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," J. Acoust. Soc. Am. **96**, 886–895.

Shannon, R. V. (**1983a**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.

Stohl, J. S., Kucukoglu, M. S., Throckmorton, C. S., and Collins, L. M. (**2007**). "Developing a spear3-based experimental psychophysics environment," in Abstracts of the 30th Annual Midwinter Research Meeting: Association for Research in Otolaryngology, Denver, CO, p. 159.

Throckmorton, C. S., and Collins, L. M. (**2001**). "A comparison of two loudness balancing tasks in cochlear implant subjects using bipolar stimulation," Ear Hear. **22**, 439–448.

Throckmorton, C. S., Kucukoglu, M. S., Remus, J. J., and Collins, L. M. (**2006**). "Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: Implications for cochlear implants," Hear. Res. **218**, 30–42.

Tong, Y., and Clark, G. (**1985**). "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients," J. Acoust. Soc. Am. **77**, 1881–1888.

Tong, Y., Dowll, R., Blamey, P., and Clark, G. (**1983**). "Two-component hearing sensations produced by two-electrode stimulation in the cochlea of a deaf patient," Sci., New Ser. **219**, 993–994.

Townshend, B., Cotter, N., Compernolle, D. V., and White, R. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–114.

van Dijk, J., van Olphen, A., Langereis, M., Mens, L., Brokx, J., and Smoorenburg, G. (**1999**). "Predictors of cochlear implant performance," Audiology **38**, 109–116.

Vandali, A., Whitford, L., Plant, K., and Clark, G. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.

Wilson, B. S., Lawson, D. T., and Zerbi, M. (**1993**). "Speech processors for auditory prostheses," Technical Report, NIH Project N01-DC-2-2401, Fifth Quarterly Progress Report.

Zeng, F.-G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech reognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.

Across-site patterns of modulation detection in listeners with cochlear implants[a)]

Bryan E. Pfingst[b)]
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109-5616

Rose A. Burkholder-Juhasz
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109-5616

Li Xu
School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701 and Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109-5616

Catherine S. Thompson
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109-5616

(Received 16 April 2007; revised 15 October 2007; accepted 21 November 2007)

In modern cochlear implants, much of the information required for recognition of important sounds is conveyed by temporal modulation of the charge per phase in interleaved trains of electrical pulses. In this study, modulation detection thresholds (MDTs) were used to assess listeners' abilities to detect sinusoidal modulation of charge per phase at each available stimulation site in their 22-electrode implants. Fourteen subjects were tested. MDTs were found to be highly variable across stimulation sites in most listeners. The across-site patterns of MDTs differed considerably from subject to subject. The subject-specific patterns of across-site variability of MDTs suggest that peripheral site-specific characteristics, such as electrode placement and the number and condition of surviving neurons, play a primary role in determining modulation sensitivity. Across-site patterns of detection thresholds (*T* levels), maximum comfortable loudness levels (*C* levels) and dynamic ranges (DRs) were not consistently correlated with across-site patterns of MDTs within subjects, indicating that the mechanisms underlying across-site variation in these measures differed from those underlying across-site variation in MDTs. MDTs sampled from multiple sites in a listener's electrode array might be useful for diagnosing across-subject differences in speech recognition with cochlear implants and for guiding strategies to improve the individual's perception.
 [DOI: 10.1121/1.2828051]

PACS number(s): 43.66.Ts, 43.66.Mk [JHG] Pages: 1054–1062

I. INTRODUCTION

Most modern auditory prostheses use both spectral cues mapped to electrode place and temporal cues mapped to the envelope of the electrical signal. Typically, the acoustic signal is received by a microphone and divided into a number of channels using bandpass filters. The temporal envelopes of the filtered signals are then extracted and used to amplitude modulate the charge per phase of pulses in continuous-interleaved pulse trains. These trains of charge-modulated pulses are then sent to individual electrodes arranged along the tonotopic axis of the cochlea or a central auditory nucleus. Spectral resolution with current auditory prostheses is limited, so listeners are strongly dependent on the temporal-envelope information in the electrical signal. Listeners can typically achieve reasonable speech recognition with as few as four spectral channels as long as there is sufficient temporal-envelope information (Shannon *et al.*, 1995; Xu *et al.*, 2005).

Temporal modulation patterns are important for the perception of voicing and manner of consonants (Van Tasell *et al.*, 1987), recognition of lexical tone (Xu *et al.*, 2002), and for sound-source segregation (Bregman *et al.*, 1990; Chatterjee *et al.*, 2006). Thus, improvements in cochlear implant users' abilities to detect temporal modulation might lead to improved perception of speech and other important auditory signals.

Modulation detection thresholds (MDTs) are a useful measure of the listener's acuity for detection of temporal modulations. They can be used to assess the listener's ability to detect relatively slow modulations in the charge per phase of electrical pulse trains, like those in the temporal-envelope encoding of speech information in an auditory prosthesis.

[a)]Initial reports of these data were presented at the Fourth Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI.
[b)]Author to whom correspondence should be addressed. Electronic mail: bpfingst@umich.edu

MDTs have been found to correlate with speech recognition across listeners using cochlear and brainstem implants (Fu, 2002; Colletti and Shannon, 2005). These studies have demonstrated considerable across-listener variability in MDTs. In comparing MDTs in two populations of patients with auditory brainstem implants, Colletti and Shannon (2005) found that a population of patients with excised acoustic tumors had higher MDTs than patients who had no acoustic tumor. They concluded that the tumor patients had higher MDTs because neural pathways important for modulation detection were disrupted.

Although these studies examining modulation detection in listeners with auditory prostheses provided informative results pointing to possible causes of across-listener variability, they did not investigate within-listener variability of MDTs across stimulation sites in the electrode array. Cochlear implants typically have up to 22 stimulation sites positioned along the length of the scala tympani. Each of these sites could feasibly differ in sensitivity to modulation. Fu's (2002) examination of modulation detection used only one stimulation site located near the center of the electrode array. In a recent report, we tested MDTs at three sites spaced evenly in the basal, middle, and apical parts of the electrode array and found that there was variation in MDTs across the three sites (Pfingst *et al.*, 2007). In the current study, we explored this variation in greater detail by measuring MDTs at all available sites along the electrode array at two listening levels.

Examining across-site variability in MDTs is important because it could provide clues about what mechanisms underlie modulation detection in cochlear implant users and this in turn could guide rehabilitation strategies. First, it is important to know if the origins of differences in MDTs across various cochlear implant users are peripheral or central. That is, do MDTs reflect the general ability of the listener to detect modulation or does modulation detection depend on conditions near the individual stimulation sites? We hypothesize that a listener's ability to detect modulation in the electrical signal depends on peripheral conditions in the implanted listener's cochlea. Such conditions might include the nerve survival pattern near each electrode and/or the location of each electrode with respect to the modiolus. Conditions in the scala tympani, such as bone and tissue growth, might also affect the current path from the electrodes to the excitable neural elements. Each of these variables could affect the number of fibers activated and the sites of activation and thus they could affect the perception of the signal. These conditions are known to vary along the length of the implanted cochlea in a deaf ear, and the pattern of this variation differs from person to person (Hinojosa and Lindsay, 1980; Nadol, 1997; Saunders *et al.*, 2002). We reason that if these conditions affect modulation detection, then we should find variation in MDTs across the individual stimulation sites within most listeners. Alternatively, if modulation detection simply relates to a listener's general ability to perceive modulated signals, then we would expect a more uniform across-site pattern of MDTs within listeners. Further, it is important to know, assuming the conditions determining MDTs are peripheral in origin, the degree to which these conditions are localized with respect to the individual stimulation sites in the cochlear implant electrode array.

Previous studies conducted in our laboratory demonstrated across-site variability in detection thresholds (*T* levels), maximum comfortable loudness levels (*C* levels), and dynamic ranges (DRs) (Pfingst *et al.*, 2004; Pfingst and Xu, 2004, 2005). Given this result, it is reasonable to expect that other psychophysical measures would also vary across stimulation sites in most cochlear implant users. If across-site variation in MDTs is found, it is important to know whether or not the across-site patterns of variation in MDTs match the across-site patterns of *T* levels, *C* levels, and DRs. Examining the relationship between *T* levels, *C* levels, and DRs could tell us whether or not the across-site patterns are due to common underlying mechanisms.

In addition, it would be clinically useful to know if the across-site patterns of MDTs and these other psychophysical measures are similar. *T* and *C* levels are routinely collected before programming a patient's processor and DRs are derived from these measures. Thus, if *T* levels, *C* levels, or DRs show the same across-site pattern as MDTs, they could serve as a more clinically convenient and less time-consuming tool for identifying stimulation sites that are weak in modulation detection.

Knowledge about the causes of variation in modulation detection and the relationship of modulation detection to other psychophysical measures is also important for the design of clinical rehabilitation strategies. If the variation in MDTs is due to peripheral physiological and anatomical variables, then it might be appropriate to design rehabilitation strategies on an individual-electrode basis. Sites that show weaknesses in modulation detection might be ineffective or even distracting during speech perception tasks. Rehabilitation strategies could involve deactivating electrodes identified as having high MDTs or adjusting the stimulation parameters (e.g., pulse rate, electrode configuration, etc.) at a given stimulation site based on parameters that are most conducive to detecting modulation at that site. On the other hand, if poor modulation detection abilities are caused by general perceptual deficits, perceptual training might be a more effective method of lowering MDTs.

In summary, across-site patterns of MDTs might reflect the mechanisms that underlie modulation detection abilities in cochlear implant users and thus inform clinicians as to the best approaches for rehabilitation. Highly variable across-site MDTs would suggest that site-specific characteristics contribute to modulation detection and direct the focus to site-specific treatments. Alternatively, if listeners have poor MDTs that vary minimally across sites, a more general deficit in recognizing temporal modulations might be present, which would suggest using a training procedure to improve the use of available cues. Thus, the knowledge gained from this study might be useful in developing clinical rehabilitation strategies to help improve temporal-envelope processing, which could lead to more accurate consonant recognition and overall improvements in perception with cochlear implants.

In this study, a 40 Hz modulation frequency was used to modulate the phase duration of pulses in a constant-rate

TABLE I. Demographic and clinical characteristics of the subjects used in the current study.

ID	Gender	Age at onset of deafness (years)	Age at testing (years)	Duration of deafness before implantation (years)	Duration of CI use	Implant	Processing strategy	Etiology of deafness	Sites tested
S45	M	44	50	1	5	CI24R(CS)	ACE	Head injury	3-22
S53	F	26	58	26	6	CI24R(CS)	ACE	Unknown	3-22
S54	F	32	71	35	4	CI24R(CS)	ACE	Ototoxicity	3, 5-22
S57	F	60	68	3	5	CI24M	SPEAK	Fever/Ménière's	3-21
S58	F	42	57	6	9	CI24M	SPEAK	Autoimmune	3-22
S60	M	62	67	2	3	CI24R(CS)	ACE	Hereditary	3-22
S64	F	50	59	1	8	CI24M	SPEAK	Unknown	3-22
S65	F	35	38	1	2	CI24R(CS)	ACE	Hereditary	2-13, 15-22
S67	M	59	65	<1	6	CI24R(CS)	ACE	Hereditary	1-5, 7-21
S71	M	57	60	<1	3	CI24R(CS)	ACE	Unknown	2-7, 9-22
S72	F	16	66	47	3	CI24R(CS)	ACE	Enlarged vestibular aqueduct	5-13, 15-22
S73	M	50	64	12	2	CI24R(CS)	ACE	Unknown	4-22
S74	F	61	66	2	3	CI24R(CS)	ACE	Unknown	3-22
S75	M	53	58	2	3	CI24R(CS)	ACE	Noise induced	2-21

pulse train. This frequency was chosen as an intermediate value in the range of modulation frequencies important for cochlear implant users. For English phoneme recognition, the most important temporal envelope information occurs in the frequency region below 50 Hz (Drullman *et al.*, 1994a, b; Shannon *et al.*, 1995; Fu and Shannon, 2000; Xu *et al.*, 2005), whereas higher-frequency cues (50–500 Hz) have been found to benefit lexical-tone recognition (Fu *et al.*, 1998; Xu *et al.*, 2002) and voice gender recognition (Fu *et al.*, 2004). Most current cochlear prosthesis speech processors provide temporal envelope cues up to about 200–400 Hz (Wilson, 2004). However, modulation detection in cochlear implant users typically begins to decline above about 100 Hz modulation frequency and most subjects cannot detect modulation above 300 Hz (Shannon, 1992).

II. METHOD

A. Subjects

Fourteen postlingually deafened adults fitted with Nucleus cochlear implants participated in the study. Eleven of the subjects used the Nucleus CI24R (Contour) array, and three of the subjects used the CI24M (straight) array. All subjects were native speakers of American English and had at least two years (mean of 4.4 years) of experience with their device before beginning the experiment. Table I lists these and other demographic and clinical characteristics of the subjects. The use of human subjects for this research was approved by the University of Michigan Medical School Institutional Review Board.

B. Hardware and software for electrical stimulation

The listeners completed psychophysical tests wearing a laboratory-owned SPrint processor (Serial Number 408594, Cochlear Corporation, Englewood, CO) connected to a Processor Control Interface (Cochlear Corporation). The input to the processor was generated through the Nucleus Implant Communicator software libraries (version 3.27) and an IF5 ISA card. Listeners' own implanted receiver/stimulators received radio frequency pulses generated by the processor, and these pulses were transmitted as electrical current to the appropriate sites in the implanted 22-electrode array. The calibration value of each listener's receiver/stimulator was obtained from Cochlear Corporation and used to calculate the stimulation levels in peak microamperes. These levels were then converted to decibels of current using the formula

$$\text{level (dB re 1000 } \mu\text{A)} = 20 \log(x/1000\ \mu\text{A}),$$

where x is the level in microamperes.

C. Identification of testable sites

Prior to testing, each listener's map was evaluated using CustomSound software (Cochlear Corporation) to determine which sites could be tested. Sites were excluded from testing if they were not used in the listener's regular clinically programmed map. Stimulation sites are commonly turned off during mapping if their stimulation results in uncomfortable or non-auditory sensations or if there is an electrical short. Stimulation sites used for each listener are listed in Table I.

D. Psychophysical testing

Listeners completed psychophysical tests to determine thresholds (T levels), comfort levels (C levels), and MDTs at all available sites in the electrode array.

T and C levels were obtained using symmetric-biphasic pulses of 50 μs/phase with an 8 μs interphase gap and a pulse rate of 250 pulses/s (pps). The stimulus burst duration was 600 ms presented in an on/off duty cycle with a 600 ms interburst interval. Monopolar stimulation (MP1+2) was used in all cases. In this configuration current is passed between a single electrode in the scala-tympani implant and two connected electrodes outside the cochlea.

Listeners used the method of adjustment to set T and C levels. Each trial started with the initiation of the on/off cy-

cling of the stimulus. To record the *T* level, listeners were instructed to adjust the level of the signal up or down until it was "just barely audible." Adjustments were made by using the mouse to click on large and small boxes on the computer screen representing 5 Cochlear Level Unit and 1 Cochlear Level Unit increases and decreases (where 1 Cochlear Level Unit equals 0.176 dB of current). Listeners recorded their *T* level by clicking on a button when they were satisfied with the level they reached. Once the *T* level was recorded, listeners began increasing the stimulus level until the *C* level was reached. Listeners were instructed to record a *C* level when they reached a level that was "the loudest they could listen to comfortably for an extended period of time."

The stimuli used for the modulation detection task were symmetric-biphasic pulses with a mean pulse duration of 50 μs/phase and an interphase gap of 8 μs. The pulse rate was 250 pps and stimulus duration was 600 ms. The interstimulus duration was 600 ms. Monopolar stimulation (MP1+2) was used in all cases. The phase duration of the pulses was modulated by a 40 Hz sinusoid which started and ended at zero phase. The positive and negative phases of the pulses were modulated equally to maintain charge balance while the interphase gap was held constant.

The modulation index (m) was defined as:

$$m = (\mathrm{PD}_{\max} - \mathrm{PD}_{\min})/(\mathrm{PD}_{\max} + \mathrm{PD}_{\min}),$$

where $\mathrm{PD}_{\max}$ and $\mathrm{PD}_{\min}$ are the maximum and minimum phase durations, respectively. We report modulation values in decibels (dB) re 100% modulation (i.e., $20 \log m$). In the results section, these values are plotted with the lowest values (most sensitive modulation thresholds) at the top of the ordinate and/or the right of the abscissa.

Phase duration modulation rather than amplitude modulation was used for these experiments because the limits of the implanted stimulators allowed finer control of charge per phase when phase duration was modulated compared to when amplitude was modulated. The smallest step size in phase duration available with these stimulators was 0.2 μs. Thus, with a mean phase duration of 50 μs the smallest achievable modulation was 0.4% or −47.96 dB re 100% modulation.

Listeners' *T* and *C* levels were used to determine their DRs for each site and to set the levels of the stimuli for the modulation detection task. MDTs were measured at 30% and 70% of DR in decibels of current. Fu (2002) has suggested that measurement of MDTs at multiple levels is necessary to adequately characterize the listener's modulation-detection ability. The decision to measure MDTs at only two levels was based on an analysis of data from a previous study (Pfingst *et al.*, 2007). In that study, MDTs were measured in 12 cochlear implant users at five stimulus levels (10%, 30%, 50%, 70%, and 90% of DR in decibels of current) and three stimulation sites: one basal, one middle, and one apical site. MDTs were measured for two carrier rates: 250 pps and 4 kpps. For this analysis, we used the data for the 250 pps carrier to match the rate used in the current study. A bivariate, across-listeners, correlation of mean MDTs taken at two levels and mean MDTs taken at five levels was significant at all three sites ($r=0.99$, $p<0.001$ in each case). Based on this analysis, we concluded that MDTs could be adequately characterized by measuring MDTs at only two levels: 30% and 70% of DR.

MDTs were obtained using a two-interval, forced-choice paradigm with flanking cues. On each trial, listeners were presented with four sequential observation intervals marked by squares on the computer screen. These squares were illuminated in sequence as the electrical stimuli were presented to the implant. The interstimulus interval was 600 ms. The first and fourth interval contained identical unmodulated pulse trains which served as flanking cues. One of the other intervals (interval 2 or interval 3, chosen at random on each trial) also contained this unmodulated signal. The modulated pulse train occurred in the remaining interval. Listeners were instructed to choose the interval (interval 2 or interval 3) containing the stimulus that sounded different from the other three. Selections were made by using the computer mouse to click on the desired square.

A two-down, one-up adaptive procedure (Levitt, 1971) was used, starting with a modulation depth of 50% and ending when 14 reversals were recorded. Modulation depth was increased or decreased in steps of 6 dB to the first reversal, 2 dB for the next two reversals, and 1 dB for the next 10 reversals. The MDT was defined as the mean of the levels at the last 8 reversal points. MDTs were measured in each listener at all available stimulation sites and at both stimulus levels (30% and 70% of DR) two times in random order. If the difference between the two estimates obtained for a condition exceeded 7 dB, a third estimate of MDT was made for that condition and the outlier was discarded (see Pfingst *et al.*, 2007). A third measurement was required for only 5.9% of the total measurements.

III. RESULTS

A. Across-site patterns of MDTs

There was considerable across-subject and across-site (within-subject) variability in MDTs. Further, the pattern of variation across sites differed markedly from subject to subject. Figure 1 shows the across-site patterns for the 14 subjects. The across-site variances of each listener's MDTs for each of the two levels tested are shown in the lower right-hand or upper left-hand corner of each panel. The top number is the across-site variance in square decibels for MDTs at 70% of DR and the number below is the across-site variance for MDTs at 30% of DR. The subjects demonstrated a wide range of variability in MDTs measured at 30% of DR (σ =4.59–114.97 dB^2) and 70% of DR (σ=0.23–70.30 dB^2). Many listeners had near minimal MDTs at the majority of sites when the stimulus level was 70% of DR (M= −34.93 dB re 100% modulation, SD=7.62) and thus the across-site variance was lower at this level. MDTs were higher (M=−24.91 dB, SD=7.73) and more variable at 30% of DR.

Across-site variation in MDTs could be due to localized variation in physiological and biophysical variables that result from subject-specific patterns of pathology along the length of the cochlea. However, systematic normal physiological differences along the length of the cochlea, system-

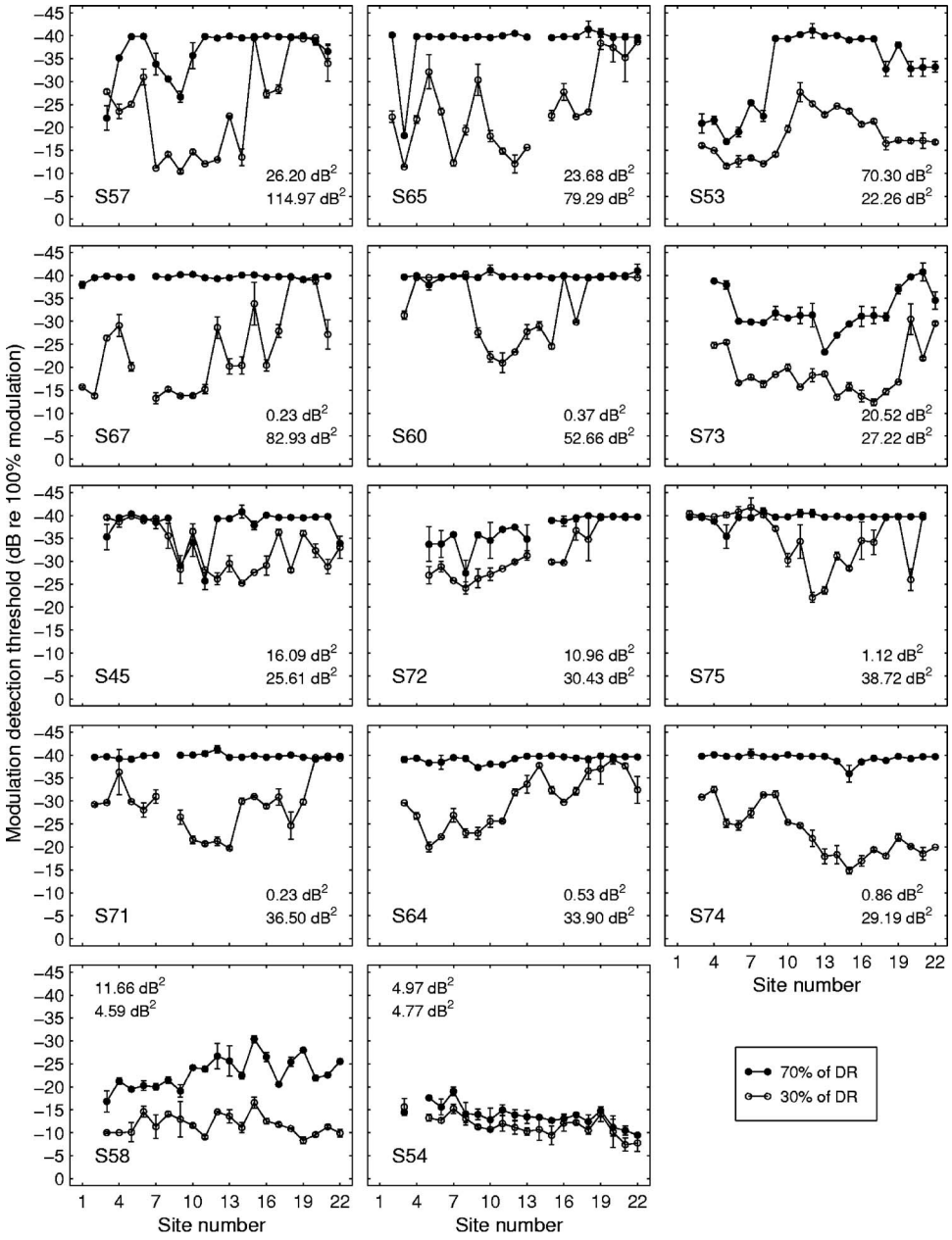


FIG. 1. Across-site patterns of MDTs for 14 subjects. Each panel shows data for one subject. MDTs measures at 70% of DR (filled symbols) and 30% of DR (open symbols) are plotted as a function of stimulation site. Error bars show the two MDTs obtained for each level at each site and the open and filled circles show the means of these two values. Stimulation sites are numbered in order from the most basal electrode in the 22-electrode array (Site 1) to the most apical electrode (Site 22). Subject numbers are given in the lower left corner of each panel. The variances shown in each panel (dB^2) are the across-site variance for MDTs measured at 70% of DR (top number) and the across-site variance for MDTs measured at 30% of DR (bottom number). The mean of these two variances was used to order the panels with the data for the subject with the highest mean variance shown in the upper left-hand panel and the subject with the lowest mean variance shown in the lower right-hand panel.

atic differences in susceptibility to pathology along the cochlear length, or systematic differences in medial-lateral electrode location as a function of cochlear length could also contribute to across-site variation in MDTs. To determine if there were systematic differences in MDTs along the cochlear length, the MDT data were divided into three groups based on the segment of the implant where the stimulation occurred: Basal (Sites 1–8), middle (Sites 9–15), and apical (Sites 16–22). Across-subject mean MDTs for each segment are shown in Fig. 2. A repeated-measures analysis of variation (ANOVA) indicated that there was an effect of segment on MDTs when the MDTs were measured at 30% of DR [$F(2,41)=4.695$, $p=0.0182$]. A post-hoc Tukey test indicated that the mean of the MDTs from the apical segment ($M=-27.76$ dB re 100% modulation, SD=9.84 dB) was lower than the mean of the MDTs from the middle segment ($M=-22.12$ dB, SD=6.13 dB). The mean of the MDTs from the basal segment ($M=-24.92$ dB, SD=9.48 dB) was not

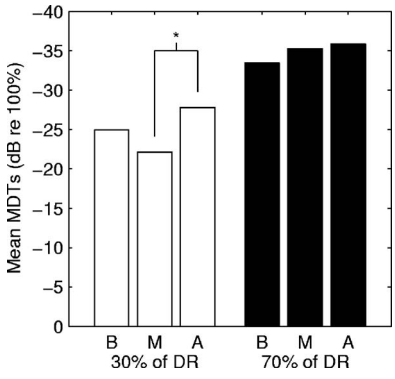


FIG. 2. Across-subject mean MDTs for each of three segments of the implant. Basal segment (B)=Sites 1–8; middle segment (M)=Sites 9–15; apical segment (A)=Sites 16–22. MDTs for all tested sites within each segment were averaged. Mean MDTs measures at 30% of DR and 70% of DR are shown. The asterisk (*) indicates a significant difference between the mean MDTs for the apical and middle segments at 30% of DR in a post-hoc Tukey test.

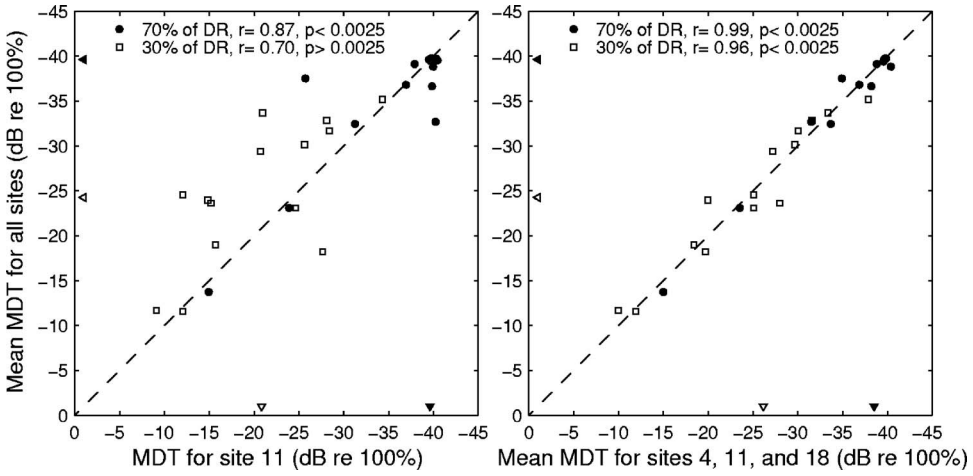


FIG. 3. Scatterplots illustrating the correlations between mean MDTs for all sites and MDTs at Site 11 (left-hand panel) or the mean MDT for Sites 4, 11 and 18 (right-hand panel) at 70% of DR (filled circles) and 30% of DR (open squares). The filled and open triangles pointing to the abscissa represent the medians of the MDTs across all 14 subjects at Site 11 (left-hand panel) and at Sites 4, 11, and 18 (right-hand panel) for 70% and 30% of DR, respectively. The filled and open triangles pointing to the ordinate represent the medians of the MDTs across all 14 subjects at all sites for 70% and 30% of DR, respectively.

significantly different from those from the two other segments. There was no statistically significant effect of segment on MDTs when measurements were made at 70% of DR [$F(2,41)=2.345$, $p=0.1158$].

Although statistically significant regional differences in MDTs were seen in the data averaged across all listeners at 30% of DR, these differences were small and they were not found in all individuals. Notable exceptions to the pattern shown in Fig. 2 are the data for subjects S53, S74, and S54 shown in Fig. 1.

B. Characterizing the listeners

Although across-site patterns of modulation detection are useful for many purposes, it might also be useful in some cases to have a single number to characterize a listener's modulation-detection acuity relative to that of other listeners. This can be done by averaging the MDTs across all sites, averaging across a subsample of sites within the electrode array or by determining the MDT for only a single site (as was done by Fu, 2002). Presumably, averaging MDTs across all sites would result in the most accurate assessment of a listener's overall modulation-detection acuity. However, this approach is very time consuming. If MDTs from a small sample of sites or a single site are closely related to, and accurately reflect the across-site mean of, MDTs at all sites, then the subjects' overall modulation detection skills could be more efficiently assessed by testing fewer sites. To determine the relationship between these three methods of acquiring a single value to quantify listeners' modulation-detection acuity, we calculated correlations between MDTs at each individual site (each available site from 3 to 22) and the mean MDTs across all sites. In most cases, the correlations calculated at both 30% and 70% of DR were large and highly significant after the Bonferroni method of correction was used to adjust the criterion p-value ($0.05/20=0.0025$). At 70% of DR, the r values of all correlations ranged from 0.74 to 0.98, and the p-values for all but one of these correlations were 0.001 or less. At 30% of DR, r values ranged from 0.63 to 0.92. All but five of these correlations resulted in significant p-values. One example of the relationship between MDTs for a single site (Site 11) and the means for all sites is shown in the left-hand panel of Fig. 3.

This analysis indicates that single-site measures of MDTs are highly related to the across-site mean MDTs. Listeners with relatively high MDTs at single sites also had high MDTs when MDTs from all sites were averaged together. Thus, measuring MDTs at only one site seems to provide a reasonable estimate of a listener's modulation-detection acuity relative to that of other listeners. However, simple correlations between single-site measures of MDTs and across-site means of MDTs do not indicate if measuring at a single site will result in comparable quantitative values or if measuring at a single site could significantly under- or overestimate listeners' modulation-detection acuity. For example, although MDTs measured at Site 11 were significantly correlated with mean MDTs across all sites, the two measures were not equivalent for many of the listeners. At 30% of DR, the MDTs measured at Site 11 tended to underestimate the mean modulation-detection acuity for all sites (points above the diagonal in the left-hand panel of Fig. 3). In some instances, it might be desirable to have more precise estimates of listeners' MDTs rather than just an estimated rank order of their modulation-detection acuity. To determine if MDTs measured at single sites were comparable to the across-site means for all sites, we calculated paired-samples t-tests between MDTs measured at single sites (each available site from 3 to 22) and the across-site mean MDTs. Because many listeners performed at ceiling when 70% of DR was used, these comparisons were only done for the data collected at 30% of DR. The results of these comparisons indicate that, in this group of listeners, MDTs measured at 6 of the 20 sites were significantly different from the across-site mean MDTs. MDTs measured in the middle of the array at Site 10 [$t(13)=2.79$, $p<0.05$], Site 11 [$t(13)=2.67$, $p<0.05$], and Site 13 [$t(13)=2.32$, $p<0.05$] significantly underestimated listeners' overall modulation-detection acuity. MDTs measured at the apical end of the array at Site 19 [$t(13)=-2.86$, $p<0.05$], Site 20 [$t(13)=-2.43$,$p<0.05$], and Site 21 [$t(13)=-2.39$, $p<0.05$] significantly overestimated listeners' overall modulation-detection acuity.

We also examined if MDTs averaged over three sites would be correlated to, and quantitatively equal to, the mean MDT calculated across all sites. For these analyses, the mean

TABLE II. Bivariate correlations between individual subjects' modulation detection thresholds (MDTs) and detection threshold levels (T levels; column 2), maximum comfortable loudness levels (C levels; column 3), and dynamic ranges (DRs; column 4) across all available stimulation sites. Asterisks (*) indicate significance at the $p<0.001$ level.

Subject	MDT vs. T level	MDT vs. C level	MDT vs. DR
45	0.14	0.54	0.25
53	0.00	−0.72*	−0.76*
54	0.35	0.11	−0.18
57	−0.20	−0.42	−0.55
58	−0.17	0.18	0.35
60	−0.02	0.21	0.23
64	−0.20	0.28	0.53
65	−0.52	−0.34	−0.38
67	−0.37	0.06	0.23
71	−0.23	0.38	0.39
72	0.85*	−0.24	−0.85*
73	0.01	0.44	0.67*
74	0.20	0.66*	0.34
75	0.10	0.42	0.38

MDTs for a basal site (Site 4), a middle site (Site 11), and an apical site (Site 18) were averaged and compared to the mean MDTs across all sites. Figure 3 (right panel) illustrates that MDTs averaged across the three sites along the electrode array were significantly correlated to MDTs averaged across all sites at both 30% ($r=0.99$, $p<0.001$) and 70% ($r=.96$, $p<0.001$) of the dynamic range. A paired-samples t-test was conducted to compare quantitatively the mean MDTs of three sites to the mean MDTs of all sites. There was no significant difference found between the mean MDT for Sites 4, 11, and 18 and the mean MDT of all sites at either 30% of DR [$t(13)=0.02$, $p=0.99$] or 70% of DR [$t(13)=0.60$, $p=0.56$]. Thus, sampling at three sites (Fig. 3, right-hand panel) clearly gave a better estimate of the mean MDT for all sites than just sampling at the middle site (Fig. 3, left-hand panel) or several other single sites along the electrode array.

C. Relation to across-site patterns of other measures

Correlational analysis was used to determine if across-site patterns of MDTs could be predicted from measures of T and C levels or DRs. In each listener, bivariate correlations were conducted between MDTs at the individual sites and T levels, C levels, and DRs at the individual sites. This analysis showed only sporadic relationships between the psychophysical measures within listeners. The sign of the correlation was also variable across listeners. For some listeners, values of T levels, C levels, and DRs increased as a function of MDT. For other listeners, these values decreased. Table II lists the r-values obtained for each listener for correlation of the MDT at each site with the C level, T level, and DR at each site. Significance levels were determined using the Bonferroni correction ($p=0.05/42=0.001$). Only 6 of the 42 correlations were significant by this criterion and those did not all have the same sign: Three were positive and three were negative. Thus, it was not possible to consistently predict MDTs from T levels, C levels, or DRs.

IV. DISCUSSION

This study demonstrated that MDTs are highly variable across stimulation sites of cochlear-implant electrode arrays, consistent with the hypothesis that the MDTs are affected by local variation in the number of neurons activated by each site and/or the condition of those neurons. Most of the listeners in our sample had MDTs that were highly variable across sites at a stimulus level that was 30% of DR.

These results suggest that strategies to improve a listener's modulation detection performance might best be directed at individual stimulation sites in the electrode array. Such strategies might include adjusting stimulation parameters at individual sites to optimize modulation detection for each site or simply removing weaker sites from the processor map in order to improve the overall mean modulation detection scores and/or reduce any negative effects of the weak sites or the high across-site variability. One caution regarding such an approach is that strategies to improve modulation detection could potentially degrade other important features of perception. Before proceeding with such strategies, we need to know more about the relationship between across-site patterns of MDTs and across-site patterns of other perceptual features of electrical stimulation and to better understand the importance of these across-site patterns for speech recognition and other aspects of perception with auditory prostheses.

Potential rehabilitation strategies based on modulation-detection acuity rely on the assumption that modulation-detection acuity is closely related to speech recognition with auditory prostheses. There is some information about these relationships in the literature (Cazals *et al.*, 1994; Fu, 2002; Colletti and Shannon, 2005), but additional studies are needed to determine how the patterns of MDTs across the whole implant electrode array relate to speech recognition. It is also important to determine how these across-site patterns are affected by the presence of interleaved stimulation at other stimulation sites. Modulation detection interference is known to occur with multichannel stimulation (Richardson *et al.*, 1998; Chatterjee and Oba, 2004), and this could have a large effect on the across-site pattern of modulation detection and its relationship to speech recognition with auditory prostheses.

The MDTs in the experiments reported here were measured using monopolar stimulation, which is the electrode configuration used in most contemporary auditory prostheses. Monopolar stimulation has been shown to activate a broad spatial extent of neurons at levels a few decibels above the neural threshold (Merzenich and White, 1980; Hartmann *et al.*, 1984; Bierer and Middlebrooks, 2002). This suggests that there is considerable overlap in the neural populations activated by stimulation of sites in the implant that are close together. However, the large variation seen in MDTs across adjacent or nearby stimulation sites in the current study suggests that the neural responses for stimulation of nearby sites are not identical in most cases. It is possible that stimulation using electrode configurations that produce more restricted activation patterns, such as tripolar stimulation, would result in even greater across-site variation in MDTs. However,

given that these configurations are not currently used in most speech processors, the practical significance of using these configurations for evaluation of a subject's temporal acuity is limited.

This study, as well as previous studies, showed that MDTs improve as a function of stimulus level in almost all cases. However, the shape of the MDT versus level function varies considerably from listener to listener (Fu, 2002; Galvin and Fu, 2005; Pfingst *et al.*, 2007). In some listeners, these functions have steep slopes at low levels but reach a ceiling and plateau at higher levels. Other listeners show steady improvements in MDTs throughout the dynamic range. These variations are also seen across stimulation sites within listeners (e.g., see Fig. 4 in Pfingst *et al.*, 2007). In many cases the differences across sites are reduced at higher levels in the dynamic range because the MDTs have reached ceiling performance levels. However, in some cases, the across-site differences are maintained over a large extent of the dynamic range. For example, see the data for S53 in Fig. 1 from the current study where MDTs at Sites 2–5 at 70% of DR are poorer than those at Sites 10–14 at 30% of DR. Because of the differences in growth of MDT versus level functions, it would be possible in some cases to reduce across-site variation in MDTs by making small site-specific adjustments in stimulus level, i.e., changing the input–output functions for each channel. On the other hand, some across-site differences persist over such a large range of levels that this approach would not be practical.

The large variability in MDTs across stimulation sites suggests that measurements made at a single site might not provide an adequate assessment of cochlear implant users' modulation-detection acuity. For example, although MDTs obtained at a single site in the middle of the electrode array (Site 11) at 30% of DR were significantly correlated with the mean MDTs obtained at all sites, substantial differences were found between the MDTs at Site 11 and the mean MDTs for all sites in several individual listeners. Thus, measuring MDTs at a single site could result in over- or underestimations of temporal processing acuity in many cochlear implant users. Correspondence to the mean MDT for all sites was much closer when the mean of MDTs measured at three sites spaced throughout the electrode array (Sites 4, 11, and 18) was used. Thus, measuring MDTs at more than one site is advantageous for obtaining accurate estimates of cochlear implant users' modulation-detection acuity. Such an estimate of a listener's mean modulation-detection acuity might be useful for predicting cochlear implant performance (e.g., Fu, 2002) and for identifying whether or not the weaknesses in speech recognition or other complex perceptual abilities are due to deficits in temporal-envelope perception (e.g., Colletti and Shannon, 2005). However, assuming that the source of a listener's perceptual weakness is related to temporal-envelope perception, a more detailed analysis of MDTs across the whole electrode array would be needed to identify the specific stimulation sites where changes are required.

In this study, we also found a weak dependence of MDTs on position of the stimulation sites along the apical-basal dimension of the scala tympani. MDTs were significantly lower at apical sites than at sites located in the middle of the array. A number of variables could contribute to this regional difference including (1) more hair cells present in the cochlear apex (Kiefer *et al.*, 2004) that contribute to spontaneous activity in the neurons and alter sensitivity to temporal modulations; (2) systematic differences in temporal response properties of neurons from base to apex in the cochlea (Adamson *et al.*, 2002; Liu and Davis, 2006); (3) better nerve survival in the apex of the cochlea (Nadol, 1997); and (4) systematic variation in the positions of the electrodes with respect to the modiolus along the cochlear-implant electrode array (Saunders *et al.*, 2002). However, although regional differences in MDTs were seen in the average data across all listeners at 30% of DR, they were not found in all individuals. The more prominent characteristic of these data was the listener-specific variation in the across-site patterns of MDTs. This finding suggests that localized listener-specific patterns of the pathology of deafness and/or electrode placement play the dominant role in determining the across-site patterns of modulation detection.

If across-site patterns of MDTs are determined by the pattern of variation in physiological and biophysical variables along the length of the electrode array, then we might expect that other measures of implant function would show similar patterns of across-site variation. It is known that *T* levels, *C* levels, and DRs do vary as a function of stimulation site along the electrode array (Pfingst and Xu, 2004, 2005). However, in this study we found that, for most listeners, the across-site patterns of these psychophysical measures were not correlated with across-site patterns of MDTs. This suggests that the specific mechanisms underlying the across-site patterns of variation differ across these various psychophysical measures. The weak and inconsistent correlations between MDTs and *T* levels, *C* levels, and DRs make it improbable that these clinical measures would be successful predictors of modulation-detection acuity.

ACKNOWLEDGMENTS

The authors express appreciation to their research subjects for their cheerful participation in these studies, to Ian Hsu for assistance with data analysis and presentation, to Thyag Sadasiwan for his computer programming work, and to the associate editor and reviewers of JASA for insightful comments and helpful suggestions on earlier drafts of this article. This work was supported by NIH NIDCD Grant Nos. R01 DC04312, T32 DC00011, and P30 DC05188.

Adamson, C. L., Reid, M. A., Mo, Z. L., Bowne-English, J., and Davis, R. L. (**2002**). "Firing features and potassium channel content of murine spiral ganglion neurons vary with cochlear location," J. Comp. Neurol. **447**, 331–350.

Bierer, J. A., and Middlebrooks, J. C. (**2002**). "Auditory cortical images of cochlear-implant stimuli: Dependence on electrode configuration," J. Neurophysiol. **87**, 478–492.

Bregman, A. S., Levitan, R., and Liao, C. (**1990**). "Fusion of auditory components: Effects of the frequency of amplitude modulation," Percept. Psychophys. **47**, 68–73.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **96**, 2048–2054.

Chatterjee, M., and Oba, S. I. (**2004**). "Across- and within-channel envelope interactions in cochlear implant listeners," J. Assoc. Res. Otolaryngol. **5**,

360–375.
Chatterjee, M., Sarampalis, A., and Oba, S. I. (**2006**). "Auditory stream segregation with cochlear implants: A preliminary report," Hear. Res. **222**, 100–107.
Colletti, V., and Shannon, R. V. (**2005**). "Open set speech perception with auditory brainstem implant," Laryngoscope **115**, 1974–1978.
Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). "Effect of temporal envelope smearing on speech perception," J. Acoust. Soc. Am. **95**, 1053–1064.
Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.
Fu, Q.-J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.
Fu, Q.-J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.
Fu, Q.-J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589–597.
Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.
Galvin, J. J., and Fu, Q.-J. (**2005**). "Effects of stimulation rate, mode and level on modulation detection by cochlear implant users," J. Assoc. Res. Otolaryngol. **6**, 269–279.
Hartmann, R., Topp, G., and Klinke, R. (**1984**). "Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea," Hear. Res. **13**, 47–62.
Hinojosa, R., and Lindsay, J. R. (**1980**). "Profound deafness. Associated sensory and neural degeneration," Arch. Otolaryngol. **106**, 193–209.
Kiefer, J., Gstoettner, W., Baumgartner, W., Pok, S. M., Tillein, J., Ye, Q., and von Ilberg, C. (**2004**). "Conservation of low-frequency hearing in cochlear implantation," Acta Oto-Laryngol. **124**, 272–280.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liu, Q., and Davis, R. L. (**2006**). "From apex to base: How endogenous neuronal membrane properties are distributed in the spiral ganglion," Assoc. Res. Otolaryngol. Abstr. **29**, 305.
Merzenich, M. M., and White, M. (**1980**). "Coding considerations in design of cochlear prostheses," Ann. Otol. Rhinol. Laryngol. Suppl. **89**, 84–87.
Nadol, J. (**1997**). "Patterns of neural degeneration in the human cochlea and auditory nerve: Implications for cochlear implantation," Otolaryngol.-Head Neck Surg. **117**, 220–228.
Pfingst, B. E., and Xu, L. (**2004**). "Across-site variation in detection thresholds and maximum comfortable loudness levels for cochlear implants," J. Assoc. Res. Otolaryngol. **5**, 11–24.
Pfingst, B. E., and Xu, L. (**2005**). "Psychophysical metrics and speech recognition in cochlear implant users," Audiol. Neuro-Otol. **10**, 331–341.
Pfingst, B. E., Xu, L., and Thompson, C. S. (**2004**). "Across-site threshold variation in cochlear implants: Relation to speech recognition," Audiol. Neuro-Otol. **9**, 341–352.
Pfingst, B. E., Xu, L., and Thompson, C. A. (**2007**). "Effects of carrier pulse rate and stimulation site on modulation detection by subjects with cochlear implants," J. Acoust. Soc. Am. **121**, 2236–2246.
Richardson, L. M., Busby, P. A. and Clark, G. M. (**1998**). "Modulation detection interference in cochlear implant subjects," J. Acoust. Soc. Am. **104**, 442–452.
Saunders, E., Cohen, L., Aschendorff, A., Shapiro, W., Knight, M., Stecker, M., Richter, B., Waltzman, S., Tykocinski, M., Roland, T., Laszig, R., and Cowan, R. (**2002**). "Threshold, comfortable level and impedance changes as a function of electrode-modiolar distance," Ear Hear. **23**, 28S–40S.
Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.
Shannon, R. V., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
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.
Wilson, B. S. (**2004**). "Engineering design of cochlear implants," in *Cochlear Implants: Auditory Prostheses and Electrical Hearing*, edited by F.-G. Zeng, A. N. Popper, and R. R. Fay (Springer, New York), pp. 14–52.
Xu, L., Thompson, C., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.

Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task[a)]

Michael A. Stone[b)] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 16 November 2006; accepted 11 November 2007)

These experiments are concerned with the intelligibility of target speech in the presence of a background talker. Using a noise vocoder, Stone and Moore [J. Acoust. Soc. Am. **114**, 1023–1034 (2003)] showed that single-channel fast-acting compression degraded intelligibility, but slow compression did not. Stone and Moore [J. Acoust. Soc. Am. **116**, 2311–2323 (2004)] showed that intelligibility was lower when fast single-channel compression was applied to the target and background after mixing rather than before, and suggested that this was partly due to compression after mixing introducing "comodulation" between the target and background talkers. Experiment 1 here showed a similar effect for multi-channel compression. In experiment 2, intelligibility was measured as a function of the speed of multi-channel compression applied after mixing. For both eight- and 12-channel vocoders with one compressor per channel, intelligibility decreased as compression speed increased. For the eight-channel vocoder, a compressor that only affected modulation depth for rates below 2 Hz still reduced intelligibility. Experiment 3 used 12- or 18-channel vocoders. There were between 1 and 12 compression channels, and four speeds of compression. Intelligibility decreased as the number and speed of compression channels increased. The results are interpreted using several measures of the effects of compression, especially "across-source modulation correlation."
 [DOI: 10.1121/1.2821969]

PACS number(s): 43.66.Ts, 43.66.Mk, 43.71.Gv [KWG] Pages: 1063–1076

I. INTRODUCTION

Fast-acting dynamic range compression is widely used in electro-acoustic systems such as hearing aids or cochlear implants to keep the levels of signals at the output within prescribed limits. The fast compression increases the audibility of low-level signal components (Villchur, 1973), which sometimes leads to improved intelligibility of speech (Moore *et al.*, 1992; Yund and Buckles, 1995b), but does not always do so (De Gennaro *et al.*, 1986; Boothroyd *et al.*, 1988; Drullman and Smoorenburg, 1997). A single-channel compressor reduces temporal contrast (related to the depth of modulation in the signal envelope) but has no effect on spectral contrast (related to the magnitude of peaks and dips in the spectrum). Multi-channel compression, because it can change gain across frequency as well as time, reduces both spectral and temporal contrast (Plomp, 1988). Reduction of either type of contrast can lead to reduced intelligibility of speech, especially when background sounds are present (ter Keurs *et al.*, 1992; 1993; Baer and Moore, 1993; 1994; Stone, 1995; Chi *et al.*, 1999; Turner *et al.*, 1999). This may partly account for the fact that fast-acting compression sometimes has deleterious effects on speech intelligibility when background sounds, such as a competing talker, are present. Hereafter, we focus on the task of identifying the speech of a target talker in the presence of a background talker.

Stone and Moore (2004) identified a further factor that may contribute to the reduced intelligibility sometimes produced by fast-acting compression when a background talker is present, which they called "comodulation." Here, following Stone and Moore (2007), we use the term "across-source modulation correlation" (ASMC) to describe this factor, since it is used to refer to situations where the amplitude-modulation patterns of one signal influence the amplitude-modulation patterns of another signal. When two or more people speak at the same time, the auditory system is faced with the task of determining which frequency components in the mixture emanated from the target talker, and which components emanated from the background talker(s). The patterns of amplitude modulation of the speech from a single talker are partially correlated across different frequency regions (Steeneken and Houtgast, 1999), and the auditory system appears able to exploit this to group together the frequency components arising from that talker and to separate them from the (independently modulated) components arising from other talkers (Bregman, 1990; Bregman *et al.*, 1985; Rappold *et al.*, 1993). ASMC arises when the variable gain produced by a single-channel compressor is applied to a mixture of speech signals of nearly equal level. Peaks in one signal cause a reduction in gain that reduces the levels of all the signals. Signals that were previously independently am-

[a)] Some of these data were presented at the British Society of Audiology conference "Experimental studies of hearing and deafness" held at Cardiff University, United Kingdom, 12th and 13th September 2005.
[b)] Author to whom correspondence should be addressed. Electronic mail: mas19@cam.ac.uk

plitude modulated, and therefore unlikely to fuse perceptually, acquire a common component of modulation from the gain control, and their independence is thereby reduced. The reduced independence leads to undesirable perceptual fusion of the target and background, and this may contribute to reduced intelligibility.

Stone and Moore (2007) proposed a quantitative measure of ASMC, which is described here in slightly simplified form (as are the other measures described later). For the mixture of target and background, the gain-control signal produced by the compressor is calculated (separately for each channel in the case of multi-channel compression). This gain-control signal is applied independently to the original target (giving $\text{Target}_{\text{postcomp}}$) and to the original background (giving $\text{Background}_{\text{postcomp}}$). The Pearson correlation, r, is calculated between the envelope of channel i of $\text{Target}_{\text{postcomp}}$ and the envelope of channel i of $\text{Background}_{\text{postcomp}}$. This is repeated for all i and averaged across i to give the measure of ASMC. Mathematically, this is expressed as

$$\text{ASMC} = \frac{1}{N}\sum_{i=1}^{N} r(a_i,b_i), \tag{1}$$

where a_i and b_i are the envelopes of $\text{Target}_{\text{postcomp}}$ and $\text{Background}_{\text{postcomp}}$, respectively, in the ith channel. For this measure, and the other measures discussed below, the logarithm of the envelope is used in the calculations, rather than the linear value, for reasons discussed in Stone and Moore (2007).

Stone and Moore (2004) used a noise-vocoder simulation of the information conveyed by a cochlear implant (Shannon *et al.*, 1995) to demonstrate the influence of ASMC. An advantage of using the vocoder is that it conveys only envelope information in a limited number of frequency bands, and is therefore likely to reveal deleterious effects of compression on the representation of the envelope. In experiment 2 of Stone and Moore (2004), single-channel compression was applied to the speech of a target and an interfering talker either before mixing (condition INDEP) or after mixing (condition COMOD) of the two signals at target-to-background ratios (TBRs) near 0 dB. In condition INDEP, the temporal contrast of both signals was reduced, but there was no ASMC. In condition COMOD, temporal contrast was reduced for both signals (by a similar amount as for condition INDEP), but ASMC was also introduced because the gain-control signal was derived from the mixture of the target and background. Intelligibility was lower for condition COMOD, and Stone and Moore (2004) attributed this difference to ASMC.

Stone and Moore (2007) described two other effects of fast-acting compression that might have influenced the results of Stone and Moore (2004). The first of these is related to within-source modulation correlation (WSMC). If the speech of a single talker is filtered into frequency channels, and the envelope is extracted for each channel, then the envelope fluctuations are correlated to some extent in different frequency channels. Stone and Moore (2007) defined WSMC as the degree of correlation of the envelope (on a dB scale) of a single source across different frequency regions (i.e., after filtering into frequency channels). Let $r(a_i,a_j)$ be the Pearson correlation of the envelopes of the ith channel of signal a, and the jth channel of signal a. The value of WSMC is defined as

$$\text{WSMC} = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} r(a_i,a_j), \quad i \neq j, \tag{2}$$

where N is the number of channels. Effectively this quantity gives a measure of the similarity of the envelope of each channel with that of all the other channels, averaged across channels.

Correlated fluctuations across the different frequency channels of the speech of a single talker may help to bind the speech into a single perceptual stream, and when competing speech is present they may help to segregate the target speech from the background speech. For example, sinewave speech (Remez *et al.*, 1981), in which the formants in speech are replaced by three or four sinewaves tracking the lower formant frequencies, has relatively low WSMC. However, the WSMC can be increased by amplitude modulating all of the sinewaves in the same way, and this leads to improved intelligibility (Carrell and Opie, 1992; Carrell, 1993).

WSMC may have influenced the results of experiment 2 of Stone and Moore (2004) since changes in WSMC produced by compression were different for conditions INDEP and COMOD. However, Stone and Moore (2007) reported that the measures of WSMC were slightly lower for condition INDEP than for condition COMOD, but intelligibility was better for condition INDEP than for condition COMOD. This suggests that WSMC was not the main factor affecting intelligibility for these stimuli.

The other factor that may have influenced intelligibility in the experiments of Stone and Moore (2004) is related to fidelity of envelope shape (FES), which was defined by Stone and Moore (2007) as the degree to which the envelope shape of the target in different frequency channels is preserved following compression. Stone and Moore suggested that FES could be quantified by determining the correlation between the envelope of channel i for the original target and the envelope of channel i of the target after compression (denoted b_i), repeating this for all i, and averaging the results across i. This is expressed by

$$\text{FES} = \frac{1}{N}\sum_{i=1}^{N} r(a_i,b_i), \tag{3}$$

Fast-acting compression can markedly alter the envelope of speech from a single talker. The envelope may be distorted in shape (Stone and Moore, 1992), and abrupt changes in envelope magnitude can lead to "overshoot" and "undershoot" effects (Verschuure *et al.*, 1996). The degree to which these effects occur depends on the design of the compressor, for example on whether the audio signal is delayed so as to time-align it with the gain-control signal (Robinson and Huntington, 1973). When compression is applied to a mixture of the speech from different talkers, the envelope shape associated with the target speech is altered, since the gain changes produced by the compressor are determined by the

mixture rather than just by the target speech. However, Stone and Moore (2007) reported that FES was lower for condition INDEP than for condition COMOD. This suggests that FES was not the main factor affecting intelligibility for these stimuli.

The experiments of Stone and Moore (2003; 2004) used only single-channel compression. In the present paper we report experiments evaluating the effect of fast-acting multi-channel compression on speech intelligibility, using stimuli processed using a noise vocoder. The results are interpreted taking into account all three of the measures discussed above, ASMC, WSMC, and FES.

II. EXPERIMENT 1: EFFECTS OF APPLYING MULTI-CHANNEL COMPRESSION BEFORE AND AFTER MIXING THE SPEECH FROM TWO TALKERS

A. Speech stimuli and equipment

The female target talker (BKB sentences, Bench and Bamford, 1979) and male background talker, equipment, and signal-processing method were the same as for experiment 2 of Stone and Moore (2004). The background talker produced speech that was highly modulated in both the long and short term. In all the experiments described in this paper, the speech of the background talker started about 1 s before and ended 0.5–1 s after each target sentence. As in our earlier work, the stimuli were "pre-emphasized" by applying a gain rising at 3.3 dB/octave 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 input level of the pre-emphasized target speech to the processing was 67 dB sound pressure level (SPL) (unweighted), and the presentation level of the mixed signal after processing was 68 dB SPL, for both conditions, as measured when the target and background were present at the same time. Replay of pre-processed material was through a Lynx ONE soundcard hosted in a PC, under control of MATLAB. Signal buffering of the soundcard output was provided by a Mackie 1202-VLZ PRO mixing desk. Presentation was diotic via Sennheiser HD580 headphones. Subjects were tested in a double-walled, sound-attenuating booth.

B. Conditions, compressor type and procedure

The target talker and background talker were compressed with a fast-acting multi-channel compressor either before (condition "BEFORE") or after (condition "AFTER") mixing. The compression was applied independently in each and every channel. Six and 11 channels of compression were used with TBRs of +8 and +3 dB, respectively. These ratios led to moderate intelligibility in experiment 2 of Stone and Moore (2004). Since compression alters the histogram of short-term levels within a signal, estimating the levels of signals before and after compression requires care in order match TBRs. Rather than measure the energy of the signal above a fixed threshold, the mean level was estimated by summing the energy from the 80% of 10 ms frames with the highest levels. The output signal from each compressor was subsequently noise vocoded as described by Stone and Moore (2003). The lowpass filter used to extract the envelope signal for each channel had a cutoff frequency of 45 Hz and negligible response by 100 Hz. The number of vocoder channels was equal to the number of compression channels.

The compressor was the "envelope" compressor, as used in experiment 1 of Stone and Moore (2004), which has near-symmetric attack and release times of about 10 ms. The effect of a compressor on envelope modulation can be characterized using the fractional reduction in modulation, f_r, in each channel, defined as $(CR_e-1)/CR_e$, where CR_e is the effective compression ratio for a specific modulation frequency (Stone and Moore, 1992); f_r represents the fraction of modulation removed by the compressor (Stone and Moore, 2003). It is dependent on modulation frequency, and reaches its maximum value for very low modulation frequencies. A plot of f_r versus modulation frequency for the envelope compressor is shown as line 4 of Fig. 2, presented later. This "f_r plot" is almost flat up to about 14 Hz, and falls off rapidly above that, reaching zero by 100 Hz: compression or expansion for modulation rates exceeding 100 Hz is negligible.

Each processing scheme was assessed using 33 sentences, giving a maximum score of 99 keywords correct. The four processing schemes (conditions BEFORE or AFTER and two numbers of channels) were tested in a counterbalanced design.

C. Equating effective compression for single- and multi-channel compression

To allow comparison of the results with those obtained earlier using single-channel compression, the compression ratio (CR) was independently selected for each channel so as to match the compression produced by the single-channel wideband compressor used in experiment 2 of Stone and Moore (2004). The CR for each channel was chosen to equate the mapping of short-term (125 ms) signal levels between input and output of the channel compressor to that produced in the same frequency band by a single-channel compressor of the same speed operating on the same speech signal. The single-channel envelope compressor had a CR of 2.78 and a compression threshold of 55 dB SPL, as used in experiments 1 and 2 Stone and Moore (2004). The compression thresholds for the multi-channel compressor were set 13 dB below the mean level in each channel, to ensure that the channel compression was activated frequently. The resulting CRs for conditions BEFORE and AFTER are given in Table I, separately for each vocoder system. The table also shows maximum values of f_r for each channel (values applicable for very low modulation rates).

D. Subjects and training

Twelve subjects (4M, 8F, mean=21.3 years, standard deviation (SD)=2.8 years, range 20–30 years), all university undergraduates, were selected on the basis of their having audiometric thresholds ≤15 dB hearing level (HL) at octave frequencies between 125 and 8000 Hz and at 3000 and 6000 Hz. All were native speakers of British English. They were paid for their attendance. All subjects had prior experience of a generic form of the processing, in an experiment

TABLE I. Compression ratios (CRs) used in the channel compressors of experiment 1. Part A is for the six-channel system, which used a TBR of +8 dB. Part B is for the 11-channel system, which used a TBR of +3 dB. The first line in each section shows the channel index. Lines two and three show the static CR and the corresponding value of f_r, the fractional reduction in modulation, that was set for each channel of the system with compression BEFORE mixing, in order to match the effective compression in each channel to that produced by the single-channel compressor, whose CR was 2.78. Lines four and five show CRs for the system with compression AFTER mixing and the corresponding f_r values.

Condition	Channel number										
A. Six-channel system		1		2		3		4		5		6	
BEFORE: channel CR		2.49		1.48		1.47		1.59		1.38		1.03	
f_r		*0.60*		*0.32*		*0.32*		*0.37*		*0.28*		*0.03*	
AFTER: channel CR		2.17		1.54		1.52		1.63		1.25		1.14	
f_r		*0.54*		*0.35*		*0.34*		*0.39*		*0.20*		*0.12*	
B. 11-channel system	1	2	3	4	5	6	7	8	9	10	11		
BEFORE: channel CR	2.07	1.66	1.46	1.29	1.41	1.51	1.50	1.61	1.29	1.14	1.00		
f_r	*0.52*	*0.40*	*0.32*	*0.22*	*0.29*	*0.34*	*0.33*	*0.38*	*0.22*	*0.12*	*0.00*		
AFTER: channel CR	1.94	1.60	1.42	1.30	1.36	1.40	1.47	1.45	1.18	1.18	1.00		
f_r	*0.48*	*0.38*	*0.30*	*0.23*	*0.26*	*0.29*	*0.32*	*0.31*	*0.15*	*0.15*	*0.00*		

that had occurred within the preceding two to seven days, but additional training was given before data collection began. Subjects attended one session, which lasted just over 1 h.

The training involved subjects reporting back sentences with increasing degree of difficulty as the training progressed. The increase in difficulty was achieved by two means: reducing the number of vocoder channels, from a starting value of 16, and reducing the TBR progressively towards the values used in the data collection. To ensure that the training was generic in form, different target and background talkers were used in part of the training from those used in the data collection.

E. Results

The scores were transformed into rationalized arcsine units (RAU, Studebaker, 1985). To check for a possible training effect across successive conditions, the time-ordered averaged scores were expressed relative to the average score obtained in the last condition tested. The resulting relative scoring rates for the first three conditions tested (in time order) were 0.96, 0.93 and 0.99 of the rate in the final condition tested. Using a t test, the means for the first and last conditions were compared. There was no significant difference between the two ($t=0.45$, 11 df, $p>0.25$, one tailed), implying that a learning effect was not present, consistent with previous results obtained using subjects with similar amounts of training.

The mean scores, transformed from RAUs back into percentages, are plotted in Fig. 1. The error bars show SDs across subjects. For both numbers of channels, intelligibility was higher for condition BEFORE, where the target and background were compressed before mixing, than for condition AFTER. A two-way repeated-measures analysis of variance (ANOVA) was conducted on the RAU-transformed scores with factors condition (BEFORE or AFTER) and number of channels (6 or 11). The comparison of scores across channel numbers is not meaningful, as the TBR was different for the six- and 11-channel systems. There was a significant effect of condition, $F(1,11)=23.1$, $p<0.001$; post hoc tests showed that the effect of condition was significant for both channel numbers ($t>2.61$, 21.8 df, $p<0.02$, two tailed). The interaction between condition and number of channels was not significant, $F(1,11)=0.70$, $p=0.42$.

F. Discussion

Table II shows values of the three measures, WSMC, FES and ASMC, for the two conditions (AFTER and BEFORE) and for the uncompressed stimuli. The table also shows mean intelligibility scores. The values of WSMC were decreased slightly by the application of compression, and for each number of channels the WSMC measures were very slightly higher for condition AFTER than for condition BEFORE. Since higher values of WSMC are assumed to be a "good thing," and since intelligibility was higher for condition BEFORE than for condition AFTER, is appears that WSMC was not the factor responsible for the higher intelligibility in condition BEFORE.

The values of the measures FES also decreased following compression, although they remained rather high (greater than 0.97). This confirms that the envelope compressor did not markedly change the shape of the envelope of the com-

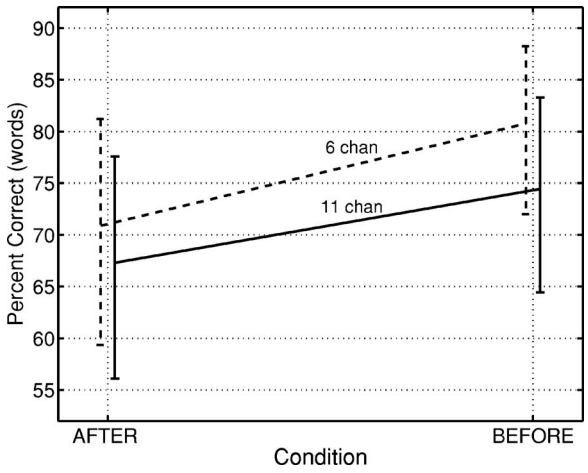


FIG. 1. Results of experiment 1 for the six-channel vocoder (dashed line) and the 11-channel vocoder (solid line). Error bars show ±1 SD. The target and background were subject to multi-channel compression either AFTER or BEFORE mixing. Comparisons of performance across channel numbers are not meaningful, as the TBR differed for the six- and 11-channel systems.

TABLE II. Values of the various measures of envelope distortion for the stimuli used in experiment 1. The bottom row shows the mean intelligibility scores for these stimuli. Intelligibility was not measured for the unprocessed (uncompressed) stimuli.

	Unprocessed		AFTER		BEFORE	
	6 channel	11 channel	6 channel	11 channel	6 channel	11 channel
WSMC	0.288	0.253	0.273	0.243	0.263	0.237
FES	1.0	1.0	0.977	0.980	0.978	0.983
ASMC	0.012	0.012	–0.111	–0.070	0.014	0.016
Intelligibility, %			70.9	67.3	80.8	74.4

pressed signal. The FES values were very similar for conditions AFTER and BEFORE, indicating that FES was also not the factor responsible for the difference in intelligibility between the two conditions.

The values of ASMC became negative following application of compression in condition AFTER. These negative values are as expected (Stone and Moore, 2004; 2007). They occur because peaks in the speech of one talker result in a reduction in gain for the speech of the other talker. Hence, increases in level of the speech from one talker result in decreases in level of the speech from the other talker, and vice versa. ASMC appears to be the only measure that is consistently related to the obtained intelligibility scores. Note that the reduction in modulation depth produced by the compressors was equated for conditions BEFORE and AFTER, so the differences in intelligibility scores for the two conditions cannot be attributed to this factor.

Table III compares the intelligibility scores for this experiment and for experiment 2 of Stone and Moore (2004). The same speech material and TBRs were used for the two experiments, although the subjects differed. For each number of channels the scores are similar for condition COMOD and condition AFTER. However, scores are higher for condition INDEP than for condition BEFORE, especially for the six-channel vocoder. Hence, the difference between scores for conditions INDEP and COMOD is greater than the difference between scores for conditions BEFORE and AFTER. The fact that performance was worse for condition BEFORE (multi-channel compression) than for condition INDEP (single-channel compression) can probably be attributed to the deleterious effect of multi-channel compression on short-term spectral contrast (Plomp, 1988).

III. EXPERIMENT 2: ASSESSING THE EFFECTS OF THE SPEED OF MULTI-CHANNEL COMPRESSION USING A NOISE VOCODER

A. Background and rationale

Experiments 2 and 3 of Stone and Moore (2003) demonstrated that single-channel fast-acting compression consistently degraded intelligibility relative to either slow-acting compression or no compression for stimuli processed through a noise vocoder with a range of channel numbers. The fast and slow compression systems exhibited very different f_r plots: parts of these are shown in Fig. 2 as the dashed lines marked "F" and "S," respectively. The slow system led to large values of f_r below 0.1 Hz but had no effect (f_r=0) for modulation rates greater than 0.5 Hz. The fast system led to large values of f_r for modulation rates up to about 1.4 Hz, but to f_r values near zero for modulation rates greater than about 20 Hz. Experiment 2 reported here explored the effects on intelligibility of multi-channel compression systems with speeds bracketing that of the fast system tested earlier. This experiment had similarities to that of Drullman *et al.* (1994a), in which the signal was filtered into narrow frequency bands and each band was represented as a carrier (temporal fine structure) modulated by an envelope, similar to Dudley's vocoder (Dubley, 1939). The envelope modulations of speech in a continuous background noise were reduced by high-pass filtering of the envelope. Drullman *et al.* concluded that, for normal-hearing listeners, the envelope modulation depth for rates below 4 Hz could be reduced without affecting speech intelligibility, independent of spectral resolution (0.25- or 1-octave wide bands). A complementary experiment (Drullman *et al.*, 1994b), in which the envelopes were subject to low-pass filtering, indi-

TABLE III. Comparison between the scores (in percent correct) of experiment 2 of Stone and Moore (2004) and experiment 1 here. Part A is for the six-channel system. Part B is for the 11-channel system. The second line in each section shows the scores in experiment 2 of Stone and Moore (2004) which used single-channel compression. Line three shows the scores in experiment 1 here which used multi-channel compression. The final column shows the difference in scores between conditions INDEP/BEFORE and COMOD/AFTER.

Condition	COMOD/AFTER	INDEP/BEFORE	Difference
A. Six-channel system			
Single-channel	65.6	89.0	23.4
Multi-channel	70.9	80.8	9.9
B. 11-channel system			
Single-channel	68.7	79.7	11.0
Multi-channel	67.3	74.4	7.1

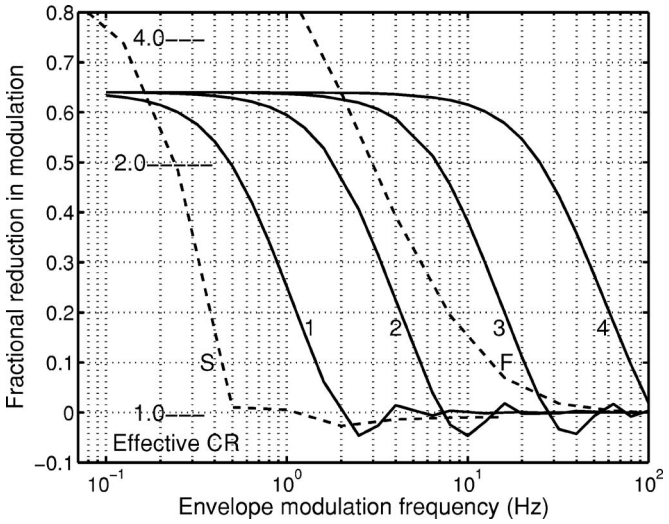


FIG. 2. Plot of fractional modulation reduction, f_r, for the four compressor speeds used in experiment 2 (solid lines labeled 1–4; compression speed increases progressively from 1 to 4). For comparison purposes, the dashed lines marked S and F are for the slow and fast compressors, respectively, used by Stone and Moore (2003). The numbers 1.0, 2.0 and 4.0 plotted adjacent to the 0.2 Hz grid line indicate effective compression ratios.

cated that reducing the modulation depth for rates above 16 Hz had only a very small effect on speech intelligibility. However, Ghitza (2001) pointed out a potential flaw in the studies of Drullman *et al.* Because the temporal fine structure of the channel signals was preserved, the recombination of the channel signals would have reintroduced some of the channel envelope modulations that were supposed to have been removed. Our use of a noise "carrier" for each channel prevented this from happening. A further difference from the work of Drullman *et al.* is that we used a modulated rather than a steady background.

Experiment 1 used multi-channel compression, but, because we wanted to match the effective amount of compression to that for the single-channel compressor, the compression ratios were very low in some of the channels, so the effect on spectro-temporal contrast varied across frequency. Experiment 2 was designed to reduce temporal modulations by the same degree in all channels, so the same compression ratio was used for each channel.

Since multi-channel compression degrades spectro-temporal contrast and the noise vocoder also degrades spectral resolution, there was a need to ensure that the compression processing did not materially affect the intelligibility of the target speech in quiet and thereby limit the performance that could be obtained in the main experiment. Hence, a "reference experiment" was included in which the intelligibility of the target speech was measured in quiet for a subset of the compression speeds tested.

We wished to explore the effect of compression on a "real-world" mixture of signals so the compression was applied after mixing of the target and background (when present).

B. Subjects, speech material and equipment

Two sets of subjects were recruited, six for the reference experiment (3M, 3F, mean=25.7 years, SD=8.2 years, range 21–42 years) and 20 for the main experiment (8M, 12F, mean=23.2 years, SD=5.5 years, range 19–42 years). All were university undergraduates or graduates and were selected on the same basis as for experiment 1. An exception was that one subject in each group was a native speaker of American rather than British English. All subjects were paid for their attendance.

For this experiment we used the IEEE sentences (IEEE, 1969) which are longer and more complex in structure than the ASL or BKB sentences used previously. The target sentences were spoken by a male speaker of British English. The background, when present, was the same male speaker as used for experiment 1. An additional segment of the background talker was used at the start of each list to be processed: this allowed the compressors to stabilize before the target sentences started (see below for details), but was removed before presentation to the subject. The equipment, signal-processing method, input and output levels and presentation method were the same as for experiment 1.

C. Choice of compression method, speed and ratio

To assess the effects of compressing the entire range of modulation frequencies found to be important for speech intelligibility by Drullman *et al.* (1994a; 1994b), the fastest compressor was the very fast envelope compressor used by Stone *et al.* (1999). Although having very short attack and release times, the effective compression produced by this compressor was very small at speech fundamental frequencies (see curve 4 in Fig. 2). This compressor has the highest possible compression speed while avoiding significant distortion of the signal. The slowest compressor was chosen to produce no compression for modulation rates of 2 Hz and above (curve 1 in Fig. 2), so that modulation depth was preserved over the whole range of modulation frequencies proposed to be important by Drullman *et al.* (1994a; 1994b). In our envelope compressor, the envelope is extracted using a near-linear-phase filter, and the audio signal is delayed to compensate for the time delay of the gain signal produced by envelope filtering (Robinson and Huntington, 1973). As a result, the compressor introduces very little asymmetry in the envelope of the compressed signal and, if FES is important, this should lead to greater intelligibility than compressors with short attack times and longer release times (Stone and Moore, 2004).

A condition without compression and four conditions using compressors of different speeds were implemented by scaling the corner frequency of the two-pole low-pass filter used for envelope extraction in the compressor. The corner frequencies were spaced equally on a logarithmic scale at 0.46, 1.71, 6.4 and 24.0 Hz, equivalent to attack/release times of 546, 147, 39 and 10.5 ms, respectively. The required compensating delays for the audio channel were 326.1, 87.7, 23.4 and 6.3 ms, respectively. The longer delay times would be impractical for "real-time" processing systems (Stone and Moore, 2002; 2005), but were used here to avoid distortion of the envelope shape of the signal. The f_r plots of the compressors are shown in Fig. 2 and labeled 1–4 for the slowest to the fastest compressor, respectively.

TABLE IV. Edge frequencies in Hz for the band-pass filters used to create channels in experiments 2 and 3. For experiment 2 (rows marked SII), edge frequencies were chosen to produce an equal contribution to the SII (ANSI, 1997) from each band, and compression channel edge frequencies equaled vocoder channel edge frequencies. For experiment 3, the compression processing used 1, 3, 6, or 12 channels while the noise vocoder used 12 or 18 channels, and edge frequencies had equal spacing on the ERB_N-number scale (Glasberg and Moore, 1990).

Number of channels	Channel edge frequencies (Hz)																		
3	100						725						2539						7800
6	100			331			725			1396			2539			4485			7800
8 (SII)	100				484			759	1093		1530			2111	2920		4142			7800
12 (SII)	100				407		567	759	975	1227	1530	1897	2348	2920	3662		4792			7800
12	100		200	331		502	725		1016	1396		1892	2539		3383	4485		5923		7800
18	100	164	240	331	440	570	725	910	1132	1396	1712	2089	2539	3076	3718	4485	5401	6494		7800

For comparison, Fig. 2 also shows f_r plots for the slow (S) and fast (F) compressors used by Stone and Moore (2003); these systems used a CR of 7. The CR for all systems tested here was 2.78 (f_r=0.64 at low modulation frequencies), the same as used before with the "very fast" envelope compressor in a wideband configuration (Stone and Moore, 2004). The compression threshold was set in the same way as for experiment 1.

D. Choice of number of vocoder channels and target-to-background ratio

Pilot trials with a six-channel vocoder at +8 dB TBR and with an 11-channel vocoder at +3 dB TBR, similar to values used in previous experiments, revealed intelligibility below 50% for all conditions. Although we could have increased the TBR to increase intelligibility, the effect of compression on WSMC and ASMC would then have decreased. Increasing the number of channels was an alternative, but we wished to use channel numbers representative of the effective number of channels in modern cochlear implants, which is typically between eight and 11 (Friesen *et al.*, 2001). We therefore chose to use eight channels at +8 dB and 12 channels at +4 dB TBR, ensuring a reasonable range of channel numbers and effect of compression on WSMC and ASMC. The edge frequencies of the vocoder channels are shown in Table IV (rows marked SII). They were chosen so that each channel would make an equal contribution to intelligibility according to calculations based on the Speech Intelligibility Index (ANSI, 1997).

E. Procedure and training

The subjects for the reference experiment had experienced noise-vocoder processed speech in the preceding 4–18 months, so they attended only a single session, which lasted about 1 h. The subjects for the main experiment were all naïve to the processing and attended two sessions, the first of which was solely dedicated to training and lasted over 1 h. The second, test, session lasted just over 1 h and typically took place about one week (range 1 day–2 months) after the training session. Our previous work has shown that naïve subjects usually give stable responses after training of 1 h, provided that some further training is provided within the session involving data collection.

The training method was similar to that for experiment 1, involving the reporting back of sentences with increasing degree of difficulty as the training session progressed. For the subjects of the main experiment, an additional stage of listening to vocoded prose passages without reporting back was included at the start of their training session. Again, they listened to, and reported back, sentence material produced by two different combinations of speaker and background. This procedure was intended to familiarize the subjects with the general nature of the processing, to a similar degree as for the subjects who had been tested in other experiments with other sentence material. In the testing session, both groups were presented with ten practice sentence lists (100 sentences, 500 keywords) produced by the same target and background speakers as used for data collection. For the reference experiment, all training and testing was performed using sentences in quiet. Each condition was assessed using 20 sentences, giving a maximum score of 100 keywords correct. There were six conditions (no compression and compression speeds 2 and 4 each with two numbers of channel). For the main experiment, there were ten conditions (no compression and four compression speeds each with two channel numbers). For both experiments, the different conditions were tested in a counterbalanced order across subjects.

F. Results and discussion

The scores were transformed into RAUs and checked for a possible training effect across successive conditions using the same procedure as for experiment 1. No training effect was present, for either the reference or main experiment. The mean scores in percent correct are plotted in Fig. 3. The key to the abscissa markings is *N* for no compression and 1–4 for compression speeds 1–4, respectively. For the reference ex-

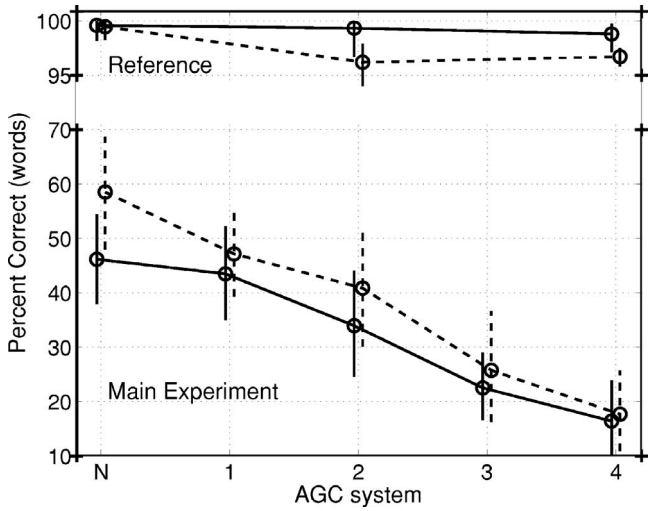


FIG. 3. Results of experiment 2: the dashed lines show results for eight channels and the solid lines show results for 12 channels. Comparisons of performance across channel numbers are not meaningful, as the TBR differed for the eight- and 12-channel systems. The error bars show ±1 SD. Note the broken scale and different values on the ordinate between the results for the reference experiment (upper lines) and main experiment (lower lines). For system *N*, no compression was applied; f_r plots for compression systems 1–4 are shown in Fig. 2.

periment (upper two traces) scores were generally very high, ranging from 96.2 to 99.6%, indicating that neither the linguistic complexity of the sentences nor the compression had a substantial effect on the intelligibility of the target talker in quiet. The SDs are also very small. An ANOVA was conducted on the RAU-transformed scores for the reference experiment with factors speed (no compression, 2 and 4) and number of channels (8 or 12). There was a significant effect of number of channels, $F(1,5)=17.99$, $p<0.008$. The effect of compression speed was not significant, $F(2,10)=3.65$, $p=0.065$, nor was the interaction between number of channels and speed, $F(2,10)=3.95$, $p=0.054$.

For the main experiment (lower two traces in Fig. 3), intelligibility decreased as the compressor speed increased, for both numbers of channels. An ANOVA was conducted on the RAU-transformed scores with factors speed (*N*, 1–4) and number of channels (8 or 12). There was a significant effect of speed, $F(4,76)=149.51$, $p<0.001$, and of number of channels, $F(1,19)=36.67$, $p<0.001$. There was a significant interaction between speed and number of channels, $F(4,76)=3.09$, $p=0.021$. This interaction reflects the fact that the difference between scores for speed *N* (no compression) and speed 1 was not significant for the 12-channel vocoder ($t=1.07$, 76 *df*, $p>0.25$, one tailed), but was significant for the eight-channel vocoder ($t=4.57$, 76 *df*, $p<0.0005$, one tailed). Since speed 1 reduced the envelope modulation depth only for modulation rates below 2 Hz, the significant effect for the eight-channel vocoder leads to the conclusion that envelope modulation rates below 2 Hz contributed to intelligibility. This is surprising, since Drullman *et al.* (1994a) concluded that modulation at rates below 4 Hz did not contribute to intelligibility. The discrepancy may have occurred because the eight-channel vocoder processing provided very limited spectral resolution, while the processing used by Drullman *et al.* provided higher spectral resolution. It appears to be the case that a wider range of modulation rates is useful when spectral resolution is low. This issue is discussed in more detail later on. For all other comparisons of the effect of speed, increasing the speed of the compressor always significantly decreased intelligibility ($t \geq 2.54$, 76 *df*, $p<0.01$, one tailed).

TABLE V. Values of the various measures of envelope distortion for the stimuli used in experiment 2. The bottom row of each section of the table shows the mean intelligibility scores. N indicates the condition with no compression. The numbers 1–4 are labels for the different compression speeds.

		Compression speed			
	N	1	2	3	4
8 channel					
WSMC	0.297	0.296	0.287	0.255	0.238
ASMC	0.008	−0.028	−0.097	−0.162	−0.171
FES	1.00	0.940	0.878	0.880	0.907
% intelligibility	58.5	47.2	40.9	25.8	17.7
12 channel					
WSMC	0.261	0.259	0.244	0.213	0.201
ASMC	0.011	−0.024	−0.087	−0.145	−0.153
FES	1.00	0.936	0.869	0.867	0.892
% intelligibility	46.2	43.5	34.0	22.5	16.4

Table V shows the values of the various envelope measures for each condition tested. Mean intelligibility scores are shown at the bottom of each section of the table. For a given number of channels, the values of WSMC decreased with increasing compressor speed, so, in this case, the pattern of the changes in WSMC is consistent with the intelligibility scores. The values of ASMC also decreased (became more negative) with increasing compressor speed, so they too were consistent with the intelligibility scores. While the values of FES were lower with compression than without compression (condition *N*), the FES values did not decrease monotonically with increasing compressor speed; the values were lowest for compressor speeds 2 and 3. Thus, the pattern of the FES values is not consistent with the intelligibility scores. For this experiment it seems reasonable to assume that the worsening in intelligibility with increasing compressor speed was partly produced by the greater reduction in envelope modulation depth produced by the faster compressors, and partly by decreases in ASMC and perhaps WSMC with increasing compressor speed.

In summary, speech intelligibility worsened markedly with increasing compression speed for both numbers of channels. The results indicate that envelope modulation at rates below 2 Hz contributed to intelligibility for the eight-channel but not for the 12-channel system. The effects of compressor speed were consistent with changes in the measures ASMC and WSMC, but not with FES. The reduction of modulation depth produced by the fast compression may also have contributed to the reduction of intelligibility with increasing compression speed.

IV. EXPERIMENT 3: DETERMINING THE EFFECT OF INDEPENDENTLY VARYING THE NUMBER AND SPEED OF COMPRESSION CHANNELS AND THE NUMBER OF VOCODER CHANNELS

A. Rationale

Cochlear implant users are known to have difficulty discriminating changes in stimulation rate for rates above 300 Hz, indicating that they are insensitive to temporal fine structure for rates above about 300 Hz (Simmons, 1966; Eddington *et al.*, 1978; Fourcin *et al.*, 1979; Moore and Carlyon, 2005). Cochlear implant users also have poor access to spectral cues, since the effective number of channels is usually only eight to 11 (Friesen *et al.*, 2001). Since the noise vocoder removes fine spectral and temporal cues, it provides a reasonable simulation of the information conveyed via a cochlear implant.

The ability to use temporal fine-structure information has also been shown to be reduced in some people with moderate cochlear hearing loss and is often essentially absent for greater losses (Van Tasell *et al.*, 1987; Moore and Moore, 2003; Buss *et al.*, 2004; Lacher-Fougère and Demany, 2005; Rossi-Katz and Arehart, 2005; Lorenzi *et al.*, 2006; Moore *et al.*, 2006; Hopkins and Moore, 2007). With degraded access to temporal fine-structure cues, such people rely more on temporal-envelope and spectral-envelope cues: but even these cues may be poorly transmitted. Temporal cues may be degraded by the compression used to overcome the reduced dynamic range of impaired hearing. Spectral cues may be poorly represented since people with cochlear hearing loss also have reduced frequency selectivity relative to normal hearing (Pick *et al.*, 1977; Zwicker and Schorn, 1978; Glasberg and Moore, 1986). For normally hearing people, the auditory system has about 30 independent channels between 100 and 7800 Hz (Glasberg and Moore, 1990), the frequency span of the processed speech signal used here. Noise-vocoder processing implicitly simulates broadening of the auditory-filter bandwidths; for example, the eight-channel vocoder simulates broadening by a factor of 3.8 relative to normal, a degree of broadening at the upper limit of that found in hearing-impaired subjects (Moore and Glasberg, 2004). The noise vocoder can therefore be seen as a way of simulating the hearing loss of people with little ability to use temporal fine structure: the number of channels chosen determines the spectral resolution, while the use of noise carriers degrades the temporal fine structure information.

In experiment 3 we investigated the effect on speech intelligibility of the number of compression channels and (separately) of the number of vocoder channels. We included systems with higher levels of spectral resolution (more vocoder channels) than in experiment 2, to simulate the spectral resolution of people with moderate-to-severe hearing loss. The speech signals were processed in two distinct stages. First, we used "m" channels of independent multi-channel compression; this was expected to reduce both spectral and temporal modulations, by amounts that increased with increasing number of channels and increasing compression speed. Then, the signal was processed by an "n"-channel noise vocoder ($m \leq n$) which produced a further reduction of spectral information and removed the temporal fine structure of the signal. The value of m was chosen to cover the range typically found in hearing aids.

B. Speech materials, conditions and equipment

The target speech was taken from the same corpus as for experiment 2, using lists that had not been used in that experiment. The background talker was the same as for experiments 1 and 2. Some of the conditions were chosen to be similar to those of experiment 2. Pilot trials indicated that an 18-channel vocoder with a +4 dB TBR and a 12-channel vocoder with a +7 dB TBR would produce mean intelligibility around 60–70%. The 18 and 12 channels give frequency selectivity equivalent to a broadening of the auditory filters by factors of 1.67 and 2.5 relative to normal. These broadenings approximate the loss of selectivity commonly found with moderate and moderate-severe losses, respectively (see Fig. 2.14 of Moore, 1995). Since the vocoder channels were partly intended to simulate the auditory filters in hearing-impaired ears, the edge frequencies of the channels were chosen to give the same width in ERB_N-number for each channel (Glasberg and Moore, 1990). This is not very different from equal SII per channel in the mid-frequency range, but departs from this at low and high frequencies. Table IV shows edge frequencies for each number of channels. The equipment, signal-processing method, input and output levels and presentation method were the same as for experiment 1.

C. The choice of number of compression channels and speed

The number of compression channels used in a hearing aid can be chosen as a compromise between the need to provide flexibility in fitting a range of hearing losses and the need to preserve a reasonable degree of spectral contrast, and thereby intelligibility. Here, one, three, six and 12 channels of compression were chosen. Edge frequencies for the compression channels, shown in Table IV, were chosen to give the same width in number for each channel (Glasberg and Moore, 1990).

The same envelope compressor as in experiment 2 was used here, but corner frequencies of 0.18, 0.45, 1.13 and 2.81 Hz were chosen, equivalent to attack/release times of 1395, 558, 223 and 89 ms, respectively. The required compensating delays for the audio channel were 833, 333, 133 and 53 ms, respectively (again, not suitable for real-time implementation). With a smaller ratio between each speed (of 2.5) than for experiment 2, we were able to investigate more closely the range of conditions where intelligibility starts to be affected by the multi-channel compression. The slowest compressor was similar to the slow compressor of Stone and Moore (2003), which produced intelligibility not significantly different from that without compression for noise vocoders with eight, 11 or 16 channels. The compression system with a corner frequency of 0.45 Hz was very similar to that for speed 1 of experiment 2. The fastest compressor was chosen to be similar to the fast (single-channel) compressor of Stone and Moore (2003), which consistently

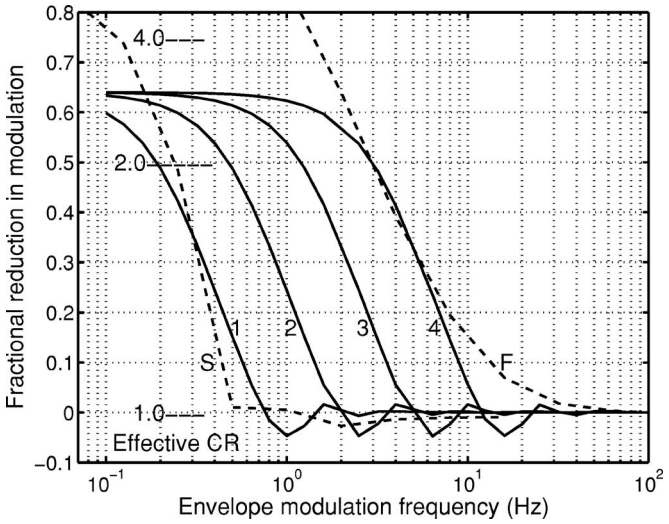


FIG. 4. Plots of f_r versus modulation frequency for the four compressor speeds used in experiment 3 (solid lines, labeled 1–4). For comparison purposes, the dashed lines marked S and F are for the slow and fast compressors, respectively, used by Stone and Moore (2003).

reduced intelligibility by about 5% relative to no or slow compression, independent of the number of noise-vocoder channels. The compression ratio was 2.78 for all channels and the compression threshold for each channel was 13 dB below the rms level in that channel.

The f_r plots for the four compressor speeds are shown in Fig. 4 and labeled 1–4 for the slowest to the fastest compressors, respectively, again with the equivalent traces for the slow (S) and fast (F) compressors of Stone and Moore (2003). Following the application of compression in each of the m channels, the channel signals were combined to create the signal that was used as input to the n-channel vocoder.

D. Subjects, training and procedure

Thirty-two subjects (16 M, 16 F, mean=22.5 years, SD =4.7 years, range 19–42 years), all university undergraduates or graduates, were selected on the same basis as for experiments 1 and 2. One was a native speaker of American English. All subjects were paid for their attendance. Subjects fell into three categories in terms of their experience with noise-vocoder processing. Those in the first category were totally naïve to the processing and attended a 1 h training session prior to testing. Those in the second category had prior experience of the processing, but had not taken part in experiment 2. Those in the third category had taken part in experiment 2. In the first data-collection session, all subjects received further training, using sentences materials of increasing difficulty, as described earlier. The training lasted about 30 min for the first two groups and 15–20 min for the third group. All subjects were tested in a randomized design spread across two separate data-collection sessions, each of about 1 h duration. The interval between sessions ranged from a few hours to a few weeks.

Each condition was assessed using ten sentences, giving a maximum score of 50 keywords correct per subject. The experiment involved 32 different conditions (four numbers of compression channels, m, by four speeds of compression by two numbers of spectral resolution, n). To reduce fatigue, 12 conditions were assessed in the first data-collection session and the remaining 20 in the second. The order of testing of conditions was counterbalanced across subjects.

E. Results

No training effect was present in the time-ordered results. Mean scores in percent correct (and SDs) are given in Table VI. The general pattern for both channel numbers was that, as the compressor speed or the number of compression channels increased, intelligibility decreased. For the 18-channel vocoder, there was no clear effect of compression speed for the single-channel compressor, although the worst performance did occur for the fastest speed. Similarly, there was no clear effect of number of channels for speed 1.

An ANOVA was conducted on the RAU-transformed scores with factors number of vocoder channels (12 or 18), number of compressor channels (1, 3, 6 or 12) and compressor speed (1–4). There were significant effects of number of vocoder channels, $F(1,31)=155.04, p<0.001$, number of compression channels, $F(3,93)=8.54, p<0.001$, and compressor speed, $F(3,93)=40.68, p<0.001$. The significant ef-

TABLE VI. Scores in percent correct for experiment 3 for each combination of compressor speed and number of compression channels. Part A is for the 12-channel vocoder and part B is for the 18-channel vocoder. Numbers in parentheses are SDs. Standard errors would be a factor of 5.6 smaller.

	Compressor speed			
	1	2	3	4
A. 12-channel vocoder				
Number of compression channels				
12	66.9 (10.2)	59.4 (11.0)	57.4 (11.7)	48.9 (11.1)
6	64.2 (11.5)	65.3 (10.4)	57.2 (12.9)	51.0 (12.3)
3	65.4 (13.2)	61.6 (16.0)	61.8 (12.6)	56.0 (12.6)
1	65.2 (13.6)	65.9 (11.7)	64.4 (13.2)	59.3 (12.2)
B. 18-channel vocoder				
Number of compression channels				
12	52.2 (11.4)	53.2 (9.8)	48.9 (11.1)	40.3 (12.7)
6	54.1 (12.3)	52.1 (10.8)	51.7 (12.6)	44.6 (11.2)
3	56.8 (12.7)	53.9 (10.7)	52.2 (12.4)	46.2 (13.6)
1	53.6 (15.1)	57.9 (14.2)	54.6 (14.3)	51.9 (11.2)

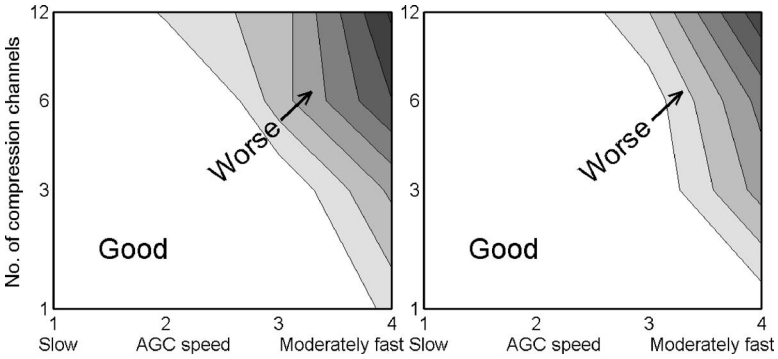


FIG. 5. Results of experiment 3 represented as contours of equal intelligibility. The left-hand panel shows results for the 12-channel vocoder and the right-hand panel shows results for the 18-channel vocoder. The abscissa shows the compressor speed, as defined in Fig. 4, while the ordinate shows the number of channels of independent compression (logarithmically spaced).

fect of number of vocoder channels is not meaningful since the TBR was different for each number of channels. The interaction between number of compression channels and compressor speed just failed to reach significance, $F(9,279) = 1.83, p = 0.063$.

The results in Table VI are represented as contours of equal intelligibility in Fig. 5; contours were obtained using the MATLAB "contourf" function. The left- and right-hand panels show results for the 12- and 18-channel vocoders, respectively. The abscissa shows the compressor speed, while the ordinate shows the number of channels of compression (on a logarithmic scale). In each panel, a "white" area indicates good performance; its edge is defined by a contour referenced to (5.5% below) the average of the scores for three conditions for which the compression would be expected to have little effect on intelligibility, namely speeds 1 and 2 for one-channel compression, and speed 1 for three-channel compression. Darker areas indicate poorer performance, as indicated by the arrows marked "Worse." The first three contour lines approximate the differences from the reference contour required to achieve changes in performance at significance levels of $p < 0.025$, $p < 0.005$ and $p < 0.0005$ (one tailed). The contours are spaced in steps of 1.8% in terms of intelligibility. It is clear that performance worsens when either the number of compression channels or the compression speed is increased, and that the worst performance is obtained with the largest number of channels and the highest speed.

Figure 6 shows contour plots for the various envelope measures, in a similar format to the contour plots of intelligibility in Fig. 5. The contour plots for WSMC show a very different pattern to the contour plots for intelligibility; for WSMC, the values increase (improve) with increasing number of compression channels, while the intelligibility scores decrease. Thus, for these results, WSMC was probably not the main factor affecting intelligibility. The contour plots for both FES and ASMC are similar in form to the contour plots for intelligibility, indicating that either of these two aspects of the envelope could have affected intelligibility. However, the slopes of the contour lines are different from those found for intelligibility (Fig. 5), suggesting that FES and ASMC do not fully account for the effects of compressor speed and number of channels on intelligibility. Changes in effective envelope modulation depth also probably played a role.

V. DISCUSSION

A. Disruption of cues to segregation

Previously, we showed that applying single-channel fast-acting compression after mixing the speech from two talkers with similar levels led to lower intelligibility than applying compression before mixing (Stone and Moore, 2004). That

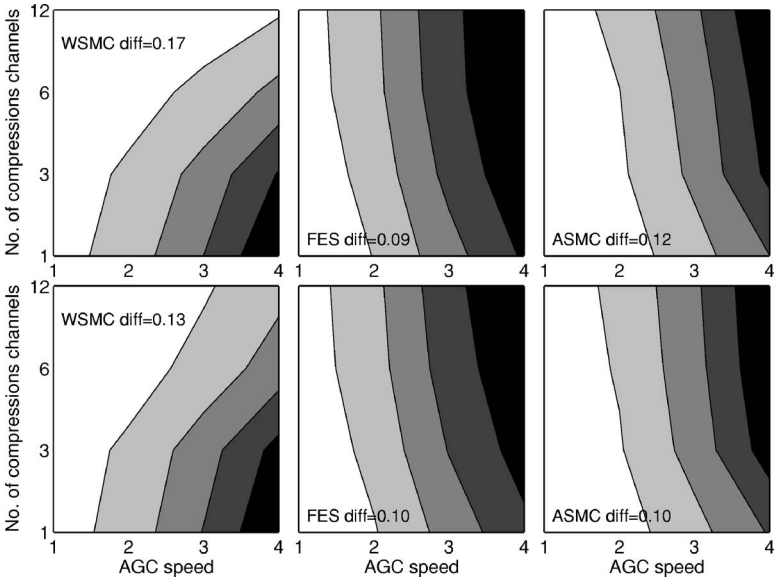


FIG. 6. Contour plots of the measures WSMC, FES and ASMC, plotted in a similar format to that for Fig. 5. The upper row shows plots for the 12-channel vocoder and the lower row shows plots for the 18-channel vocoder. The "diff" values indicate the difference between the highest and lowest values in each contour plot. The contours are spaced at 0.2 times the "diff" values.

effect was attributed to ASMC; the idea was that the previously independent sources acquired a common component of modulation, and that this made it more difficult to use across-frequency similarities and differences in modulation pattern to group together parts of the signal belonging to one source, and segregate them from parts belonging to the other source. In principle, the difference between the two cases might have been partly due to differences in WSMC or FES across the two conditions. However, Stone and Moore (2007) showed that the values of WSMC and FES were both lower when the signals were compressed before mixing than when they were compressed after mixing; these changes are inconsistent with the changes in intelligibility. Therefore, the effects in our earlier experiment can be attributed mainly to ASMC.

In experiment 1 of the present paper, we used multi-channel fast compression, with each compression channel corresponding to one vocoder channel, and showed that intelligibility was better when the compression was applied before mixing than when it was applied after mixing. This pattern of results was reflected in the ASMC values which were negative for condition AFTER. However, the WSMC values were higher for condition AFTER than for condition BEFORE, while the FES values were almost the same for the two conditions. Therefore, it seems likely that the poorer intelligibility in condition AFTER can mainly be attributed to changes in ASMC. Across all the experiments reported here, the only measure that consistently showed the same pattern as the intelligibility scores was ASMC. Thus, while we cannot rule out a role for WSMC and FES under some conditions, it seems that the negative ASMC introduced by applying compression to a mixture of speech from a target and a background talker is usually associated with reduced intelligibility.

Whatever the underlying mechanism, it seems clear that either single-channel or multi-channel fast compression can have deleterious effects on speech intelligibility when applied to mixtures of sounds and when there is a heavy reliance on envelope cues, as is the case for users of cochlear implants and users of hearing aids with severe or profound hearing loss.

B. Spectro-temporal trade-offs in multi-channel compressor performance

Xu *et al.* (2005) investigated the perception of consonants and vowels in quiet as they systematically varied the envelope cutoff frequency and the number of channels in a noise vocoder, using normal-hearing listeners. Good performance in consonant perception required high temporal resolution (high envelope cutoff frequencies) but only modest spectral resolution (small number of channels). The converse was true for vowel perception. Their results showed separately for both vowels and consonants that there were large areas in the spectro-temporal domain where similar levels of performance could be obtained. For a given level of performance, temporal resolution could be decreased provided that there was a compensating increase in spectral resolution. Although we investigated the intelligibility of sentences in the presence of a background talker, the results of experiment 3, as illustrated by the contour plots in Fig. 5, can be interpreted in a similar way. The effect of increasing compressor speed is to reduce the envelope modulation depth over a greater range of modulation rates, extending from lower towards higher rates. The effect of increasing the number of compression channels is to reduce spectral contrast. The fact that the contours of constant performance have a downward slope indicates that reduction of envelope modulation produced by fast-acting compression can be compensated by decreasing the number of compression channels, thus preserving spectral contrast. A measure of the trade-off between spectral and temporal information when using multi-channel compression can be derived from the slope of a typical constant-performance contour in Fig. 5. If the number of compression channels is doubled, then the compression speed needs to be reduced by a factor of about 0.63 to maintain constant performance. This trading relationship is the same for both numbers of vocoder channels.

A practical consequence of this is that, if a person with a cochlear implant has a relatively small number of effective channels, it is important to preserve temporal resolution. This may mean that less compression should be used in the mapping of signal levels to currents applied to the electrodes, and/or that slow-acting compression should be used. The results of experiment 2 showed that modulations at rates below 2 Hz contributed to intelligibility, but only for the smaller of the two channel numbers used (i.e., eight). Eight effective channels are commonly encountered in cochlear implants and listeners with profound hearing impairment (Friesen *et al.*, 2001; Faulkner *et al.*, 1990).

C. Implications for the design of hearing prostheses

The performance of compression systems for use in hearing prostheses is affected by many factors, including the characteristics of the input signals (spectral and temporal variability and dynamic range), whether or not competing signals and reverberation are present, and the characteristics of the user (absolute thresholds, dynamic range, frequency selectivity, and so on). Not surprisingly, the literature is conflicting about the aspects of compression (speed, number of channels, compression threshold, etc.) that lead to the "best" performance (Moore, 1990, 2007; Dillon, 1996; Souza, 2002). For people with moderate hearing loss, fast-acting compression can lead to better performance than linear amplification (Laurence *et al.*, 1983; Moore *et al.*, 1992), even when the number of compression channels is large (Yund and Buckles, 1995a; 1995b). Also, performance with multi-channel fast-acting compression has been found to be comparable to that obtained using single-channel slow-acting compression (Stone *et al.*, 1999). However, for people with severe to profound hearing loss, fast-acting compression does not always lead to better performance than linear amplification, and sometimes it leads to worse performance (De Gennaro *et al.*, 1986; Boothroyd *et al.*, 1988; Drullman and Smoorenburg, 1997). It appears that improving the audibility of low-level speech sounds does not always lead to increased intelligibility.

As described earlier, it seems likely that people with severe to profound hearing impairment have a reduced abil-

ity to use temporal fine structure cues, and hence rely heavily on the spectro-temporal envelope of speech. Under these conditions, degradation of the spectro-temporal envelope by the use of multi-channel compression may adversely affect speech intelligibility, especially when fluctuating background sounds are present. Fast-acting multi-channel compression may work well for people with mild to moderate hearing loss, but as the degree of hearing impairment increases, or as the necessity to hear in the presence of background noise increases, the use of slower-acting compression may be necessary in order to maintain intelligibility. At the same time, short-term audibility may have to be sacrificed.

VI. CONCLUSIONS

In all three experiments, the intelligibility of speech from a target talker was measured in the presence of speech from a background talker, and all stimuli were noise vocoded so that information was conveyed mainly by envelope cues in a limited number of frequency channels. Experiment 1 showed that multi-channel fast-acting compression (with one compression channel for each vocoder channel) led to worse performance when the compression was applied after mixing the speech from the two talkers than when it was applied before mixing. This resembles the effect found earlier for single-channel compression, which was attributed to the compression introducing ASMC when applied after mixing (Stone and Moore, 2004). The effect found here also appears to be attributable to ASMC. When the compression was applied after mixing the speech from the two talkers, the values of ASMC were negative, potentially making it more difficult to perceptually separate the speech of the target and background talkers.

Experiment 2 used multi-channel compression applied after mixing the target and background speech, again with one compression channel for each vocoder channel, and showed that intelligibility worsened progressively with increasing compression speed. For the eight-channel system, even the slowest compressor used led to lower intelligibility than for no compression. This result indicates that envelope modulation at rates below 2 Hz contributes significantly to intelligibility when spectral resolution is limited. The pattern of intelligibility changes was consistent with the changes in WSMC and ASMC, but not with the changes in FES.

Experiment 3 used multi-channel compression applied after mixing the target and background speech, but the number of compression channels (m) was varied independently of the number of vocoder channels (n, where $m \leq n$). The results showed that intelligibility worsened with increasing compressor speed and with increasing number of compression channels. Contour plots of the intelligibility scores resembled contour plots of the FES values and ASMC values, but differed from contour plots of the WSMC values.

It is suggested that, for people with a limited ability to use temporal fine structure and with limited frequency resolution (i.e., users of cochlear implants and hearing-aid users with severe to profound hearing loss), the amount and speed of compression, and the number of compression channels should be limited, so as to avoid loss and distortion of information contained in the patterns of spectral and temporal modulation of speech.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). We thank Brian Glasberg for assistance with statistical head banging. We also thank Ken Grant and two anonymous reviewers for helpful comments on an earlier version of this paper.

ANSI (**1997**). *ANSI S3.5-1997, Methods for the Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in the presence of noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Boothroyd, A., Springer, N., Smith, L., and Schulman, J. (**1988**). "Amplitude compression and profound hearing loss," J. Speech Hear. Res. **31**, 362–376.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge, MA).

Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. (**1985**). "Spectral integration based on common amplitude modulation," Percept. Psychophys. **37**, 483–493.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Carrell, T. (**1993**). "The effect of amplitude comodulation on extracting sentences from noise: Evidence from a variety of contexts," J. Acoust. Soc. Am. **93**, 2327.

Carrell, T. D., and Opie, J. M. (**1992**). "The effect of amplitude comodulation on auditory object formation in sentence perception," Percept. Psychophys. **52**, 437–445.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2731.

De Gennaro, S., Braida, L. D., and Durlach, N. I. (**1986**). "Multichannel syllabic compression for severely impaired listeners," J. Rehabil. Res. Dev. **23**, 17–24.

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.

Drullman, R., and Smoorenburg, G. F. (**1997**). "Audio-visual perception of compressed speech by profoundly hearing-impaired subjects," Audiology **36**, 165–177.

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.

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. (**1978**). "Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers," Trans. Am. Soc. Artif. Intern. Organs **24**, 1–5.

Faulkner, A., Rosen, S., and Moore, B. C. J. (**1990**). "Residual frequency selectivity in the profoundly hearing-impaired listener," Br. J. Audiol. **24**, 381–392.

Fourcin, A. J., Rosen, S. M., Moore, B. C. J., Douek, E. E., Clark, G. P., Dodson, H., *et al.* (**1979**). "External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings," Br. J. Audiol. **13**, 85–107.

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.
Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.
Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.
Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.
Laurence, R. F., Moore, B. C. J., and Glasberg, B. R. (**1983**). "A comparison of behind-the- ear high-fidelity linear aids and two-channel compression hearing aids in the laboratory and in everyday life," Br. J. Audiol. **17**, 31–48.
Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.
Moore, B. C. J. (**1990**). "How much do we gain by gain control in hearing aids?," Acta Oto-Laryngol., Suppl. **469**, 250–256.
Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, Oxford).
Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues* (Wiley, Chichester).
Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York).
Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.
Moore, B. C. J., and Moore, G. A. (**2003**). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," Hear. Res. **182**, 153–163.
Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure information," Hear. Res. **222**, 16–27.
Moore, B. C. J., Johnson, J. S., Clark, T. M., and Pluvinage, V. (**1992**). "Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss," Ear Hear. **13**, 349–370.
Pick, G., Evans, E. F., and Wilson, J. P. (**1977**). "Frequency resolution in patients with hearing loss of cochlear origin," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans, and J. P. Wilson (Academic Press, London).
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.
Rappold, R. W., Mendoza, L., and Collins, M. J. (**1993**). "Measuring the strength of auditory fusion for synchronously and non-synchronously fluctuating narrow-band noise pairs," J. Acoust. Soc. Am. **93**, 1196–1199.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.
Robinson, C. E., and Huntington, D. A. (**1973**). "The intelligibility of speech processed by delayed long-term averaged compression amplification," J. Acoust. Soc. Am. **54**, 314.
Rossi-Katz, J. A., and Arehart, K. H. (**2005**). "Effects of cochlear hearing loss on perceptual grouping cues in competing-vowel perception," J. Acoust. Soc. Am. **118**, 2588–2598.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Simmons, F. B. (**1966**). "Electrical stimulation of the auditory nerve in man," Arch. Otolaryngol. **84**, 2–54.
Souza, P. E. (**2002**). "Effects of compression on speech acoustics, intelligibility, and sound quality," Trends Amplif. **6**, 131–165.
Steeneken, H. J. M., and Houtgast, T. (**1999**). "Mutual dependence of the octave-band weights in predicting speech intelligibility," Speech Commun. **28**, 109–123.
Stone, M. A. (**1995**). "Spectral enhancement for the hearing impaired," Ph.D. thesis, University of Cambridge, England.
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., and Moore, B. C. J. (**2002**). "Tolerable hearing-aid delays. II. Estimation of limits imposed during speech production," Ear Hear. **23**, 325–338.
Stone, M. A., and Moore, B. C. J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am. **114**, 1023–1034.
Stone, M. A., and Moore, B. C. J. (**2004**). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," J. Acoust. Soc. Am. **116**, 2311–2323.
Stone, M. A., and Moore, B. C. J. (**2005**). "Tolerable hearing-aid delays: IV. Effects on subjective disturbance during speech production by hearing-impaired subjects," Ear Hear. **26**, 225–235.
Stone, M. A., and Moore, B. C. J. (**2007**). "Quantifying the effects of fast-acting compression on the envelope of speech," J. Acoust. Soc. Am. **121**, 1654–1664.
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.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelope smearing on speech reception. I," J. Acoust. Soc. Am. **91**, 2872–2880.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Effect of spectral envelope smearing on speech reception. II," J. Acoust. Soc. Am. **93**, 1547–1552.
Turner, C. W., Chi, S.-L., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **42**, 773–784.
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.
Verschuure, J., Maas, A. J. J., Stikvoort, E., de Jong, R. M., Goedegebure, A., and Dreschler, W. A. (**1996**). "Compression and its effect on the speech signal," Ear Hear. **17**, 162–175.
Villchur, E. (**1973**). "Signal processing to improve speech intelligibility in perceptive deafness," J. Acoust. Soc. Am. **53**, 1646–1657.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
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.
Zwicker, E., and Schorn, K. (**1978**). "Psychoacoustical tuning curves in audiology," Audiology **17**, 120–140.

Stop consonant voicing and intraoral pressure contours in women and children[a)]

Laura L. Koenig[b)]
Haskins Laboratories, New Haven, Connecticut 06511 and Long Island University, Brooklyn, New York 11201

Jorge C. Lucero[c)]
Department of Mathematics, University of Brasilia, Brasilia, Brazil

(Received 19 February 2007; revised 30 November 2007; accepted 4 December 2007)

Previous authors have established that stop consonant voicing is more limited in young children than adults, and have ascribed this to immature vocal-tract pressure management. Physical development relevant to speech aerodynamics continues into adolescence, suggesting that consonant voicing development may also persist into the school-age years. This study explored the relationship between stop consonant voicing and intraoral pressure contours in women, 5 year olds, and 10 year olds. Productions of intervocalic /p b/ were recorded from eight speakers at each age. Measures were made of stop consonant voicing and δ, a measure designed to characterize the time course of intraoral pressure increase in stops, following Müller and Brown [*Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic, Madison, 1980), Vol. **4**, pp. 318–389]. Age effects for stop consonant voicing and δ were not statistically significant, but correlations between δ and stop voicing were less often significant and sometimes reversed in the children, providing some evidence of immature aerodynamic control. The current data, as well as those of Müller and Brown, also show that the δ measure may yield some paradoxical values, indicating that more work is needed on methods of assessing time-varying characteristics of intraoral pressure.
 [DOI: 10.1121/1.2828065]

PACS number(s): 43.70.Aj, 43.70.Ep, 43.70.Gr [BHS] Pages: 1077–1088

I. INTRODUCTION

A. Purpose

This paper investigates stop consonant voicing and intraoral pressure (P_{io}) trajectories in normal female and child speakers of American English. Although previous authors have speculated that developmental voicing patterns reflect limited aerodynamic control in children, no studies have directly compared voicing and pressure data from child speakers. Moreover, data on time-varying characteristics of P_{io} are very limited in adults as well as children.

The next section reviews the literature on stop consonant voicing and its relationship to P_{io}, focusing particularly on the work of Müller and Brown (1980). These authors introduced methods of assessing P_{io} trajectories and used aerodynamic modeling to explore how P_{io} signals might reflect underlying articulation. Many of their analysis methods have been adopted in the current work. The subsequent section reviews the developmental literature and considers reasons why one may expect differences in voicing and P_{io} control between children and adults.

B. Voicing of stop consonants and P_{io} control

In many languages, one class of stop consonants is produced with vocal-fold vibration during most or all of the closure. Such stops are typically called voiced, in contrast to voiceless stops, which have silent closures. When the vocal tract is closed for an oral stop, P_{io} builds up behind the occlusion. Since subglottal pressure (P_{sub}) does not vary greatly over the time scale of individual speech segments (Löfqvist, 1975; McGlone and Shipp, 1972; Netsell, 1969), this increase in P_{io} yields a reduced transglottal pressure (P_{trans}) differential. Diminishing P_{trans} can ultimately lead to cessation of voicing, even when the vocal folds are adducted. Voicing will persist during a consonantal closure (indeed, during any speech sound) only so long as P_{trans} remains above a threshold value (e.g., Ishizaka and Matsudaira, 1972; Lindqvist, 1972; Stevens, 1991; Titze, 1988).

Past work has documented several ways in which adult speakers increase supraglottal volumes during voiced stop closures, slowing the buildup of P_{io} and allowing phonation to last longer. Volume-expanding maneuvers may include lowering the larynx, elevating the soft palate, advancing the tongue root, depressing the tongue body, and/or expanding the cheeks or lateral pharyngeal walls (Bell-Berti, 1975; Kent and Moll, 1969; Perkell, 1969; Riordan, 1980; Svirsky *et al.*, 1997; Westbury, 1983). The degree to which tissues passively deform in response to P_{io} can also be altered: reducing the level of muscular contraction increases tissue compliance (Westbury, 1983). The specific methods of ex-

[a)]Portions of this work were presented at the first Pan-American/Iberian meeting on Acoustics, Cancun, Mexico.
[b)]Electronic mail: koenig@haskins.yale.edu
[c)]Electronic mail: lucero@unb.br

panding supraglottal volume during a given stop closure appear to vary both within and across speakers (Bell-Berti, 1975; Westbury, 1983).

Müller and Brown (1980) conducted an influential study on P_{io} trajectories during stop closures and the underlying vocal-tract conditions. These authors collected intraoral pressure data during oral stops produced by five men, and evaluated the data both qualitatively and quantitatively. The qualitative analysis involved visually classifying each token by shape, primarily in an attempt to indicate whether pressure pulses showed a fast initial rise (convex tokens) or had more slowly rising trajectories (concave and linear tokens).[1] Schematic examples of concave and convex tokens are shown in Fig. 2 (Sec. II F). The results indicated that voiceless stops were more likely to have convex contours, whereas voiced stops more typically had linear or concave pressure signals. For the quantitative analysis, Müller and Brown calculated a δ measure as the difference between the slopes to average and peak pressures (Pa and Pk). Cases of $\delta > 0$ should be associated with a convex pressure contour; $\delta < 0$ would correspond to a concave contour, and $\delta = 0$ would indicate a linear contour, i.e., no slope change. These results showed higher δ values for voiceless consonants than voiced. In short, both the qualitative and quantitative analyses indicated that the initial P_{io} rise is slower in voiced stops than voiceless.

Müller and Brown (1980) next adapted the aerodynamic model of Rothenberg (1968) to investigate the articulatory conditions that could produce the slow-rising pressure traces observed in the men's voiced stops. The simulations showed that modeling relaxed (compliant) vocal-tract walls and adjusting the timing and magnitude of glottal opening affected the rate of P_{io} buildup somewhat, but did not produce concave trajectories. Only expanding supraglottal volumes resulted in concave P_{io} patterns. Westbury (1983), also working from Rothenberg's model, estimated that oral closures produced with tense vocal-tract walls would have voicing for only about 10 ms, whereas lax walls produced voicing for 80 ms or longer. Modeling active volume increases permitted voicing to continue "virtually indefinitely" for tense as well as lax wall characteristics (Westbury, 1983, p. 1333). Taken together, these studies imply that (a) stop consonant voicing is facilitated by supraglottal volume increases; (b) the effects of such volume changes may be evident in P_{io} trajectories; and (c) P_{io} contours can therefore provide insight into the degree to which speakers manage vocal tract pressures in service of maintaining stop consonant voicing.

The above-outlined experimental and modeling studies considered only adult speakers. As detailed in Sec. I C, there are several reasons to suspect that the results of this literature do not accurately characterize the speech production patterns of children.

C. Child–adult differences

Children differ from adults not only in absolute size, but also in the relative dimensions of their speech production systems and in aerodynamic quantities. Respiratory system variables such as lung volumes, capacities, and recoil pressures do not reach adult-like levels until adolescence (de Troyer *et al.*, 1978; Hoit *et al.*, 1990; Mansell *et al.*, 1977; Polgar and Weng, 1979). Laryngeal proportions in children are more similar to those of women than men; prior to puberty, sex differences in the larynx are minimal, but pubertal changes in males lead to disproportionate increases in vocal fold length and mass (Hirano *et al.*, 1983; Kahane, 1982). Differentiation of the laryngeal tissue layers continues into adolescence in both sexes, and growth of the vocal folds, both during childhood and at puberty in males, adds length mainly in the anterior, membranous region, which has lower stiffness characteristics than the posterior, cartilaginous region (Hirano *et al.*, 1983; Titze, 1989). As for the vocal tract, high laryngeal positions in early childhood yield shorter overall vocal tract lengths and particularly small pharyngeal cavities (Bosma, 1975; Crelin, 1987; Fitch and Giedd, 1999; Goldstein, 1980; Sasaki *et al.*, 1977; Vorperian *et al.*, 2005).

Given such differences in respiratory, laryngeal, and supralaryngeal systems, there is little reason to expect that the transglottal pressures needed to maintain vocal-fold vibration should be the same in children and adults, or that the mechanisms adults use to prolong stop consonant voicing would be equally effective for a child. Lucero and Koenig (2005) recently observed that scaling a laryngeal model down to a size appropriate for a 5 year old increased the phonation threshold pressure, as a result of less glottal tissue area absorbing the aerodynamic energy that fuels oscillation. Several past studies have found higher subglottal pressures in children compared to adults (Arkebauer *et al.*, 1967; Bernthal and Beukelman, 1978; Brown, 1979; Netsell *et al.*, 1994; Stathopoulos and Sapienza, 1993; Stathopoulos and Weismer, 1985; Subtelny *et al.*, 1966). Some researchers have pointed out that the increased resistance afforded by children's smaller airways will, all else being equal, lead to higher pressures (Stathopoulos and Sapienza, 1993; Stathopoulos and Weismer, 1985). Stathopolous and Weismer (1985) also proposed that children may simply select higher speaking intensities than adults. Whether higher pressures in children arise as a matter of anatomy or choice, they may be, to some extent, necessary in order to generate the requisite transglottal pressures for achieving phonation in a child-sized larynx (Lucero and Koenig, 2005).

Higher subglottal pressures notwithstanding, the developmental data suggest that young children do not produce many fully voiced (or prevoiced) stops. Early studies of voice onset time (VOT) by Preston and colleagues (Preston *et al.*, 1968; Preston and Port, 1969) indicated that toddlers from both English- and Arabic-speaking homes produced mainly voiceless unaspirated stops in syllable-initial position. In Arabic, the voicing contrast is between voiced and voiceless unaspirated stops (Yeni-Komshian *et al.*, 1977). In English, the contrast in utterance-initial, prestressed position is mainly between voiceless unaspirated /b d g/ and voiceless aspirated /p t k/; the "voiced" series may have optional closure voicing in this context for some speakers (e.g., Flege, 1982; Flege and Brown, 1982; Lisker and Abramson, 1964). Kewley-Port and Preston (1974), analyzing voicing acquisition in three young English-learning children, also found mostly voiceless unaspirated stops, and argued that this class

of stops is, in effect, the easiest to produce because it requires neither volume-compensation movements to prolong closure voicing nor precise laryngeal-supralaryngeal timing to achieve aspiration. Zlatin and Koenigsknecht (1976), studying 2 and 6 year olds, and Barton and Macken (1980), studying 4 year olds, similarly noted a rarity of prevoiced stops in their data.

Stop consonant voicing patterns in noninitial positions provide further support for age-related differences. Smith (1979) found that 2- and 4-year-old English-learning children devoiced final /b d g/ both more extensively (as a percentage of closure duration) and frequently (as a percentage of the total number produced) than did adults. Allen (1985) observed that young children learning French (which contrasts voiced and voiceless unaspirated stops) tended to produce words beginning with /b d g/ in intervocalic rather than utterance-initial positions. The results of both of these studies are consistent with Westbury and Keating (1986), who used modeling to investigate how well aerodynamic considerations predicted cross-linguistic patterns of stop consonant voicing. The simulations showed that conditions for voicing were most favorable in intervocalic position, and least so in utterance-final position, where subglottal pressure often decreases (e.g., Atkinson, 1978; Gelfer *et al.*, 1983; Lieberman, 1967). Westbury and Keating concluded that "articulatory ease," defined as the likelihood of a stop being voiced in the absence of compensatory maneuvers, accounted for some, but not all, of the cross-linguistic data: Fully voiced stops do not appear to be particularly rare in languages, as the simulations suggested, but the contrast is often neutralized in final position, as predicted. The authors did note, however, that issues of "naturalness" or "case of production" might exert a greater influence on the speech of children than adults.

D. Research goals

Müller and Brown (1980) explicitly proposed that children's P_{io} trajectories might differ systematically from those of adults. Other authors (viz., Kewley-Port and Preston, 1974) have implicated P_{io} control in explaining developmental voicing data. Yet no studies have used time-varying measures to characterize P_{io} in children. Indeed, dynamic measures of speech aerodynamics are rare in adults as well. Finally, few studies, and none on children, have assessed the relationship between P_{io} and the actual degree of stop consonant voicing.

Thus, the purpose of this work was to determine whether children differ from women in their stop consonant voicing behavior and intraoral pressure characteristics as assessed using Müller and Brown's (1980) δ measure. Women were used as the comparison group because their laryngeal and vocal tract proportions are more similar to those of children than are men's. Anatomical and aerodynamic considerations lead to the prediction that children should produce less closure voicing than adults. Past empirical work has established this for children as old as 6 years of age (Zlatin and Koenigsknecht, 1976). Reduced differentiation of children's stops according to voicing would suggest less P_{io} differentiation as well. On the other hand, children may use articulatory maneuvers (e.g., more extreme laryngeal lowering, pharyngeal expansion, velar elevation, etc.) to achieve closure voicing patterns similar to those of adults.

II. METHODOLOGY

A. Speakers

Data were recorded from eight speakers in each of three groups: 4- to 5-year-old children (age range=4;2–5;11, μ=5;2), 9- to 10-year-old children (range=9;1–10;6, μ=9;9), and women (range=32;4–46;5, μ=39;5). For simplicity, the two child groups will henceforth be referred to as 5 year olds and 10 year olds. The children's age groups were chosen to include speakers (a) as young as possible given the requirements of the experimental task, and (b) as old as possible without introducing effects of puberty. The child groups each had four girls and four boys. The adult group was composed of mothers of child participants.

All children and women were normal, healthy, native monolingual speakers of American English from the New York City metropolitan region. None of the speakers had a strong regional accent. Children were required to have a normal birth and developmental history, and no speech or language intervention (current or past). The children also passed a hearing screening to establish binaural thresholds of 25 dB or less at 500, 1000, 2000, 4000, and 8000 Hz. The women had normal speech, language, and hearing by self-report, and speech characteristics within normal limits as judged informally by the first author. Parents filled out a questionnaire on their child's developmental milestones, including, for the older age group, questions about pubertal changes. None of the children included here evidenced onset of puberty. Finally, a short narrative or conversational sample was obtained from all children. These samples were analyzed to ensure that (a) any sound substitutions were developmentally appropriate for the child's age, and (b) the child's language development was within normal limits for his/her age. Adults and parents of child subjects provided informed consent for study participation, and all children provided assent.

In a short post-hoc study, two normal adult male native speakers of American English (unrelated to the children) were recorded using the same instrumentation and methods. Their data were analyzed to verify that small differences between the current methods and those of Müller and Brown (1980) did not systematically influence the results.

B. Instrumentation

Three signals were recorded from each participant: (a) acoustics, obtained using an AKG C420 head-mounted microphone hung around the speaker's neck; (b) oral airflow, collected using an undivided pneumotachograph (Glottal Enterprises) sized appropriately for the speaker; and (c) intraoral air pressure, obtained via a catheter-tip pressure transducer (Gaeltek CT/S). This device has a sensitivity of 5 μV/volt/mm Hg and a frequency response of 1 kHz. The sensor was screwed tightly within a piece of sterile medical tubing inserted through a hole in the airflow mask to rest inside the speaker's oral cavity during bilabial closure. The acoustic signal was used to verify that the speaker's utter-

ances were perceptually accurate (see Sec. II D); all measures were made from the aerodynamic signals.

Data were collected using a PowerLab (4SP) connected to a laptop computer. Before recording began, all inputs were adjusted to yield adequate signal-to-noise levels. Acoustic signals were low-pass filtered at 10 kHz and sampled at 20 kHz. Acrodynamic signals were low-pass filtered at 5 kHz and sampled at 10 kHz. The signals and speakers were monitored throughout recording to ensure that the airflow mask was pressed tightly to the face, and that the pressure tube was not clogged with saliva. At the beginning or end of each recording session, pressure and flow calibration values were obtained using a water manometer for the pressure and a rotameter for the airflow.

The airflow recording methods used here differed from Müller and Brown (1980) in using a standard, unaltered pneumotachograph. Those authors, in contrast, adjusted their airflow mask to limit dead air space and avoid flows associated with lip and jaw movement. Specifically, they filled the area between the upper lip and nose with foam rubber and caulk, and affixed a sheet of dental dam material to the lower lip. These modifications were not implemented here partly because they were impractical for recording young children, and most studies on speech aerodynamics have used unaltered masks. Most important, perhaps, the signals obtained from the standard mask permitted reliable measurement of the two events necessary, namely oral closure and release. The relevance of this methodological difference for directly comparing this work to that of Müller and Brown is considered in Sec. III D.

C. Speaking task

Pictured stimuli were used to elicit multiple productions of the utterances "Poppa Popper" and "Poppa Bopper." (Poppa Popper had a bowl of popcorn, and a large P on his shirt; Poppa Bopper was "bopping" a large block, inscribed with a B, with a mallet). The stimuli were introduced to the speakers before recording took place. During recording, a picture was presented to the speaker, who then repeated the utterance five times. A research assistant presented the pictures and maintained the children's interest throughout the task. Each picture was presented five times during the recording session, in randomized order, resulting in approximately 25 productions per consonant.

D. Exclusion criteria

Tokens were discarded when (a) monitoring of the speaker or the airflow data suggested that an airflow leak may have occurred; (b) the target consonant was not perceptually accurate in voicing, place, and manner; (c) the flow signals for the target stop did not show abrupt changes associated with vocal-tract closure and release, suggesting lenition or spirantization, and/or if second derivative peaks defining closure or release could not be identified with certainty (described further in the next section); or (d) P_{io} did not return to near-baseline during the unstressed vowel preceding the target consonant, with "near-baseline" defined as 1 cm H_2O. Cases with nonbaseline pressures primarily occurred in a few 10 year olds who used faster speech rates. Most other exclusions in the children owed to airflow leaks (e.g., the child moved and mask seal was not maintained) or lenition. In total, 1017 pressure traces were analyzed (510 for /b/ and 507 for /p/); there were 282, 305, and 430 tokens for the 5 year olds, 10 year olds, and women, respectively. Subjects contributed an average of 21 tokens each of /p/ and /b/ to the data pool (range=10–31 tokens per consonant per speaker).

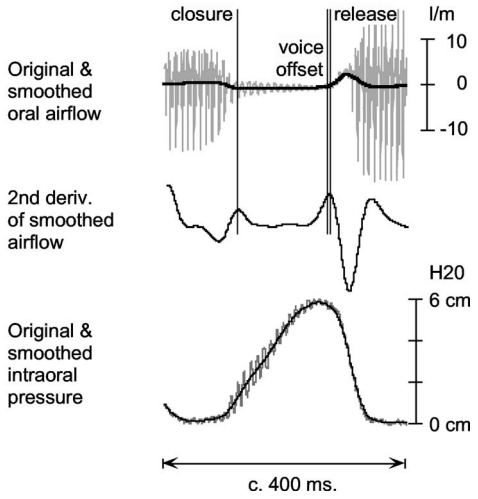


FIG. 1. Sample production of /əbɑ/ ("Poppa Bopper") showing events used for measurement. Top panel: Original (grey) and smoothed (black) oral airflow. The three vertical lines indicate the times of (a) closure, (b) voicing offset, and (c) release. In fully voiced tokens like the one shown here, the time of voicing offset was set just preceding the time of release. Second panel: Second time derivative of smoothed oral flow. Third panel: Original (grey) and smoothed (black) intraoral pressure (P_{io}). All signals are temporally aligned.

E. Measurement of voicing and closure duration

Initial data analyses used the CHART™ software accompanying the PowerLab system. A sample token from a 10 year old is given in Fig. 1. All measures were made on the tokens of /p, b/ initiating the third, stressed syllable of the utterance. The times of voicing offset during the consonantal closure were identified visually from the unsmoothed air pressure and airflow signals. To determine the times of articulatory closure and release, the airflow signals were smoothed twice with a 201-point window, the signal was differentiated twice, and the resultant was smoothed. Release and closure were defined by finding those peaks in the second derivative signal that corresponded to the corners of the flat region in the flow signal demarcating the consonantal closure. This method of determining stop closure and release was particularly useful when the flow signals showed effects of supraglottal movement (i.e., baseline shifts) during stop closures, but it differed from that of Müller and Brown (1980), who made closure and release judgments visually from the unsmoothed flow signals. Since smoothing the flow and its derivatives could introduce some temporal noise, the data from one woman (38 tokens) were remeasured using visual inspection of the unsmoothed flow signals, to verify that the two measurement procedures yielded similar results. The closure and release times obtained via the two methods

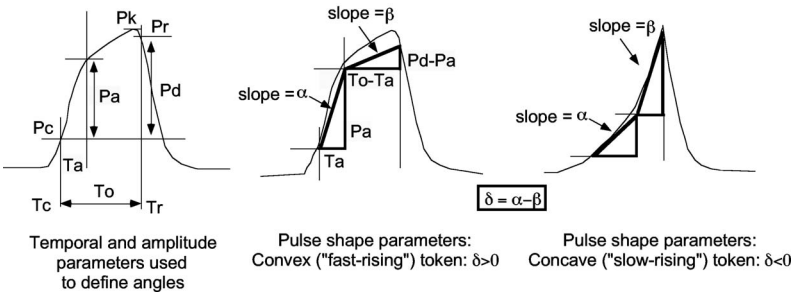


FIG. 2. Schematic signals showing time and amplitude measures made to quantify the shape of the pressure pulse, following Müller and Brown (1980).

differed by average of 2 ms (s.d.=5 ms). This aspect of our methods, therefore, appears not to be of concern for a comparison with Müller and Brown.

Times of oral closure, oral release, and voicing offset permitted calculation of (a) closure duration, (b) voicing duration, and (c) percentage of the closure that was voiced. Absolute duration of voicing was included following Westbury and Keating (1986). Percentage of voiced closure was included because children tend to have slower speech rates and longer segment durations than adults (e.g., Kent and Forner, 1980; Smith, 1978, 1992). Thus, even if children produce the same duration of voicing as adults, they may show proportionally less voicing during a stop, and be perceived to devoice more frequently. Finally, the number of devoiced stops productions of /b/ was assessed, following Smith's (1979) suggestion that these may be more prevalent in children than adults. Drawing on studies that have quantified closure voicing in both voiced and voiceless stops (Docherty, 1992; Flege and Brown, 1982; Westbury, 1979) a cutoff point of 30% voiced closure was selected for designating voiced stops as "devoiced" in production. (In the cited work, voicing was observed to persist into target voiceless stops up to 30% of the closure.) Using this cutoff, each subject's /b/ tokens were classified as (partially) voiced (30+% voiced) or devoiced (<30% voiced).

Since voicing offset was determined visually, a randomly chosen 20% of all tokens was remeasured by the first author to assess reliability. Although most of the original measures were made by graduate-student research assistants, a portion was made by the first author; of the remeasured data, 78% represented inter-rater reliability (original measures made by students) and 22% represented intra-rater reliability. Correlations between the original and remeasured data showed high reliability, both for voicing duration calculated in milliseconds and as a percentage of closure, and for inter- and intra-rater reliability (all r values 0.96 or higher and all p values <0.0001). For intra-rater reliability, the mean absolute measurement error was 5.2 ms for voicing duration (s.d.=6.9) and 0.06 (i.e., 6%) for the proportion of closure voiced (s.d.=0.09). For inter-rater reliability, mean voicing duration error was 3.0 ms (s.d.=6.6) and for voicing percentage it was 0.03 (s.d.=0.08).

F. Pulse shape calculations

The unsmoothed pressure signal for each token of /p/ and /b/ was extracted in the CHART program, using a time window wide enough to include the stop closure and release labels as defined earlier, and read into MATLAB for pressure pulse calculations. Following Müller and Brown (1980), the signals were low-pass filtered at 43 Hz using a sixth-order Butterworth filter, filtering both forwards and backwards to minimize temporal delay. From this smoothed signal, several parameters were calculated to permit characterization of the pressure pulse shape. These parameters are shown for a schematic example at the left of Fig. 2: Tc=time of closure (identified from the second derivative of the flow signal, as indicated earlier). Pc=pressure at closure. Tr=time of release (from second derivative of flow signal). Pr=pressure at release. To=closure duration (Tr−Tc). Ta=time to pressure average. Pa=pressure average relative to Pc. The average was obtained by integrating the smoothed pressure signal between Tc and Tr and Pc was then subtracted from the resultant. Pk=pressure peak. Pd=pressure difference between closure and release (Pr−Pc). α=slope of the hypotenuse of the triangle formed by Ta and Pa. β=slope of the hypotenuse of the triangle formed by (To−Ta) and (Pd−Pa). $\delta=\alpha-\beta$.

As shown in Fig. 2 (middle and right panels), a pressure pulse that rises quickly and then levels off should have a high α slope, low β slope, and $\delta>0$, whereas a slow initial rise followed by more rapidly increasing pressure should yield $\delta<0$.

In addition to the α, β, and δ measures, normalized slopes ($\alpha_n, \beta_n, \delta_n$) were calculated by multiplying α, β, and δ by To/Pd. Müller and Brown (1980) proposed this procedure to adjust for differences in closure and pressure pulse duration. Specifically, normalization corrects for the fact that lengthening a closure (e.g., as a result of a slower speech rate), given a constant rate of pressure increase, will lead to a higher pressure at release.

G. Statistical analysis

Repeated measures analyses of variance (ANOVAs) were conducted on voicing duration, percentage of closure voicing, δ, and δ_n measures, with consonant as the within-subject factor and age as the between-subject factor. A univariate ANOVA, with age as the independent variable, was used to analyze the percentage of devoiced /b/ for each speaker. The Bonferroni-corrected α level for determining significance in the five ANOVAs was 0.05/5=0.01. To quantify the relationship between voicing and P_{io} characteristics, within-subject correlations were run for δ and the duration of closure voicing. For 24 correlations, the adjusted α level was 0.05/24=0.002.

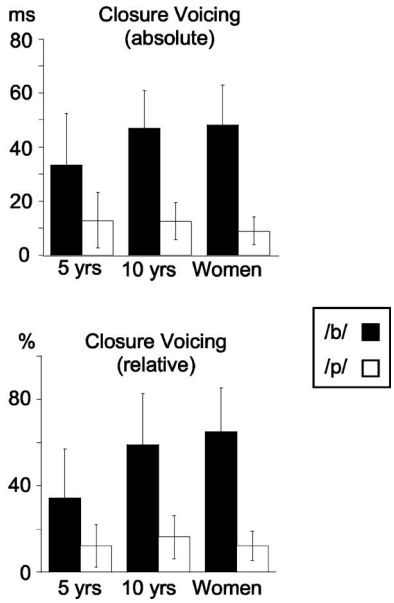


FIG. 3. Amount of closure voicing, in milliseconds (top), and as a percentage of the closure duration (bottom). Closed bars represent /b/; open represent /p/.

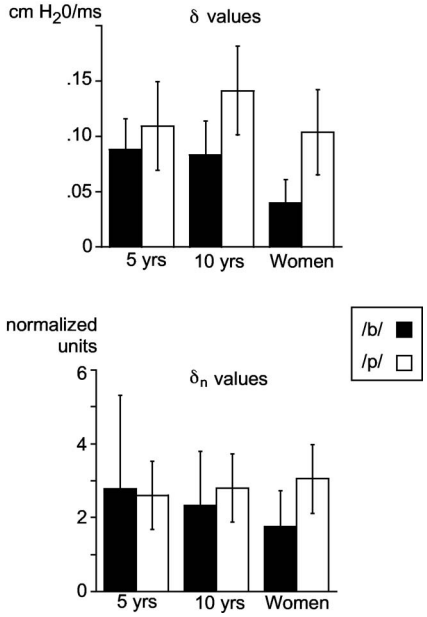


FIG. 4. δ (top) and normalized δ_n (bottom) values for each speaker group. Closed bars represent /b/; open represent /p/.

III. RESULTS

Sections III A and III B address group (i.e., age) effects in voicing and pressure measures. Section III C assesses the relationship between voicing and pressure data. Section III D compares the current data with those of Müller and Brown (1980), incorporating the post-hoc analysis of two men.

A. Stop closure and voicing durations

Closure voicing durations and percentages of closure voicing are shown for each speaker group in Fig. 3. ANOVA results for the two measures are given in Table I. As expected, voicing was more extensive for /b/ than /p/, measured both as a duration and as a percentage of the closure. The consonant effect was highly significant in both ANOVAs. A main effect of age was not significant for either analysis. The group-by-consonant interactions did not reach significance either, although the one for voicing percentage approached it ($p=0.018$). Qualitatively, the percentage of closure voicing for /p/ was similar across ages, whereas for /b/ it increased somewhat over age (see bottom panel of Fig. 3). The actual percentages for /p/ voicing, in ascending age order, were 12%, 16%, and 12%; for /b/, they were 36%, 59%, and 64%.

TABLE I. ANOVA results for closure voicing measured as a duration and as a percentage of the closure duration.

Voicing duration			
Effect	dfs	F	p
Consonant	1, 21	115.432	<0.0001
Group	2, 21	0.678	0.518
Consonant × group	2, 21	2.955	0.074
Voicing percentage			
Effect	dfs	F	p
Consonant	1, 21	102.595	<0.0001
Group	2, 21	2.314	0.124
Consonant × group	2, 21	4.927	0.018

The proportions of devoiced stops for all speakers are given in Table II. Although devoicing was most common on average in the 5 year olds, the incidence of devoiced /b/ varied considerably within groups, and the group effect did not reach significance [$F(2,21)=3.968$, $p=.035$].

B. Pressure contours

The δ analysis was intended to quantify pulse shape, with values of $\delta>0$ indicating convexity. The δ_n values are the normalized, unitless counterparts designed to correct for speech rate differences, all else being equal. Group results for δ and δ_n are shown in Fig. 4. The ANOVA results are summarized in Table III.

For δ, the consonant effect was highly significant, but the group effect failed to reach significance ($p=0.024$), as did the group-by-consonant interaction. The results for δ_n were rather different: Neither of the main effects nor the group-by-consonant interaction approached significance. As shown in Fig. 4 (lower panel), the 10 year olds and women still had the expected pattern of higher values in /p/ compared to /b/; the 5 year olds did not, and had a markedly higher standard deviation (s.d.) for /b/.

To determine why δ_n means and s.d.s differed so much from those of δ in the 5 year olds, the two parameters used to obtain δ_n from δ, namely closure duration (To) and the pressure difference between release and closure (Pd), were inspected. The results for pressure at closure (Pc) and release (Pr) were also reviewed to clarify the Pd data. The group means and s.d.s for these four measures are given in Table IV.

Table IV shows that the 5 year old's high δ_n s.d.s for /b/ reflect variability in both Pd and To, but compared to women the s.d.s for Pd were proportionally higher than those for To

TABLE II. Proportions of /b/ productions in which 30% or more of the closure was voiced. In the child groups, the first four speakers are females; the last four are males.

	Speakers 1–8 in each group								Group average
5 year olds	0.12	0.35	0.38	0.06	0.79	0.89	1.00	0.11	0.46
10 year olds	0.55	0.75	1.00	1.00	0.88	0.23	0.63	1.00	0.75
Women	1.00	0.82	0.80	0.87	0.48	0.93	0.97	0.92	0.85

(2.1 vs 1.7). Pd variability, in turn, reflects variability in both Pc and Pr: s.d.s for both are higher in the 5 year olds than the women. As for the mean values, both Pd and To are higher in the 5 year olds than the adults, but proportionally more so for Pd (1.8 times the adult value) than for To (1.4–1.5 times the adult value). As laid out in Sec. I, age-related differences in durations and intraoral pressures were expected for a variety of reasons. The δ normalization method proposed by Müller and Brown (1980) is designed to correct for higher pressures that arise simply as a function of longer durations. The current data show both longer durations and higher pressures in 5 year olds than adults, but the pressure differences are proportionally greater. Because these two parameters do not change in parallel with age, it appears that this method of δ normalization is not appropriate for comparing across age groups.

The data presented above indicate that δ differentiated /p/ and /b/ on average for all age groups. To determine how well this held for individuals, the δ difference between /p/ and /b/ was obtained for each speaker. If intraoral pressure rises faster in /p/ than /b/, the difference should be positive. The results are given in Fig. 5. Interspeaker variability is considerable for all groups, but most values are positive; negative values occur only among the children.[2] Thus, the /p/-/b/ difference observed in the δ averages is reflective of most speakers.

C. Relations between stop consonant voicing and pressure measures

To quantify the relationship between pressure pulse trajectories and the degree of stop consonant voicing, within-subject correlations were run on δ and closure voicing duration for /p/ and /b/ combined.[3] The results are given in Fig. 6. As expected, r values are usually negative: As δ becomes more positive (as for a convex pulse), the amount of voicing decreases. The three child speakers with positive r values are the same as those with negative δ differences between /p/ and /b/. Significant correlations (using a criterion of $0.05/24=0.002$) are indicated in Fig. 6 by asterisks. The number of significant correlations increases with age; that is, the relationship between pressure pulse parameters and voicing is more consistent in adults.

TABLE III. ANOVA results for the pulse shape measure δ and its normalized counterpart δ_n.

δ			
Effect	dfs	F	p
Consonant	1, 21	35.006	<0.0001
Group	2, 21	4.497	0.024
Consonant × group	2, 21	2.742	0.087
δ_n			
Effect	dfs	F	p
Consonant	1, 21	1.853	0.188
Group	2, 21	0.153	0.859
Consonant × group	2, 21	1.199	0.321

D. Comparison with Müller and Brown

This section assesses how the current results compare with those of Müller and Brown (1980). Those authors did not explicitly measure closure voicing, so this discussion is limited to the δ measure; δ_n is not considered because of the uneven nature of age-related changes in To and Pd noted in Sec. III B.

As indicated earlier, the current methods differed in a few ways from those of Müller and Brown (1980): Here, closure and release were measured semiautomatically from the acceleration of the flow signal (addressed in Sec. II E); a standard pneumotachograph was used, without employing the adjustments used by Müller and Brown; and the adult group consisted of women rather than men. The mask difference relates only to the airflow-based measures of closure and release, and should not affect the pressure measures themselves. The gender difference, however, could affect pressure measures given that women, on average, have smaller vocal tract sizes than men. Thus, to allow a more valid comparison with Müller and Brown, equivalent data were recorded from two men and analyzed in the same manner as the women and children to obtain δ measures. Values of δ from Müller and Brown were estimated (to the nearest 0.005 cm H_2O/ms) from their plots for bilabial stops in the /ɑ/ vowel context. (That study also included alveolar consonants and a high vowel context). Since Müller and Brown's subjects all produced the same number of tokens ($N=6$), group averages could be estimated from the individual subject averages.

As shown in Table V, δ values obtained here for both /p/ and /b/ were higher than Müller and Brown's (1980), particularly for the children. A general pattern of descending δ values for /b/ across 5 year olds, 10 year olds, women, and men in Table V provides qualitative support for the hypothesis that intraoral pressure tends to rise more slowly in those with larger vocal tracts. (This is consistent with the data in Table IV: The pressure rise during closure, Pd, was higher in children than women, and to a greater degree than closure duration, To). The average δ differences between Müller and Brown's data and the current *adult* speakers appear to reflect sampling error, however. Müller and Brown's data showed considerable intersubject variability in δ, and standard devia-

TABLE IV. Descriptive data for closure duration (To), pressure difference in the closure (Pd), pressure at closure (Pc), and pressure at release (Pr). To and Pd are the two measures used to δ_n from δ; Pd=Pr−Pc. Each cell in the first three rows shows the group mean (standard deviation). The last row provides the ratio of the 5 year olds to the adults (means and standard deviations calculated separately).

	Closure duration (To)		Pressure difference (Pd)		Pressure at closure (Pc)		Pressure at release (Pr)	
	/b/	/p/	/b/	/p/	/b/	/p/	/b/	/p/
5 year olds	106.7 (17.9)	113.7 (21.4)	4.9 (1.9)	5.2 (1.8)	2.5 (0.8)	3.4 (1.1)	7.4 (2.0)	8.6 (1.9)
10 year olds	86.7 (17.9)	82.9 (18.0)	3.9 (0.9)	4.4 (0.7)	2.4 (0.8)	3.9 (1.5)	6.3 (1.2)	8.3 (1.8)
Woman	75.5 (10.4)	75.6 (9.6)	2.7 (0.9)	2.9 (1.2)	1.7 (0.5)	2.9 (0.9)	4.3 (1.1)	5.8 (1.5)
Ratio: 5 year olds to women	1.4 (1.7)	1.5 (2.2)	1.8 (2.1)	1.8 (1.5)	1.5 (1.6)	1.2 (1.2)	1.7 (1.8)	1.5 (1.3)

tions for δ were also fairly large here (refer back to Fig. 4). Review of the women's data showed four cases in which individuals had δ values at or lower than Müller and Brown's averages. (The same was true for one 10 year old.) Thus, the current adult data overlap with those of Müller and Brown, and appear to be generally consistent with the results of the earlier study. Section IV raises some issues concerning δ which may explain the wide variation in δ even in adult speakers.

IV. DISCUSSION

A. Age and stop consonant voicing

Past studies have found that children up to 6 years of age produce fewer voiced stops and less closure voicing than adults (e.g., Barton and Macken, 1980; Kewley-Port and Preston 1974; Smith, 1979; Zlatin and Koenigsknecht, 1976). Qualitatively, the current data showed the least /b/ voicing and the largest number of devoiced /b/ productions in the 5 year olds, but age effects did not reach significance for any of the voicing measures. Four factors should be considered in interpreting this result: (a) nature of the voicing contrast in English; (b) phonetic context; (c) age of the children; and (d) sample size.

Given that closure voicing is not obligatory in stressed, syllable-initial contexts in American English (e.g., Lisker and Abramson, 1964; Zlatin, 1974), all speakers have the option of producing voiceless unaspirated stops, which may be motorically less demanding (Kewley-Port and Preston, 1974). At the same time, the intervocalic frame in which the target consonants appeared here tends to facilitate closure voicing. Children might be particularly susceptible to factors favoring *devoicing* (Westbury and Keating, 1986), but to ascribe a lack of significant age effects here to the intervocalic context would be to assert that children have a stronger positional bias toward *voicing*. Given smaller vocal tract sizes and higher phonation threshold pressures in children, there is no reason to think they should be more likely to voice in an intervocalic position than adults. That is, the phonetic context may have predisposed all speakers to produce voiced stop closures, but this should not have made age differences less likely.

More likely is that the children's ages and the group sizes had the effect of minimizing age effects. Past developmental studies of consonant voicing have mostly concentrated on very young children. The requirements of the experimental task used here precluded recording children younger than about 4 years, and the younger group included individuals nearing their sixth birthday. These children were thus at the upper end of the age range at which group differences have been demonstrated. The intensive within-speaker processing here also limited group sizes. Finally, the children showed considerable intersubject variability, as has often been observed (e.g., Eguchi and Hirsh, 1969; Kent and Forner, 1980). With a larger sample size, the mean group differences may have outweighed the within-group variability to produce a significant age effect (e.g., Kirk, 1999).

In contrast to the 5 year olds, the 10 year olds were qualitatively as well as statistically comparable to the women in their voicing measures. Although a few studies have reported VOTs for voiced stops in children of this age (Kent

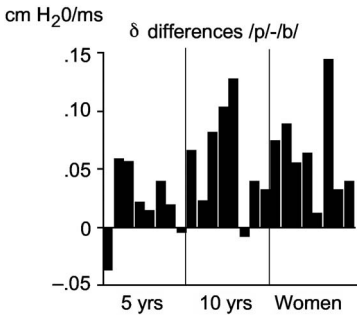


FIG. 5. Differences between δ values of /p/ and /b/ for each speaker.

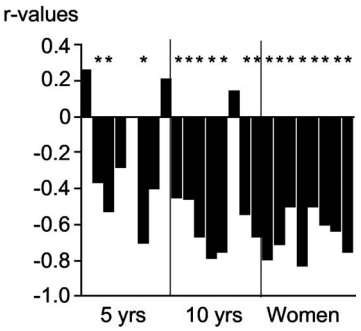


FIG. 6. Results of correlations (Pearson's r values) between δ and duration of closure voicing for each speaker. Asterisks mark speakers for whom the correlation was significant at $p<0.002$.

TABLE V. Comparison of current δ measures with those from Müller and Brown (1980, p. 342). The data from the earlier study are estimated from individual subject plots.

		/b/	/p/
Current study	5 year olds	0.088	0.109
	10 year olds	0.083	0.141
	Women	0.040	0.104
	Men	0.033	0.119
Müller and Brown		0.010	0.067

and Forner 1980; Ohde, 1985; Whiteside *et al.*, 2004), their data do not clearly indicate the degree of closure voicing, since a positive VOT may indicate a truly voiceless stop as well as one in which voicing dies out before the end of the closure. Based on the current results, it appears that 10 year olds demonstrate essentially mature closure voicing in /b/, despite continuing differences from adults in vocal tract sizes, aerodynamic quantities, speech segment durations, and token-to-token variability (e.g., Bernthal and Beukelman, 1978; Eguchi and Hirsh, 1969; Goldstein, 1980; Kent and Forner, 1980; Netsell *et al.*, 1994; Stathopoulos and Weismer, 1985; Walsh and Smith, 2002). These older children may thus have mastered methods of vocal-tract pressure management so as achieve adult-like closure voicing. The same explanation may hold for those 5 year olds whose voicing behavior was more adult-like. One difficulty with this interpretation, particularly for the younger children, is that the correlations between voicing and introral pressure (cf. Sec. IV C) were less often significant in children than adults, and sometimes reversed. Modeling of closure voicing with laryngeal and vocal tract models scaled to child sizes may provide greater insight into the conditions needed to produce voiced stops in child speakers.

B. Quantitative pressure pulse analysis and age

Past work has established that children as young as 4 years of age, like adults, produce higher peak P_{io} in voiceless stops than in voiced (Arkebauer *et al.*, 1967; Bernthal and Beukelman, 1978; Brown, 1979; Lisker, 1970; Malécot, 1966; Miller and Daniloff, 1977; Stathopoulos and Weismer, 1985; Subtelny *et al.*, 1966; Warren and Hall, 1973), thereby demonstrating some aerodynamic differentiation of consonants according to voicing category. At the same time, previous authors (particularly Kewley-Port and Preston, 1974) have attributed limited closure voicing in young children to immature control over intraoral pressure, and Müller and Brown (1980) specifically suggested that P_{io} might rise faster in children than adults. These results led to the prediction that δ and its normalized counterpart δ_n would differ between children and adults.

The data for δ showed a significant consonant effect, but age effects failed to meet significance. Comparison of the women and children's data with two men recorded post-hoc and those recorded by Müller and Brown (1980) showed decreasing δ values for speakers with larger vocal tracts, providing some support for the possibility of a size-related trend, and it may be again that a larger sample size and/or a younger group of children would yield significant age effects. Alternatively, children may perform more extreme volumetric adjustments than adults in order to control P_{io} increases in their smaller vocal tracts (but see again the comments on this point in Sec. IV C).

The results for δ_n showed neither age nor consonant effects. Inspection of the measures used to derive δ_n from δ, namely closure duration (To) and pressure difference in the stop (Pd), showed proportionally greater age differences for Pd. This suggests that δ is a more valid measure for comparing across ages than δ_n. The normalized measure may, however, still be appropriate to adjust for speech rate or loudness differences within individual speakers or age groups.

C. Relationships between P_{io} and voicing

The δ differences between /p/ and /b/ were in the expected direction for most speakers, but reversed for three of the children. Further, correlations between δ and voicing duration were significant for all of the women and for 7/8 of the 10 year olds, but only 3/8 of the 5 year olds. Thus, whereas adults have established consistent relationships between stop consonant voicing and P_{io} characteristics, as measured by δ, many 5 year olds and a few 10 year olds have not. Correlations in the expected direction that were statistically insignificant suggest that some children may have implemented control over P_{io} buildup in target voiced stops which was not extensive enough to maintain voicing, or else that they exercised such control inconsistently. Correlations in the reverse direction suggest that some children have in fact not yet learned how to manipulate intraoral pressure in order to maintain voicing. Both of these cases speak against a general developmental strategy in which children actively increase supraglottal volumes more than adults in an attempt to sustain voicing: In such a situation, one might expect even stronger correlations in children than adults. It is also possible that the δ measure is not sufficient to capture all aspects of aerodynamic control relevant for stop consonant voicing. As discussed in the following, some unexpected values of δ were encountered in the course of this work, lending support to this last interpretation.

D. Comparison with Müller and Brown and critique

The δ measures obtained here were somewhat higher, on average, than those obtained by Müller and Brown (1980). The particularly high δ values for /b/ in children compared to women as well as men may represent a tendency for speakers with smaller vocal tracts to have limited possibilities for supralaryngeal volume adjustments, but the differences among the adults appear to reflect normal cross-speaker variation. Different speech materials may also contribute to δ variance between the two corpora. Whereas the current study placed the target consonants in an intervocalic, running speech context, Müller and Brown elicited their consonants in isolated, symmetrical, nonsense VCV utterances, with instructions to produce equal stress on the two syllables. Conceivably, differences in stress pattern or speaking style may influence δ measures.

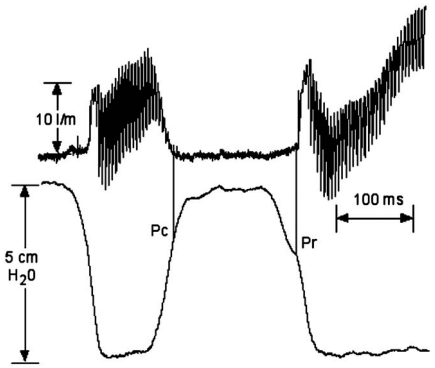


FIG. 7. A token of /b/ produced by a 5-year-old girl which yielded a negative value of Pd (difference between Pc, pressure at closure, and Pr, pressure at release).

Müller and Brown (1980) were unique in proposing methods of measuring time-varying aspects of P_{io} during stop closures. The results of their own and subsequent aerodynamic modeling (e.g., Westbury, 1983; Westbury and Keating, 1986) suggested further that the time course of P_{io} increase during stop consonants could provide information on aerodynamic management of voiced stops. For these reasons, the current work employed the δ measure, proposed by Müller and Brown as a simple metric for capturing aspects of P_{io} pulse shape. Yet the data obtained here, as well as close inspection of Müller and Brown's own data, suggest that this measurement method may have limitations. Specifically, some of the measures used to calculate δ and δ_n yielded unexpected values. These primarily reflected two situations: Cases where Pd (pressure difference between release and closure) and β (angle from average to release pressure) were negative. Pressure pulse depictions like those in Fig. 2 lead one to expect both to be positive. The next paragraphs briefly describe how these values came about.

Negative values of Pd occurred in five tokens (three from children, two from men; four of the five were in /b/). An example from a 5 year old is given in Fig. 7. (This token also had a negative β.) In this production, intraoral pressure decreases right before release. Although these examples are infrequent in our database, they appear to indicate that supraglottal movements affecting pressure contours may be initiated rather late during the closure.

Whereas negative Pd was rare, negative β was more common, occurring in 155 tokens (15% of the data for women and children combined). These were found across age groups (36%, 39%, and 26% came from women, 10 year olds, and 5 year olds respectively), and were about equally distributed across consonants (53% for /p/; 47% for /b/). Tokens with negative β also accounted for 12% of the men's data. Negative β occurs when the rise from closure to average pressure (Pa) is greater than the pressure difference between closure and release (Pd). Such values may also have occurred in Müller and Brown's (1980) data: Their plots show some speaker averages close to zero, with standard deviations consistent with negative values. The occurrence of negative β changes the interpretation of δ: Higher values of δ may result not only from convex wave forms, but also from pressure contours that fall between the time of average pressure and the time of release. To determine whether values of negative β affected the conclusions, all statistics based on δ or δ_n were rerun without tokens whose β values were negative. The results were similar or identical to the original ones, suggesting that the occurrence of negative β does not affect the general pattern of results in this work. Further, case-by-case review of these $-\beta$ tokens did not show them to be atypical in ways other than their β values; that is, they appeared to represent legitimate data points. The main observation to be made here is that Müller and Brown's measurement methods may yield some paradoxical values, and δ may reflect not only the convex/concave nature of a pressure pulse. It appears that more work is needed on how best to assess the time-varying characteristics of intraoral pressure data.

V. CONCLUSIONS

An extensive literature, using VOT as well as other measures, has indicated that stop consonant voicing is more limited in preschool children than in adults. This has been attributed to poor aerodynamic control in the context of anatomical growth and physiological maturation, processes which persist into adolescence. The current study used Müller and Brown's (1980) δ measure to assess the relationship between P_{io} characteristics and stop consonant voicing in children and adults. Unexpectedly, measures of voicing and δ did not differ significantly across age groups. For the 5 year olds, this may reflect limited sample sizes, but it appears that by 10 years of age, children have learned to produce virtually adult-like voicing patterns despite immature physical systems. The correlational patterns between δ and stop consonant voicing do suggest, however, that aerodynamic control of voicing in children is still subject to more variability than in adult speakers, and that some children, particularly at 5 years but maybe also at 10 years, have not yet learned to manipulate intraoral pressure as effectively as adults.

When interspeaker variability is taken into account, δ values for the current adult speakers seem consistent with those of Müller and Brown (1980), although varying stimulus materials may have yielded some differences between the two studies. More interesting is that issues arose relative to the δ measure itself. Paradoxical values of δ appear to have occurred in Müller and Brown's data as well as our own. Although these difficulties do not negate the conclusions of Müller and Brown regarding mechanisms of pressure control in adults, they do suggest that new methods of assessing time-varying characteristics of intraoral pressure data should be devised in order to characterize the data more fully and appropriately. Future work should explore this topic further.

ACKNOWLEDGMENTS

Several students from Long Island University, Brooklyn Campus and New York University contributed to recording and data analysis for this project. In alphabetical order: Giridhar Athmakuri, Jesse Farver, Linda Greenwald, Ingrid Katz, Karen Keung, Elizabeth Perlman, Simcha Pruss, and Gabrielle Rothman (see Rothman, *et al.*, 2002). Thanks to Elaine

Russo Hitchcock for assistance in subject recruitment and scheduling; to Linda Hunsaker for creating the stimulus pictures; to W. Einar Mencl and Scott Youmans for statistical advice; and to Anders Löfqvist, Katherine S. Harris, Arthur Abramson, Brad Story, and four anonymous reviewers for comments on earlier versions of this manuscript. This work was supported by NIH Grant No. DC-04473-03 to Haskins Laboratories and by Conselho Nacional de Desenvolvimento Cientifico e Tecnologico-CNPq (Brazil).

[1]In addition to convex, concave, and linear, Müller and Brown also categorized a few tokens as bimodal (double-peaked) or delayed (increasing slope, then a flat region, then increasing slope again). These five categories were further divided into tokens that had an abrupt slope change in the rising phase of the curve (breaking tokens) versus those that did not (smooth tokens). Although the original analysis of our data included this categorization, rater reliability was unacceptably low for the full ten-category analysis, so those results are not included here. Because the current paper only presents data for Müller and Brown's quantitative analysis, some details of the qualitative measurement scheme have been supressed in the main text.

[2]The speech of the children with negative values was not obviously remarkable in other ways; for example, it was not the case that more tokens were discarded from these children because of production errors.

[3]The two consonants were combined here in order to provide the most general picture of how voicing and pressure are related. The primary articulatory difference between /p/ and /b/, viz. vocal fold abduction, should lead to a faster pressure increase in /p/ (i.e., higher δ) as well as a reduction in voicing. Voicing in milliseconds is used here following the one study that explicitly compared closure voicing and P_{io} trajectories (Westbury, 1983). Correlations between δ and voicing as a percentage of the closure were qualitatively similar, but most r values were somewhat lower and fewer values reached significance. Finally, correlations for /b/ alone also yielded a majority of negative values (21 of 24 speakers), but only 2 of 24 p values reached significance. Some reduction in statistical significance is to be expected in both cases simply because only about half as many data points were included in each analysis.

Allen, G. D. (**1985**). "How the young French child avoids the pre-voicing problem for word-initial stops," J. Child Lang **12**, 37–46.

Arkebauer, H. J., Hixon, T. J., and Hardy, J. C. (**1967**). "Peak intraoral air pressures during speech," J. Speech Hear. Res. **10**, 196–208.

Atkinson, J. E. (**1978**). "Correlation analysis of the physiological features controlling fundamental voice frequency," J. Acoust. Soc. Am. **63**, 211–222.

Barton, D., and Macken, M. A. (**1980**). "An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children," Lang Speech **23**, 159–169.

Bell-Berti, F. (**1975**). "Control of pharyngeal cavity size for English voiced and voiceless stops," J. Acoust. Soc. Am. **57**, 456–461.

Bernthal, J. E., and Beukelman, D. R. (**1978**). "Intraoral air pressure during the production of /p/ and /b/ by children, youths, and adults," J. Speech Hear. Res. **21**, 361–371.

Bosma, J. F. (**1975**). "Anatomic and physiologic development of the speech apparatus," in *The Nervous System: Human Communication and its Disorders*, edited by D. B. Tower (Raven, New York), Vol. **3**, pp. 469–481.

Brown, W. S., Jr. (**1979**). "Supraglottal air pressure variations associated with consonant productions by children," in *Current Issues in the Phonetic Sciences: Proceedings of the IPS-77 Congress*, edited by H. Hollien and P. Hollien (Benjamins, Amsterdam), Vol. **2**, pp. 935–944.

Crelin, E. S. (**1987**). *The Human Vocal Tract: Anatomy, Function, Development, and Evolution* (Vantage, New York).

de Troyer, A., Yernault, J.-C., Englert, M., Baran, D., and Paiva, M. (**1978**). "Evolution of intrathoracic airway mechanics during lung growth," J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **44**, 521–527.

Docherty, G. (**1992**). *The Timing of Voicing in British English Obstruents* (Foris, Berlin).

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Oto-Laryngol. **57**, 1–51.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Flege, J. E. (**1982**). "Laryngeal timing and phonation onset in utterance-initial English stops," J. Phonetics **10**, 177–192.

Flege, J. E., and Brown, W. S., Jr. (**1982**). "The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance," J. Phonetics **10**, 335–345.

Gelfer, C. E., Harris, K. S., Collier, R., and Baer, T. (**1983**). "Is declination actively controlled?" in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. Titze and R. Scherer (Denver Center for the Performing Arts, Denver), pp. 113–126.

Goldstein, U. (**1980**). "An articulatory model for the vocal tracts of growing children," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Hirano, M., Kurita, S., and Nakashima, T. (**1983**). "Growth, development, and aging of human vocal folds," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 22–43.

Hoit, J. D., Hixon, T. J., Watson, P. J., and Morgan, W. J. (**1990**). "Speech breathing in children and adolescents," J. Speech Hear. Res. **33**, 33–69.

Ishizaka, K., and Matsudaira, M. (**1972**). "Theory of vocal cord vibrations," Bull. Univ. Electro-Comm. **23**, 107–136.

Kahane, J. C. (**1982**). "Growth of the human prepubertal and pubertal larynx," J. Speech Hear. Res. **25**, 446–455.

Kent, R. D., and Forner, L. L. (**1980**). "Speech segment durations in sentence recitations by children and adults," J. Phonetics **8**, 157–168.

Kent, R. D., and Moll, K. L. (**1969**). "Vocal-tract characteristics of the stop cognates," J. Acoust. Soc. Am. **46**, 1549–1555.

Kewley-Port, D., and Preston, M. (**1974**). "Early apical stop production: A voice onset time analysis," J. Phonetics **2**, 195–210.

Kirk, R. E. (**1999**). *Statistics: An Introduction* (Harcourt Brace College Publishers, Forth Worth).

Lieberman, P. (**1967**). *Intonation, Perception, and Language* (MIT, Cambridge, MA).

Lindqvist, J. (**1972**). "Laryngeal articulation studied on Swedish subjects," Quart. Stat. Prog. Rept. (Sp. Transm. Lab., Royal Inst. Tech., Stockholm) **2–3**, 10–27.

Lisker, L. (**1970**). "Supraglottal air pressure in the production of English stops," Lang Speech **13**, 215–230.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Löfqvist, A. (**1975**). "A study of subglottal pressure during the production of Swedish stops," J. Phonetics **3**, 175–189.

Lucero, J. C., and Koenig, L. L. (**2005**). "Phonation threshold pressures as a function of laryngeal size in a two-mass model of the vocal folds," J. Acoust. Soc. Am. **118**, 2798–2801.

Malécot, A. (**1966**). "The effectiveness of intra-oral air-pressure-pulse parameters in distinguishing between stop cognates," Phonetica **14**, 65–81.

Mansell, A. L., Bryan, A. C., and Levison, H. (**1977**). "Relationship of lung recoil to lung volume and maximum expiratory flow in normal children," J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **42**, 817–832.

McGlone, R. E., and Shipp, T. (**1972**). "Comparison of subglottal air pressures associated with /p/ and /b/," J. Acoust. Soc. Am. **51**, 664–665.

Miller, C. J., and Daniloff, R. (**1977**). "Aerodynamics of stops in continuous speech," J. Phonetics **5**, 351–360.

Müller, E. M., and Brown, W. S. (**1980**). "Variations in the supraglottal air pressure waveform and their articulatory interpretation," in *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic, Madison, WI), Vol. **4**, pp. 318–389.

Netsell, R. (**1969**). "Subglottal and intraoral air pressures during the intervocalic contrast of /t/ and /d/," Phonetica **20**, 68–73.

Netsell, R., Lotz, W. K., Peters, J. E., and Schulte, L. (**1994**). "Developmental patterns of laryngeal and respiratory function for speech production," J. Voice **8**, 123–131.

Ohde, R. N. (**1985**). "Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children," J. Acoust. Soc. Am. **78**, 1554–1561.

Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT, Cambridge, MA).

Polgar, G., and Weng, T. R. (**1979**). "The functional development of the respiratory system from the period of gestation to adulthood," Am. Rev. Respir. Dis. **120**, 625–695.

Preston, M. S., and Port, D. K. (**1969**). "Further results on the development of voicing in stop-consonants in young children," Haskins Labs. Status Rpts. **19/20**, 189–199.

Preston, M. S., Yeni-Komshian, G., Stark, R. E., and Port, D. K. (**1968**). "Developmental studies of voicing in stops," Haskins Labs. Status Rpts. **13/14**, 181–184.
Riordan, C. J. (**1980**). "Larynx height during English stop consonants," J. Phonetics **8**, 353–360.
Rothenberg, M. R. (**1968**). *The Breath-Stream Dynamics of Simple-Released-Plosive Production* (S. Karger, Basel), Vol. **6**.
Rothman, G. B., Koenig, L. L., and Lucero, J. C. (**2002**). "Intraoral pressure trajectories during voiced and voiceless stops in women and children," J. Acoust. Soc. Am. **108**, 2416(A).
Sasaki, C. T., Levine, P. A., Laitman, J. T., and Crelin, E. S. (**1977**). "Postnatal descent of the epiglottis in man," Arch. Otolaryngol. **103**, 169–171.
Smith, B. L. (**1978**). "Temporal aspects of English speech production: A developmental perspective," J. Phonetics **6**, 37–67.
Smith, B. L. (**1979**). "A phonetic analysis of consonantal devoicing in children's speech," J. Child Lang **6**, 19–28.
Smith, B. L. (**1992**). "Relationships between duration and temporal variability in children's speech," J. Acoust. Soc. Am. **91**, 2165–2174.
Stathopoulos, E. T., and Sapienza, C. (**1993**). "Respiratory and laryngeal measures of children during vocal intensity variation," J. Acoust. Soc. Am. **94**, 2531–2543.
Stathopoulos, E. T., and Weismer, G. (**1985**). "Oral airflow and air pressure during speech production: A comparative study of children, youths and adults," Folia Phoniatr. **37**, 152–159.
Stevens, K. N. (**1991**). "Vocal-fold vibration for obstruent consonants," in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego), pp. 29–36.
Subtelny, J. D., Worth, J. H., and Sakuda, M. (**1966**). "Intraoral pressure and rate of flow during speech," J. Speech Hear. Res. **9**, 498–518.
Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., and Wilhelms-Tricarico, R. (**1997**). "Tongue surface displacement during bilabial stops," J. Acoust. Soc. Am. **102**, 562–571.
Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.
Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (**2005**). "Development of vocal tract length during early childhood: A magnetic resonance imaging study," J. Acoust. Soc. Am. **117**, 338–350.
Walsh, B., and Smith, A. (**2002**). "Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes," J. Speech Lang. Hear. Res. **45**, 1119–1133.
Warren, D. W., and Hall, D. J. (**1973**). "Glottal activity and intraoral pressure during stop consonant productions," Folia Phoniatr. **25**, 121–129.
Westbury, J. R. (**1979**). "Aspects of the temporal control of voicing in consonant clusters in English," Doctoral dissertation, University of Texas at Austin, published in Texas Linguistic Forum **14**, 1–304.
Westbury, J. R. (**1983**). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing," J. Acoust. Soc. Am. **73**, 1322–1336.
Westbury, J. R., and Keating, P. A. (**1986**). "On the naturalness of stop consonant voicing," J. Ling. **22**, 145–166.
Whiteside, S. P., Henry, L., and Dobbin, R. (**2004**). "Sex differences in voice onset time: A developmental study of phonetic context effects in British English," J. Acoust. Soc. Am. **116**, 1179–1183.
Yeni-Komshian, G. H., Caramazza, A., and Preston, M. S. (**1977**). "A study of voicing in Lebanese Arabic," J. Phonetics **5**, 35–48.
Zlatin, M. A. (**1974**). "Voicing contrast: Perceptual and productive voice onset time characteristics of adults," J. Acoust. Soc. Am. **56**, 981–995.
Zlatin, M. A., and Koenigsknecht, R. A. (**1976**). "Development of the voicing contrast: A comparison of voice onset time in stop perception and production," J. Speech Hear. Res. **19**, 93–111.

Determination of superior surface strains and stresses, and vocal fold contact pressure in a synthetic larynx model using digital image correlation

Mychal Spencer, Thomas Siegmund,[a] and Luc Mongeau[b]
School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907

(Received 31 October 2006; revised 5 November 2007; accepted 6 November 2007)

Stresses and strains within the vocal fold tissue may play a critical role in voice fatigue, in tissue damage and resulting voice disorders, and in tissue healing. In this study, experiments were performed to determine mechanical fields on the superior surface of a self-oscillating physical model of the human vocal folds using a three-dimensional digital image correlation method. Digital images obtained using a high-speed camera together with a mirror system were used to measure displacement fields, from which strains, strain rates, and stresses on the superior surface of the model vocal folds were computed. The dependence of these variables on flow rate was established. A Hertzian impact model was used to estimate the contact pressure on the medial surface from superior surface strains. A tensile stress dominated state was observed on the superior surface, including during collision between the model folds. Collision between the model vocal folds limits the medial-lateral stress levels on the superior surface, in conjunction with compressive stress or contact pressure on the medial surface. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2821412]

PACS number(s): 43.70.Bk, 43.70.Aj, 43.70.Gr [BHS] Pages: 1089–1103

I. INTRODUCTION

Voice is produced by the dynamic interactions between lung-driven airflow, sound waves in the airway, and the deformations of the vocal folds tissue (VFs). A good understanding of the basic mechanics of phonation is needed for the development of better clinical diagnosis tools and voice prosthetic devices. Accurate measures of the mechanical stress in VFs are essential for the understanding of vocal fold damage and healing (Titze, 1994).

The present study is concerned with the deformation and stresses in a self-oscillating physical model of vocal folds. In general, the mechanical stresses within the VFs depend on: (1) the fluid pressure within the glottis; (2) the occurrence of impact between the vocal folds, and the associated change in linear momentum of the bulk of the vocal folds; and (3) elastic stresses related to the deformation of the vocal fold tissue due to posturing. These factors have been investigated in a number of previous studies. Studies on airflow through the glottis have been the focus of many investigations related to human voice production, mainly because airflow is a primary factor in the generation of sound (Titze, 2000). Flow studies are relevant for the present investigation since they provide first-order insight into the pressure applied onto the vocal fold. For modeling purposes, the glottis may be construed as an orifice with a time-varying geometry. The flow through such orifices is complex, involving jet formation, turbulence, dynamic flow separation, and asymmetry. Time varying glottal jet plume angles may be due to asymmetric orifice contours or the so-called Coanda effect. The sound pressure produced during voicing is a function of the time varying changes in the major loss coefficient of the glottal orifice, as described by Park and Mongeau (2007). Previous research using physical laryngeal models, rigid models, and computational models has yielded a good understanding of the fluid pressure distribution acting over the surface of the vocal folds (Scherer *et al.*, 2001). Typically, the fluid pressure on the inferior surface is near the subglottal pressures, and oscillates around the mean transglottal pressure (typically 0.8 kPa above atmosphere for normal voice). The fluid pressure on the superior surface oscillates, with peak values around 0.3 kPa, and a mean value slightly (around 0.1 kPa) above the ambient atmospheric pressure. The pressure distribution along the medial surface has been well documented, and involves a gradual decrease in pressure that is greater than the transglottal pressure (by 0.2 to 0.3 kPa) due to convective flow acceleration within the orifice (i.e. the so-called "Bernoulli effect"), followed by a fairly abrupt pressure recovery downstream of the region of flow separation. These flow studies have furthermore demonstrated that the geometry of the vocal fold model considered would indeed be relevant. Providing a step toward a more realistic and standardized modeling framework, Scherer *et al.* (2001) proposed a vocal fold geometry, referred to as M5, and investigated fluid flow through rigid static models. In the M5 model the vocal folds are represented as prisms with trapezoidal base. Fulcher *et al.* (2006) studied the response of spring mass models with excitation pressures based on measured M5 wall pressure data. A deformable continuum vocal fold model adopting the same basic geometry was proposed by Thomson *et al.* (2005). There experiments were conducted using deformable, self-oscillating models to study the energy

[a]Electronic mail: siegmund@purdue.edu
[b]Present address: Department of Mechanical Engineering, McGill University, Montreal, Quebec H3A 2K6, Canada.

transfer mechanisms between the flow and the radiated sound, and to verify detailed computer models of the coupled fluid–structure interactions. It was demonstrated that the cyclic deformation of the deformable vocal fold model is essential in the energy transfer from flow to solid, and the subsequent development of stresses and strain in the vocal fold. To investigate the details of the VF stress distribution due to flow induced deformation, numerical studies of fluid–structure interactions within continuum laryngeal models have been performed. Of particular interest to the present study are investigations related to vocal fold contact. The first work aiming to determine vocal fold contact pressure is that of Jiang and Titze (1994). This work outlined the three phases in vocal fold contact—an impact phase, a progressive pressure buildup phase, and an open phase. In de Oliveira Rosa *et al.* (2003) VF contact was considered, and the importance of incomplete closure was highlighted but no details on the resulting deformation fields was given. Details of dynamic contact conditions in individual vibration cycles—neglecting fluid pressure—were investigated by Gunter (2003) using finite element models with high spatial resolution. Alternatively, Horáček *et al.* (2005) employed a lumped mass mechanical representations of the VFs based on rigid mass–spring–damper systems (Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972) with contoured rigid walls to investigate vocal fold collision conditions and the associated normal tractions on the medial surfaces of the vocal folds during collision. These authors predicted impact pressure values of the order of 3 kPa. The treatment of the vocal folds as rigid bodies, however, does not allow the effects of the deformation of the vocal folds to be considered. Later, a refined analysis of impact stress was performed by Tao *et al.* (2006) using a dynamic finite element model that included contact and coupled flow–structure interactions. Contact pressures reported in that study range up to 6 kPa. Direct experimental measurements of the contact pressure have been reported (Verdolini *et al.*, 1999) using a miniature pressure transducer *in vivo*. Measured contact pressure values along the membranous portion of the vocal folds ranged from 0.5 to 3.0 kPa. Subsequently, Jiang *et al.* (2001) related pressure sensor data to photoglottography.

Imaging techniques for the measurement of VF deformations are of interest for possible clinical protocols. Progress has been made in the use of optical techniques for the study of VF vibrations using (endoscopic) high-speed imaging (Tigges *et al.*, 1996). This method demonstrated that vocal fold motion, notably the movement of the free vocal cord edge, can be tracked during the oscillatory cycle. Videokymography has been used to construct digital image sequences for the visualization of vibration patterns. Švec *et al.* (2000) demonstrated the usefulness of this method by qualitatively relating vibrational modes to the magnitude of vocal fold deformation. Attempts have been made to use laser-based optoreflectometry to determine VF displacement (Ouaknine *et al.*, 2003). This method yielded qualitative information on the inferior–superior surface motion. Methods based on triangulation algorithms, and the tracking of a single laser-generated speckle and its reflected image have allowed the measurements of the displacement of individual points on the superior surface of the VFs (Manneberg *et al.*, 2001). Full-field measurements of tissue deformation have been made using a microsuture technique on exercised canine larynges; the results were used to study the different vibrational modes of the VF medial surface (Berry *et al.*, 2001) and provided insight into mucosal wave propagation. The use of the electronic speckle pattern interferometry to analyze VFs has been proposed (Gardner *et al.*, 1995). This method has been found suitable, in principle, for noninvasive, full field strain measurements, but is limited to small deformation magnitudes.

These limitations can be overcome by the use of digital image correlation. Digital image correlation is a method that provides full field displacement fields on the surface of deformable structures. The method is noncontact and required that a speckle pattern be deposited on the structure of interest, e.g., Sutton *et al.* (1986). Mantha *et al.* (2005a, b) first used this method to study conditions on the superior surface of the vocal fold model of Thomson *et al.* (2005). Berry *et al.* (2006) used a similar VF model to study medial surface dynamics, also using a digital image correlation (DIC) method.

The goal of the present study was to conduct full field, time resolved measurements of the structural deformation of a VF model during self-oscillation using digital image correlation. This method is advantageous since it allows for large displacement magnitudes. Digital images were obtained using a high speed digital camera. Three-dimensional (3D) DIC was performed by projecting image pairs onto a single image frame using a mirror system. Images of the superior surface of a VF physical model were considered, because this surface is visible in laryngoscopic evaluations. While the displacements are the primary measured quantities in DIC, strains and strain rates can be obtained by appropriate data processing.

A deformable physical model of the human VFs was used in order to study the mechanical response in a controlled setting, without the need for animal models or excised larynges. The material used to produce the vocal fold model was silicone rubber. Silicone models are convenient for fundamental investigations, at the expense of a lack of realism in the tissue properties and the vocal fold geometry. The model vocal folds were embedded into rigid Plexiglas™ supports. The test rig supporting the laryngeal model did not allow tension adjustments of the vocal folds. This model was nevertheless deemed appropriate for the main goal of the investigation, which was to determine the usefulness of 3D DIC for the determination of mechanical states of the superior surface of the model, and to establish a process to obtain contact pressure data from superior surface data.

II. METHODS

A. The physical vocal folds model

The M5 vocal fold geometry reported by Scherer *et al.* (2001) was adopted for construction of the synthetic vocal folds. Since several other investigators have employed this model, Thomson *et al.* (2005), Fulcher *et al.* (2006), it allows for comparison between independent studies. Each fold con-

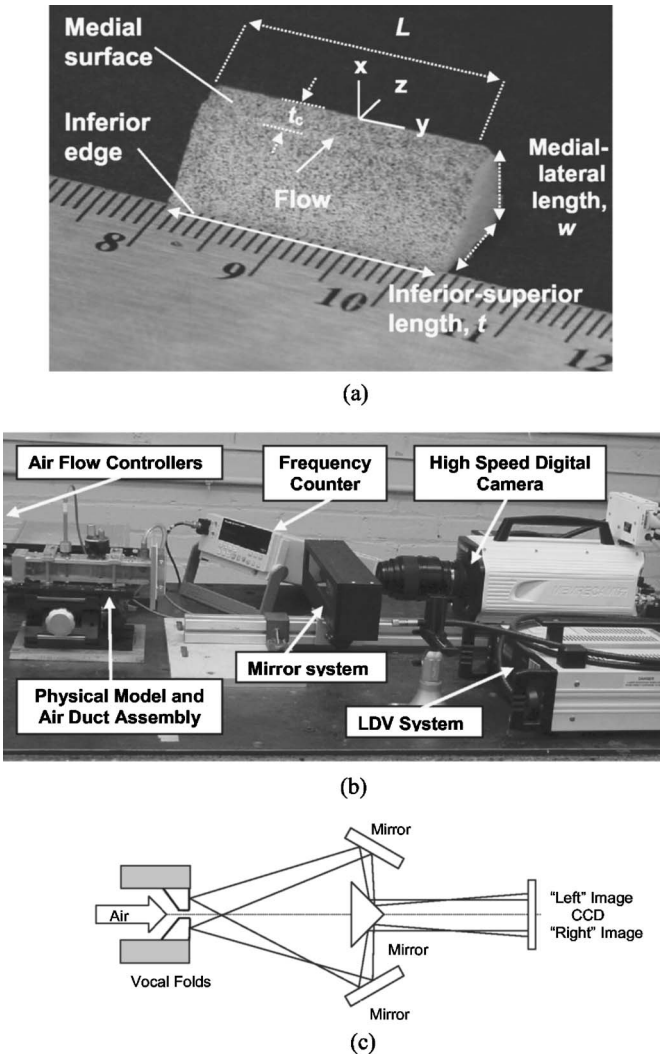


FIG. 1. (a) Characteristic dimensions of the model vocal folds, (b) experimental setup, and (c) mirror system.

sists of a prismatic extrusion of an approximately trapezoidal base with the inclined side facing the airflow, as illustrated in Fig. 1(a). The coordinate system is such that the x, y, and z axes point along the medial-lateral, the anterior–posterior, and the inferior–superior direction (i.e., the flow direction), respectively. The model inferior–superior length was $t=10.7$ mm, its medial-lateral length was $w=8.4$ mm, and the VF length was $L=22$ mm. This length of the vocal fold is on the upper bound of physiological data, Zhang *et al.* (2006). However, while the model possesses rigid boundaries an actual vocal fold does not. The anisotropic multilayer VF structure was simplified using an isotropic and homogeneous solid. The model VFs were cast from silicone rubber (Ecoflex; Smooth-On, Inc.) with a 1:4 ratio of liquid silicone rubber to silicone thinner, selected to obtain material properties with a modulus similar to that of soft tissue. The silicone rubber was characterized as an incompressible elastic solid material (Kerdok *et al.*, 2003). The shear modulus G was determined through instrumented hardness testing (ELF 3200; Bose) with a flat cylindrical indenter of radius r (Pawlak and Keller, 2002). The shear modulus was obtained from

$$G=\frac{1}{8r}\left(\frac{dP}{d\delta}\right), \tag{1}$$

where the indentation force is P, the indentation depth is δ, and $dP/d\delta$ is the slope of a P-δ regression. The indentation tests were performed for the conditions $\delta_{max}=0.85$ mm, $\dot{\delta}=0.15$ mm/s, and $r=3.0$ mm. The shear modulus was found to be $G=2.0$ kPa, based on five repeated experiments. When incompressibility is assumed, these measured properties are consistent with an elastic modulus of $E=6.0$ kPa. The model VFs were embedded into Plexiglas enclosures of thickness t with rectangular cut-outs of dimensions $L\times w$. Two symmetric enclosures were clamped using screws running in the medial-lateral direction within the plates. The outward normal of the VF superior surface coincided with the flow direction. In absence of flow, there was no visible gap between the VFs in the prephonatory state. Talc was applied to the medial surface to minimize the amount of adhesion during collision.

B. Flow supply and optical measurement setup

Airflow was generated by a centrifugal compressor. A 36-cm-diam and 48-cm-length cylindrical plenum chamber lined with 2.54-cm-thick fiberglass was used as an expansion chamber muffler to attenuate pressure fluctuations associated with noise in the flow supply. The flow path downstream of the muffler consisted of a 15-m-long flexible tube, and a rigid rectangular test section of cross section 23×22 mm and length 200 mm, to which the Plexiglas enclosure containing the model VFs was attached. Key components of the experimental setup are shown in Fig. 1(b). The tube termination was nearly anechoic. The measured reflection factor was approximately 0.15 at the self-oscillation frequencies observed in the experiment. Analog pressure regulators downstream of the compressor were used to control the airflow. A volumetric flow meter (MKS 558A; MKS Instruments) located upstream of the expansion chamber was used to measure the flow rate. Flush-mounted microphones were located 3.2 and 12 cm from model VFs upstream. The output signals from the two microphones, the volumetric flow meter, and a frequency counter (Fluke 45 DDM; John Fluke Mfg. Co., Inc.) were fed to a data acquisition system (Siglab 20-22A; Spectral Dynamics Inc.) sampling at 10 000 Hz with a voltage resolution of 0.01 V. For verification experiments, a laser Doppler velocimetry system (OFV 3000; Polytec Inc.) was employed.

Images of the superior surface of the model VF were obtained using a CMOS high-speed digital camera (Memrecam FX K3; NAC Image Technology Inc.) recording at 3000 frames/second with an open shutter. Images obtained with the high speed camera were processed to provide kymograms. Thereby, a single pixel row was obtained from each image and subsequent pixel rows were stacked on each other. The pixel size of the image chip was 0.011×0.011 mm. The pixel size relates to the resolution of the displacement field and the speckle size [the resolution of the DIC method is sensitive to the speckle size, e.g., see Hung and Voloshin (2003)]. A lens with focal length of 105 mm

(AF Micro Nikkor; Nikon Inc.) was used. A two-channel fiber optic goose-neck light with a quartz source (Model 21AC; Edmund Optics Inc.) provided uniform illumination. The orientation of the lights was optimized to reduce glare. Due to the remote location of the quartz light source a fiber optic light system avoids heating of the illuminated surface. Image pairs of the VF superior surface in undeformed and deformed configurations were obtained. A mirror system, composed of a central right angle prism with mirror coatings on the two legs and two lateral mirrors, was placed in the optical path, Fig. 1(c). Images from two viewing directions (from the left and the right side) were thus projected onto the single CMOS array.

The stereo DIC system based on a single camera with a mirror system requires a particular calibration procedure. The intrinsic parameters (i.e., focal length, aspect ratio, image center, and lens distortion) were calculated by using a grid target (a square grid of uniformly spaced white dots on a black background) positioned at various angles relative to the optical axis of the camera lens without the mirror system in place. The extrinsic parameters (i.e., the relationship between the two images produced by the mirror system) were then obtained by using a target with random speckle patterns, and two hash marks separated by a known distance, with the mirror system in place. Further details are documented in Spencer (2007).

The model VF superior surfaces were prepared with a random black/white speckle pattern. The silicone rubber was colored with a white pigment (Silc Pig; Smooth-On, Inc.). A spray of black enamel paint (1249 Flat Black Spray Enamel Paint; Testors) was deposited onto the superior surface resulting in a distribution with speckle diameter approximately 0.088 mm, corresponding to around 7 to 8 pixels.

C. Digital image correlation and analysis

The 3D-DIC analysis was performed using the commercially available program VIC-3D (Correlated Solutions Inc.). 3D DIC is a method that allows for measurements of the deformation of the surface of an object. Information regarding the algorithms used in this program can be readily found in the literature, i.e., in Chu *et al.* (1985), Sutton *et al.* (1986), Helm *et al.* (1996), on the in-plane correlation and in Schreier *et al.* (2004) on the stereoscopic aspects. A subset size, i.e., the size of the interrogation window in the numerical algorithm, was selected to be 23×23 pixels. All variables were obtained at a distance in excess of 0.5 mm away from the medial surface.

The output of the program includes the Lagrangian displacement field $\underline{u}(t)=(u,v,w)$ at each time step, obtained from $\underline{u}=\underline{x}-\underline{x}_0$ where $\underline{x}_0=(x_0,y_0,z_0)$ and $\underline{x}(t)=(x,y,z)$ represent global undeformed and deformed configurations, respectively. The velocity of points on the superior surface, $\underline{\dot{u}}=(\dot{u},\dot{v},\dot{w})$, was obtained numerically using a fourth order finite difference method. Strains $(\varepsilon_{xx},\varepsilon_{yy},\gamma_{xy})$ on the VF superior surface were calculated using (Sutton *et al.*, 2001)

$$\varepsilon_{xx}=\frac{\partial u}{\partial x}+\frac{1}{2}\left\{\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial w}{\partial x}\right)^2\right\},$$

$$\varepsilon_{yy}=\frac{\partial v}{\partial y}+\frac{1}{2}\left\{\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial w}{\partial y}\right)^2\right\},$$

$$\gamma_{xy}=\frac{1}{2}\left\{\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)+\left(\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\frac{\partial v}{\partial y}+\frac{\partial w}{\partial x}\frac{\partial w}{\partial y}\right)\right\}. \quad (2)$$

The incompressibility condition, $\varepsilon_{xx}+\varepsilon_{yy}+\varepsilon_{zz}=0$, allowed for the determination of the third strain component, $\varepsilon_{zz}=-\varepsilon_{xx}-\varepsilon_{yy}$, which yielded the complete strain vector $\underline{\varepsilon}=[\varepsilon_{xx},\varepsilon_{yy},\varepsilon_{zz},\gamma_{xy},0,0]^T$. Strain rates were again obtained from the strain time histories using a fourth-order finite difference method.

For an incompressible material, the deviatoric stress, $\underline{\sigma}^d=(\sigma^d_{xx},\sigma^d_{yy},\sigma^d_{zz},\tau_{xy},0,0)^T$, and the hydrostatic stress, σ^h, need to be determined independently [see, for example, Zienkiewicz and Taylor (2000)], in order to obtain the stress state $\underline{\sigma}=(\sigma_{xx},\sigma_{yy},\sigma_{zz},\tau_{xy},0,0)^T$. The stress state for the superior surface can be obtained by assuming the pressure on the superior surface, p_{supra}, to be known and identified with the stress component $\sigma_{zz}=p_{\text{supra}}$. Previous studies (Scherer *et al.*, 2001) of steady flows through glottis-shaped orifices have shown that p_{supra} is approximately equal to the atmospheric pressure, $p_{\text{supra}}=\sigma_{zz}\approx0$ on the superior surface. When there is no supraglottal acoustic load (i.e., if there is no confining air duct at the orifice discharge) the radiated sound pressure is also insignificant due to the poor radiation efficiency of the pulsated jet (Mongeau *et al.*, 1997). The hydrostatic stress σ^h was thus approximated as $\sigma^h=\sigma_{zz}-\sigma^d_{zz}=p_{\text{supra}}-\sigma^d_{zz}\approx-\sigma^d_{zz}$. The remaining stress components $\sigma_{xx},\sigma_{yy},\tau_{xy}$ were obtained from $\underline{\sigma}=\underline{\sigma}^d+\sigma^h\underline{m}$, with the deviatoric stress $\underline{\sigma}^d$ equal to

$$\underline{\sigma}^d=2G\left(\mathbf{I}_0-\frac{1}{3}\underline{m}\underline{m}^T\right)\underline{\varepsilon}. \quad (3)$$

The 6×6 matrix $\mathbf{I}_0$ possesses entries only along its diagonal, with the first three entries equal to one and the remainder equal to 1/2, and $\underline{m}=[1,1,1,0,0,0]^T$. As shown in the Appendix, the superior surface stresses are given by

$$\frac{\sigma_{xx}}{G}\approx4\varepsilon_{xx}+2\varepsilon_{yy} \quad (4)$$

and

$$\frac{\sigma_{yy}}{G}\approx4\varepsilon_{yy}+2\varepsilon_{xx}. \quad (5)$$

D. Evaluation of measurement accuracy

Errors in high speed DIC measurements were investigated by Siebert *et al.* (2007), and may be categorized as (1) correlation errors and (2) reconstruction errors. Correlation errors include background noise in the image (which can be enhanced by increasing the image gain) and the quantization of grayscale values. In Siebert *et al.* (2007), either one of these errors was less than 0.05 pixels, at worst. The reconstruction errors were caused by uncertainties in the calibration of the 3D-DIC system. It was shown that reconstruction errors were on the order of 0.02 pixels. It is assumed that errors were of the same magnitude and deemed to be negli-

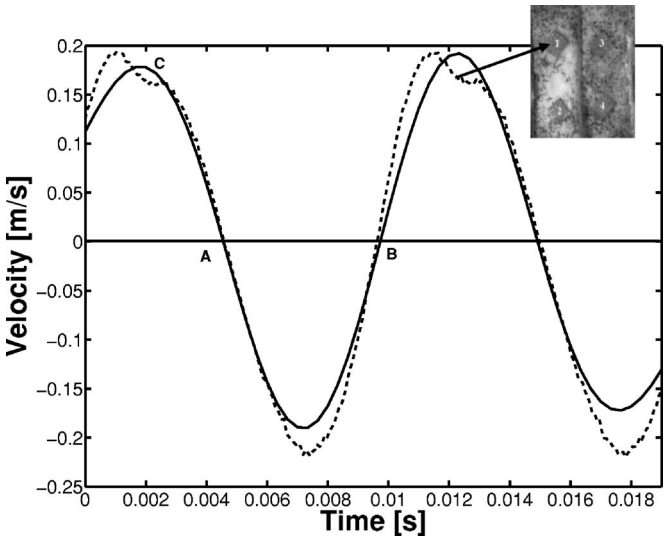


FIG. 2. Inferior–superior velocity from LDV (---) and numerical differentiation of DIC displacement data (—). The insert photograph shows the vocal fold model with the measurement locations for LDV. Results are for point (1). Instants of open (A), closed (B), and maximum surface velocity (C) are indicated.

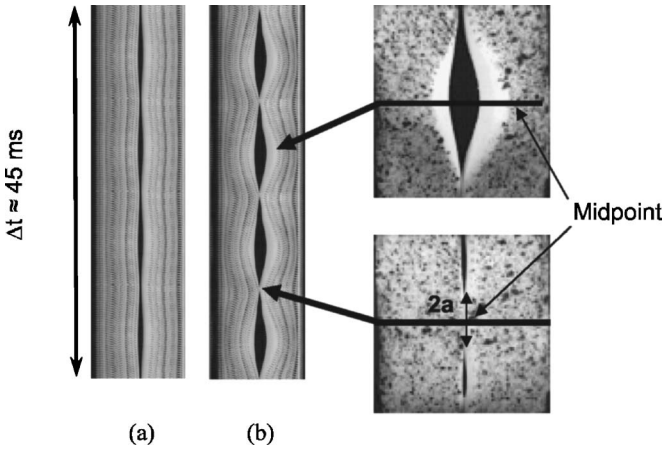


FIG. 3. Kymograms at: (a) $Q=406\ \mathrm{cm}^3/\mathrm{s}$ (the phonation onset), and (b) $Q=690\ \mathrm{cm}^3/\mathrm{s}$. Supplementary images indicate kymogram location: $z=\mathrm{const}$ at $y=0$. Distance $2a$ indicates the anterior–posterior extension of the contact area.

gible in the present study, despite the differences in optics and equipment. Comparisons between results from DIC and a strain gauge for a tensile test indicated maximum strain differences of 0.03 millistrains. The influence of speckle size on the accuracy of DIC was investigated by Hung and Voloshin (2003). It was estimated that a speckle size of roughly two to eight times the pixel size typically yields the most accurate results.

A LDV with a sampling rate of 11 800 Hz and resolution of 0.01 mm/s was used to verify the accuracy of the inferior–superior component of velocity obtained from the 3D-DIC method. Reflective patches were attached to the superior surface of a second VF model with $G=2.9$ kPa (slightly stiffer than the model described in Sec. II A), as shown in the inset photo at the upper right portion of Fig. 2. The transverse Eulerian velocity component along the direction of the laser beam, i.e., the inferior–superior direction, was measured and compared to that obtained from 3D-DIC measurements. Velocities from DIC are in the Lagrange framework (material point tracking) while LDV data are in the Eulerian framework and track a fixed region in space. In order to reduce the noise associated with the finite difference approximation in the DIC data, a low-pass filter (127 point linear-phase FIR filter using the Parks-McClellan algorithm with passband frequency of 200 Hz and stopband frequency of 400 Hz) was incorporated. Overall good agreement between the velocities from the LDV and the 3D-DIC analysis was obtained, as shown in Fig. 2. RMS values of 0.1326 and 0.1469 were found for the filtered DIC and LDV data, respectively. This implies a relative error of 9.7%. The errors were attributed to the influence of the addition of the reflecting patches, which seemed to affect the VF vibration, and the fact that LDV averages velocity over the beam diameter. The change in the orientation of the superior surface during vibration, and the errors introduced by the fourth-order finite difference scheme may also have contributed. It was felt that the errors were mostly intrinsic to the experimental design, and not representative of the accuracy of the experimental data.

III. RESULTS

A. Flow, pressure, frequency, and high-speed imaging

The volumetric flow rate, Q, was increased stepwise until self-oscillation was initiated, as detected through the use of a frequency counterfed with the output signal from one microphone located in the subglottal region. At oscillation onset, the volumetric flow rate was $Q_{\mathrm{th}}=406\ \mathrm{cm}^3/\mathrm{s}$, and the frequency was $f_{\mathrm{th}}=88.9$ Hz. The corresponding onset pressure was $p_{\mathrm{th}}=0.87$ kPa. For $Q_{\mathrm{th}}\leqslant Q\leqslant 813\ \mathrm{cm}^3/\mathrm{s}$, the relationship between pressure p and flow rate Q was characterized by the linear regression $p[\mathrm{kPa}]=0.0009Q[\mathrm{cm}^3/\mathrm{s}]+0.5263$, with $R^2=0.99$. High speed imaging indicated model VF collision from a flow rate, referred to as the collision onset, of $Q=Q_C=550\ \mathrm{cm}^3/\mathrm{s}$. The oscillation frequency reached a maximum, $\max(f)=90.1$ Hz, at $Q=Q_C$. The frequency, f, dropped very slightly with flow rate beyond the collision onset, from 90.1 Hz to $f=88.9$ Hz at a flow rate of $Q=750\ \mathrm{cm}^3/\mathrm{s}$. The energy dissipated by collision and adhesion is most likely responsible for this trend.

The repeatability and variability of the measurements were investigated. In onset experiments, repeated flow rate increases followed by 10-min-long interruptions yielded consistent Q_{th} values. Steady, constant flow rates were maintained for periods of 45 min ($Q=\mathrm{const}>Q_{\mathrm{th}}$). For $Q<Q_C$ the frequency remained constant over time with a maximum deviation less than 0.1 Hz ($\Delta f<0.1$ Hz), while for $Q>Q_C$ the frequency decreased slightly with time with a maximum deviation of 0.8 Hz ($\Delta f=-0.8$ Hz).

Kymogram images of the midline of the model ($y=0$) are shown in Fig. 3. The kymogram was constructed from one single vibration cycle; two images of this cycle—for the open and closed state—are shown in Fig. 3. The locations of the pixel lines used for the kymogram images are indicated

in Fig. 3. The kymograms clearly show differences between the vibration patterns for $Q=Q_{th}$ and for $Q=690\ \mathrm{cm}^3/\mathrm{s} > Q_C$. At $Q=Q_{th}$, Fig. 3(a), no collision between the model VFs was observed. At the higher flow rate, Fig. 3(b), the maximum glottal opening was much increased. In the closed state, collision of the medial surfaces was clearly visible; however, closure of the glottal opening was not complete. Two distinct orifice openings at the anterior and posterior ends of the model larynx were visible during the closing stage. Kymograms, however, do not provide information on motion in the inferior–superior direction.

B. Displacements, strains, and stresses

The DIC method was used to establish the time history of the superior surface displacements in order to characterize the glottal opening process. Figure 4 shows the displacements in the inferior–superior direction, w, and the medial-lateral direction, u, for flow rates $Q=Q_{th}=406\ \mathrm{cm}^3/\mathrm{s}$ and $Q=690\ \mathrm{cm}^3/\mathrm{s} > Q_C$. The displacement data are for locations along a line parallel to the medial surface, 1.5 mm away from the centerline of the undeformed model VFs. Figures 4(a) and 4(b) show that, for $Q=406\ \mathrm{cm}^3/\mathrm{s}$, the largest inferior–superior vibratory displacements occurred in the center of the model fold and VF motion, over a span of around $L/2$. The anterior and posterior extremities of the folds were almost fixed during oscillation. Figures 4(c) and 4(d) show the displacements in the medial-lateral direction. The model larynx remained in an open state over the entire cycle. The main transverse vibratory displacements occurred again in the central section of the vocal fold, over a span (of around $3L/4$) larger than that for the inferior–superior displacement. The displacements along a line parallel to the superior edge clearly show a third-order mode of vibration as visible by the three max–min values of displacement, Figs. 4(e) and 4(f).

Figures 4(e) and 4(f) show that, for $Q=690\ \mathrm{cm}^3/\mathrm{s}$, the largest inferior–superior vibratory displacements during opening again occurred in the center of the model fold. VF motion spanned a region roughly one-half the VF length during opening. During closing, a larger portion of the VF span was actively vibrating. During closing it was observed that the curvature was reversed. Figures 4(g) and 4(h) show the displacements in the medial-lateral direction. The active span was similar to that for the onset flow rate.

Time histories of the medial-lateral and inferior–superior displacements and velocities at the midpoint of the model VF, $\underline{x}_0=(0.5,0,0)$ mm, oven an oscillation period are shown in Fig. 5 for flow rates $Q=Q_{th}=406\ \mathrm{cm}^3/\mathrm{s}$ and $Q=690\ \mathrm{cm}^3/\mathrm{s} > Q_C$. Figure 5(a) shows the inferior–superior displacement component. The maximum displacement for both flow rates was reached at the instant of maximum open area, as for y=0 in Fig. 4. The deformation magnitude for $Q=690\ \mathrm{cm}^3/\mathrm{s}$ was approximately 50% larger than that for the lower flow rate. The inferior–superior velocity component is shown in Fig. 5(b). Maximum velocities were approximately 700% larger for $Q=690\ \mathrm{cm}^3/\mathrm{s}$ than for the lower flow rate. Figure 5(c) shows the medial-lateral displacement. Positive values generally indicate collision for the higher flow rate. At $Q=690\ \mathrm{cm}^3/\mathrm{s}$, the peak medial-lateral deformation was increased by 150% compared to Q_{th}. In the closed state, the positive value for $Q=690\ \mathrm{cm}^3/\mathrm{s}$ indicates that the medial motion of the midpoint crossed the midline reference location. Figure 5(d) shows the medial-lateral velocity component. Similar to the inferior–superior velocity component, the maximum surface velocity was much increased for the higher flow rate.

The dependence of the anterior–posterior ε_{yy} and medial-lateral ε_{xx} strains on flow rate was determined for the midpoint of the model and for $Q > Q_{th}$. Midpoint values at the times corresponding to maximum (open state) and minimum (closed state) glottal areas are shown in Fig. 6. At the phonation onset $Q_{th}=406\ \mathrm{cm}^3/\mathrm{s}$, the absolute values of ε_{yy}^{open} and $\varepsilon_{yy}^{closed}$ were similar. For flow rates between the phonation onset and the collision onset $Q_{th} < Q < Q_C$, magnitudes of ε_{yy} depend linearly on flow rate. Values of ε_{yy}^{open} increase monotonically with Q while $\varepsilon_{yy}^{closed}$ decreased with Q but remained tensile. For $Q \geqslant Q_C$, $\varepsilon_{yy}^{closed}$ finally was compressive and reached a nearly constant value as flow rate was increased.

Midpoint values of the medial-lateral strain ε_{xx} in the open and closed states are shown in Fig. 6(b). These quantities were nearly zero at $Q=Q_{th}$. Values of ε_{xx}^{open} were compressive and increased monotonically in magnitude with flow rate. For $Q < Q_C$, values of $\varepsilon_{xx}^{closed}$ were positive and increased linearly, with $|\varepsilon_{xx}^{closed}|=|\varepsilon_{xx}^{open}|$. For larger flow rates, above $Q=Q_C$, the values of $\varepsilon_{xx}^{closed}$ reached a maximum around a value of 0.06, and decreased slightly. Extrapolation of the curve at closed state yields a medial-lateral strain estimate for vibration in absence of collision.

Strains at the midpoint of the model VF over one oscillation period are shown in Figs. 7(a) and 7(b) at $Q=Q_{th}=406\ \mathrm{cm}^3/\mathrm{s}$ and $Q=690\ \mathrm{cm}^3/\mathrm{s}$. The medial-lateral strain component is shown versus time in Fig. 7(a). For $Q=690\ \mathrm{cm}^3/\mathrm{s}$, strain was constant over the time period where the orifice was closed (at least over the midspan region). The impact duration was estimated to be 0.03 s based on the period of constant medial-lateral strain during closure. Figure 7(b) shows the anterior–posterior strain component. Strains at the higher flow rate were much increased. Impact of the vocal folds is indicated by negative values. The duration of impact was consistent with that estimated from the medial-lateral strain history.

In the characterization of viscoelastic properties of soft materials strain rates are essential. Viscosity may generally be defined as the ratio of stress and strain rate. Strain rates obtained from a fourth-order finite difference approximation are shown in Figs. 7(c) and 7(d). For comparison purposes, mean values of strain rates can be estimated from the peak strain values of Fig. 6, and frequency, as $\dot{\bar{\varepsilon}}=2f|\varepsilon^{open}-\varepsilon^{closed}|$. This approximation is useful when measurement data are limited to states where surface velocities are small (for example, due to camera shutter speed limitations). This provided reasonable estimates of the time averaged average strain rates. For example, at a flow rate of $Q=690\ \mathrm{cm}^3/\mathrm{s}$, the mean strain rates were found to be $\dot{\bar{\varepsilon}}_{yy}=58\ \mathrm{s}^{-1}$ and $\dot{\bar{\varepsilon}}_{xx}=35\ \mathrm{s}^{-1}$, whereas the approximation from discrete values yields $\dot{\bar{\varepsilon}}_{yy}=62\ \mathrm{s}^{-1}$ and $\dot{\bar{\varepsilon}}_{xx}=33\ \mathrm{s}^{-1}$, Fig. 7. Consequently,

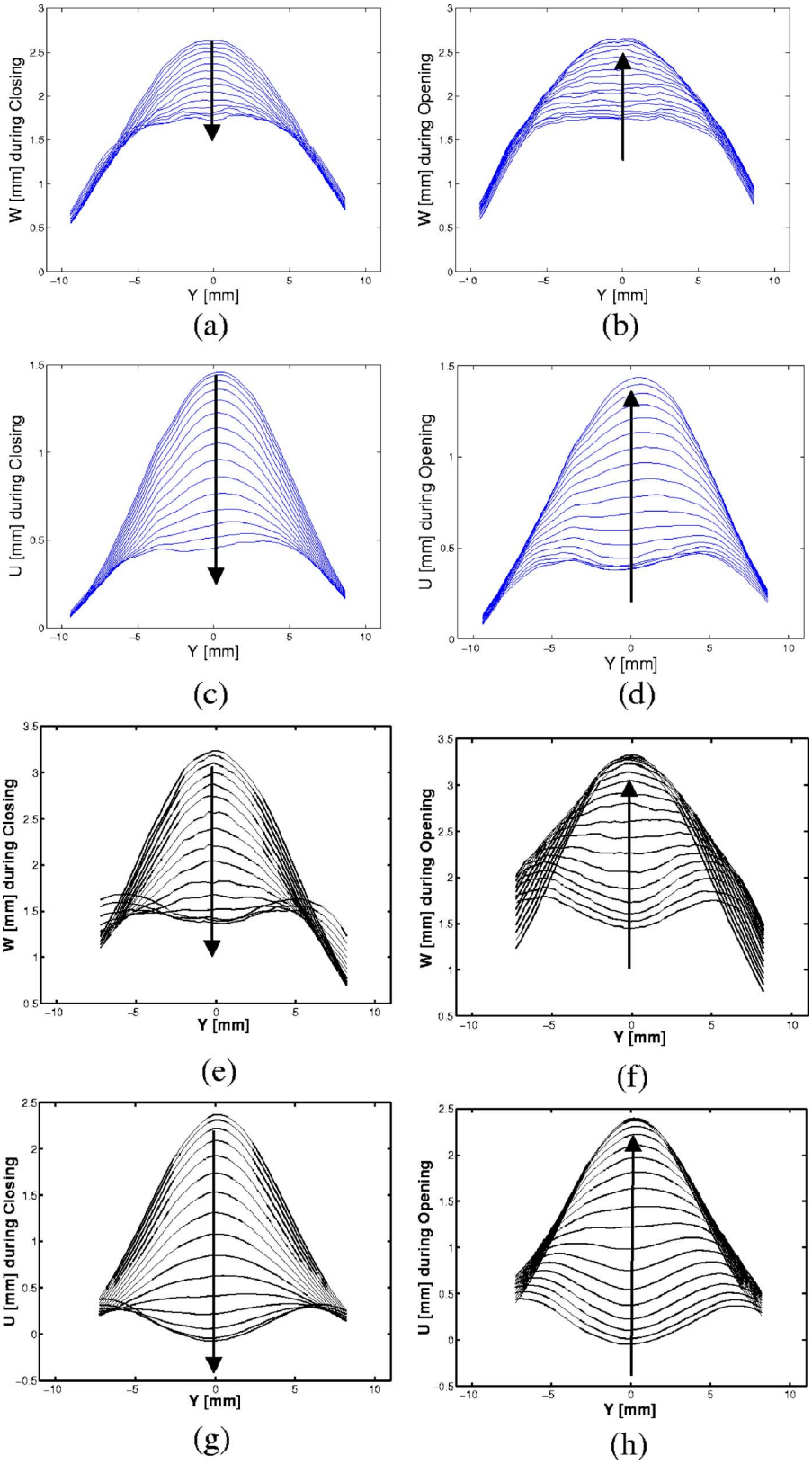


FIG. 4. (Color online) Inferior-superior (a), (b) and medial-lateral (c), (d) component of displacements along a line parallel and removed 1.5 mm from the medial surface at two flow rates: (1) Phonation onset $Q=406\ \mathrm{cm}^3/\mathrm{s}$ and (2) $Q=690\ \mathrm{cm}^3/\mathrm{s}$. The time increment between measurements is 0.363 ms.

mean strain rate values can be assessed from Fig. 6. Mean values of strain rates $\dot{\bar{\varepsilon}}_{yy}$ increase monotonically with flow rate. The medial-lateral mean strain rate $\dot{\bar{\varepsilon}}_{xx}$ reaches a limit since VF collision limits strain rate magnitudes in the medial-lateral direction. Mean strain rate values were much smaller than the peak instantaneous values, indicating a large crest factor. This may be significant since, for Newtonian materials, strain rates are often approximated as the product of the mean strain and frequency. This, along with the fact that vocal fold tissue is non-Newtonian, could lead to significant errors in the estimates of the viscosity. Direct strain rate measurement is thus beneficial and one possible advantage

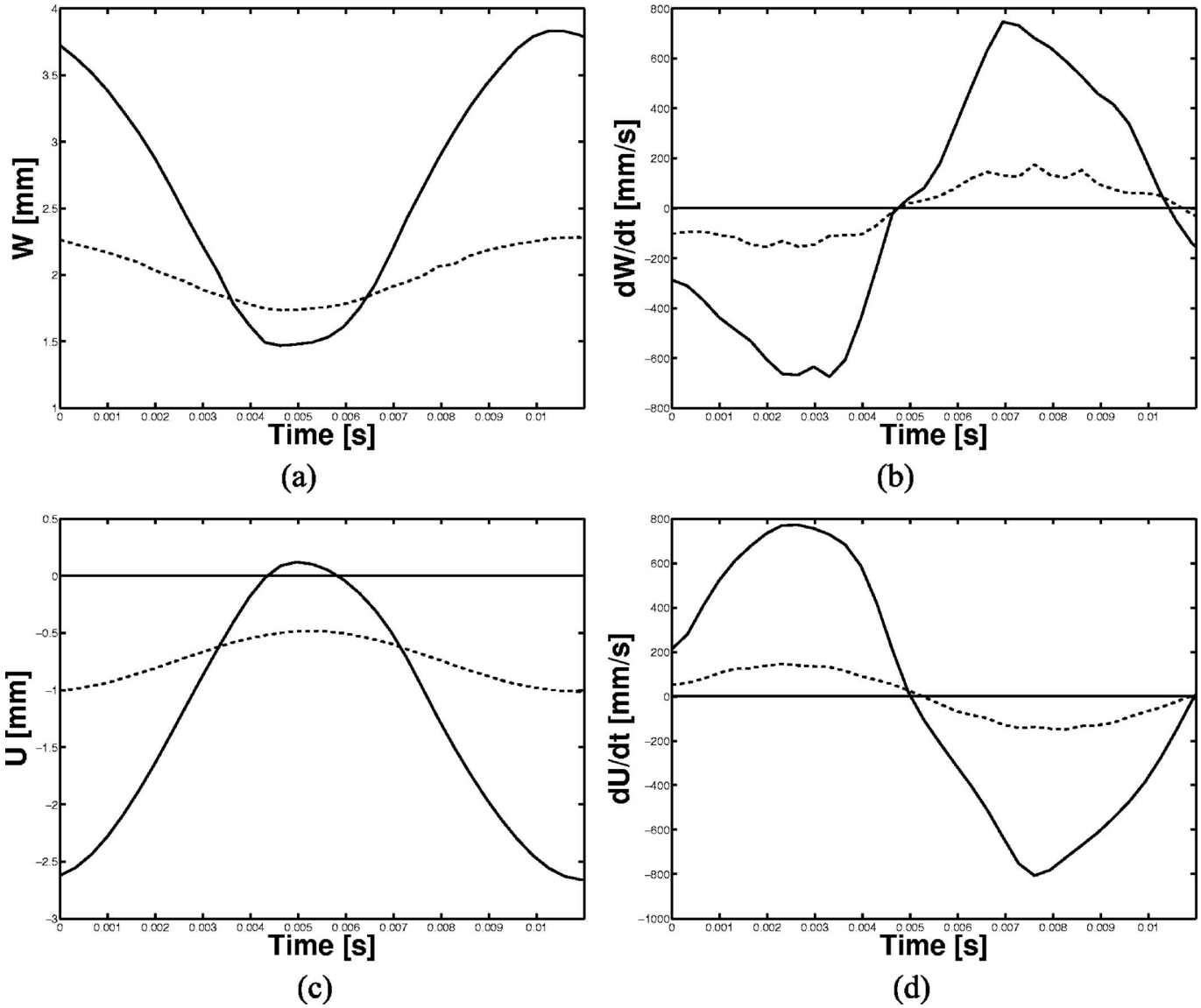


FIG. 5. Time history plots at midpoint of superior surface at phonation onset $Q=406$ cm^3/s (---) and $Q=690$ cm^3/s (—); (a) inferior–superior displacement, (b) inferior–superior velocity, (c) medial-lateral displacement, and (d) medial-lateral velocity.

of high-speed DIC (as opposed to simpler procedures based on images obtained at low sampling rates). Figure 7(c) shows the medial-lateral strain rate. The strain rate was significantly larger for $Q=690$ cm^3/s than for Q_{th}. Collision was indicated by a zero strain rate value. The maximum value at $Q=690$ cm^3/s was nearly twice the mean value. Similar trends were observed for the anterior–posterior strain rates, Fig. 7(d).

Figure 8 shows the distribution of the anterior–posterior ε_{yy} and medial-lateral ε_{xx} strain components on the deformed superior surface at a flow rate of $Q=690$ cm^3/s$>Q_C$ in the open and closed states. Figure 8 shows a composite view of the state of deformation along with strain data. In the open state, ε_{yy} was tensile for a large central part of the model VF with the maximum at the model VF center, as shown in Fig. 8(a). The region of compressive anterior–posterior strains extended to a distance $L/3$ from the frame edge. During closure, the characteristics of the ε_{yy} field changed significantly, Fig. 8(b). At the midsection of the model VFs, the curvature was reversed and ε_{yy} was compressive while tensile strains dominated in the region closer to the enclosure. The maximum tensile value of ε_{yy} during closing was well below the maximum tensile strain value observed during opening. In the open state, ε_{xx} was compressive for most of the model VF with the minimum at the model VF midsection adjacent to the enclosure, Fig. 8(c). The compressive strain level was reduced near the model VF center. Tensile strains were found only along the surfaces adjacent to the anterior and posterior faces of the enclosure. During closure, the characteristics of the ε_{xx} field changed significantly, Fig. 8(d). Most of the superior surface was found to be under tensile medial-lateral strain. These strains were a maximum in the midsection, i.e., the concave domain, with a strain value 30% above that found near VF center, see Fig. 6(b). Asymmetries in the strain fields between the left and right vocal folds are believed to be caused by imperfect placement of the model VFs in the Plexiglas enclosure.

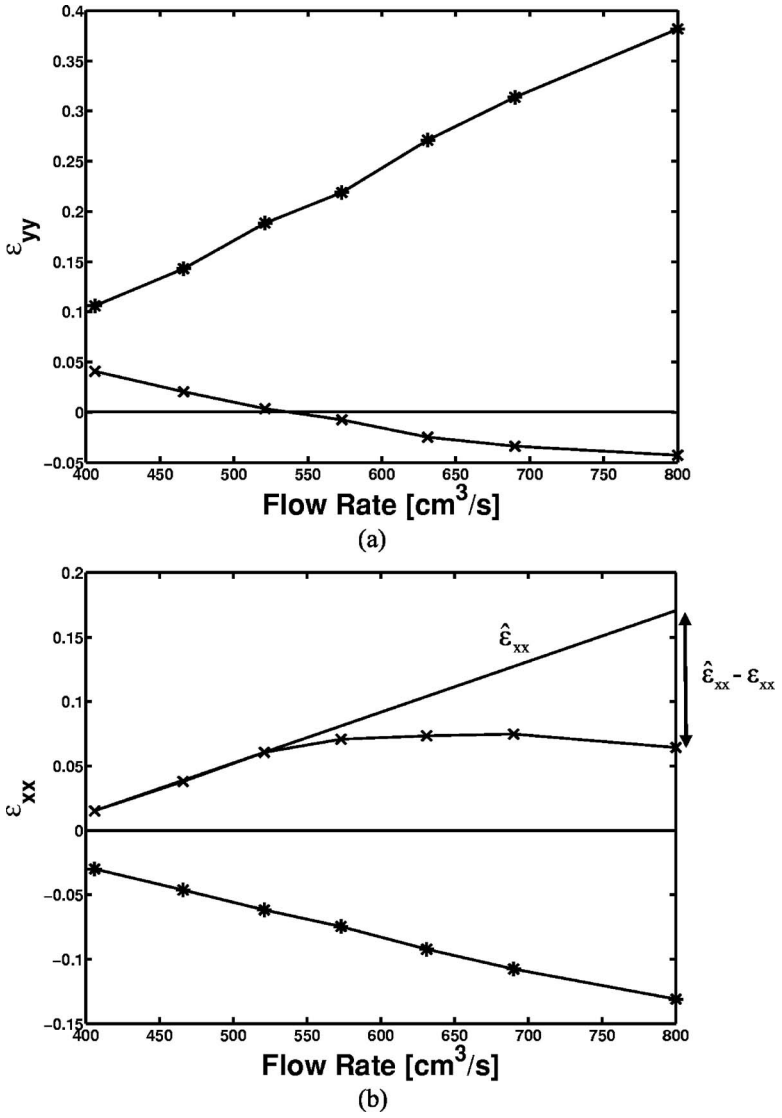


FIG. 6. Strain vs flow rate at midpoint: (a) Anterior–posterior strain ε_{yy} and (b) medial-lateral strain ε_{xx}. Symbols indicate measured data for open (*) and closed (x) states values. Strain data obtained for $y=0$. $\hat{\varepsilon}_{xx}$ is the linear extrapolation of ε_{xx} data in the closed state before collision onset.

Contours of stress on the superior surface are shown in Fig. 9. A maximum normalized anterior–posterior stress value of $\sigma_{yy}/G=1.10$ was reached at the model VF midpoint in the open state, Fig. 9(a), while compressive stresses, $\sigma_{yy}/G=-0.25$, were found at the fixed edges. Conversely, for the closed state, Fig. 9(b), $\sigma_{yy}/G=-0.05$ at the VF midpoint and $\sigma_{yy}/G=0.70$ at the fixed edges. Overall, the magnitude of stresses in the anterior–posterior direction was much reduced in the closed state when compared to the open state. In the open state, the normalized medial-lateral stress, Fig. 9(c), possessed its largest tensile value, $\sigma_{xx}/G=0.22$, at the midsection, while compressive stress, $\sigma_{xx}/G=-0.42$, was found near the medial face of the enclosure. In the closed state, Fig. 9(d), values for normalized stress were all positive. Maximum values were $\sigma_{xx}/G=0.44$ at the model VF midsection near the enclosure. The lowest value of the medial-lateral stress, $\sigma_{xx}/G=0.10$, was at the VF center. At this location, the mechanical state of the vocal folds was influenced by collision, as detailed in the following.

C. Hertzian impact model

Of special interest is the mechanical state of the model VFs in the closed state. It has been frequently hypothesized that the stresses resulting from VF collision are important factors in VF damage, although the exact injury mechanism is unknown. For VF collision to occur, the model VFs need to be stretched along the medial-lateral direction to overcome the glottal gap consisting of the initial prephonatory gap and the gap caused by the deformation due to the mean—or static—pressure on the inferior surface. Evidence of this behavior is found in the contours of medial-lateral strains and stress in the closed state, Figs. 7(d) and 9(d), respectively. The overall tensile deformation state in the medial-lateral direction may seem counterintuitive at first, but it is necessary for collision to occur. Up to the onset of collision, this tensile load state can develop freely without any constraint. Once collision occurs, collision causes new surface stresses and restricts VF motion along the midline.

A Hertzian impact model was used to estimate the collision pressure (Johnson, 2003). Horáček *et al.*, (2005) employed such a model in their computational model based on that of Stronge (2000). Considering rate independence and incompressibility, the normalized collision pressure at the center of the contact area on the medial surface, p_C/G, is given as

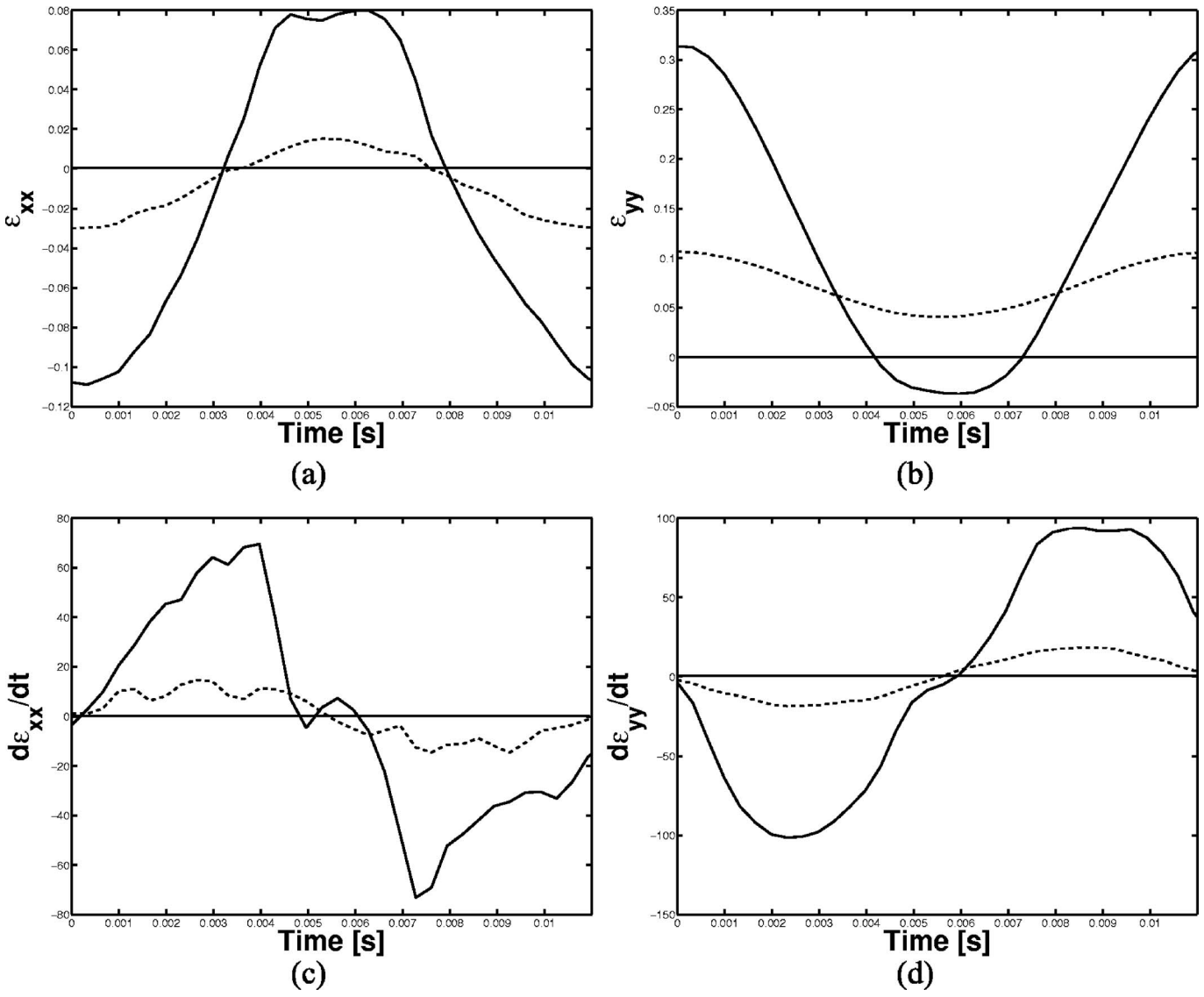


FIG. 7. Time history plots at midpoint of superior surface for phonation onset $Q=409$ cm^3/s (---) and $Q=690$ cm^3/s (—); (a) anterior–posterior strain, (b) medial-lateral strain, (c) medial-lateral strain rate, and (d) anterior–posterior strain rate.

$$\frac{p_C}{G} = \frac{4\delta_P}{\pi r_{\text{eq}}}. \tag{6}$$

In Eq. (6), r_{eq} is the radius of a circular contact region, and δ_P is the (hypothetical) penetration depth of the VFs through the contact plane. It is assumed that impacts occur at speeds that are low relative to the compressive wave speed, and thus are quasistatic (Johnson, 2003). This condition is automatically fulfilled for an incompressible material. The radius of the contact region was obtained from high speed images, as the one in Fig. 3. Contact was presumed to take place over a rectangular area of inferior–superior thickness t_c and anterior–posterior extension $2a$. Since the ellipticity of the contact is only mild $[t_c/(2a)<2]$ the use of a circular contact model with equivalent areas is assumed to introduce only insignificant errors (Greenwood, 2006). The equivalent contact radius r_{eq} was calculated by equating this contact area for the model $(2at_c)$ to that of an equivalent circular contact area (πr_{eq}^2), for consistency with Eq. (6) which was originally derived for a circular indenter. The inferior–superior thickness of the model is $t_c=3.63$ mm, see Fig. 1(a). The penetration depth of the model VFs through the contact was estimated from the measured medial-lateral strain at the model VF midpoint. Assuming that in the absence of a second model VF the medial-lateral strain could develop freely, an estimate of the unconstrained deformation of the model VF in that direction was obtained by extrapolation of the initial linear $\varepsilon_{xx}^{\text{closed}}-Q$ response, as illustrated in Fig. 6(b). The difference between the extrapolated strain, $\hat{\varepsilon}_{xx}$, and the actual measured medial-lateral strain, $\varepsilon_{xx}^{\text{closed}}$, provided the contact constraint imposed on VF deformation. For example for $Q=690$ cm^3/s, the difference between the extrapolated strain and the measured strain was $\hat{\varepsilon}_{xx}-\varepsilon_{xx}^{\text{closed}}=0.053$. The magnitude of the penetration was then obtained using

$$\delta_P = 2(\hat{\varepsilon}_{xx} - \varepsilon_{xx}^{closed})l^*, \tag{7}$$

where l^* is the depth of the VF influenced by the contact. The factor two accounts for the presence of the two VFs, assuming symmetry. Combining Eqs. (6) and (7) yielded the

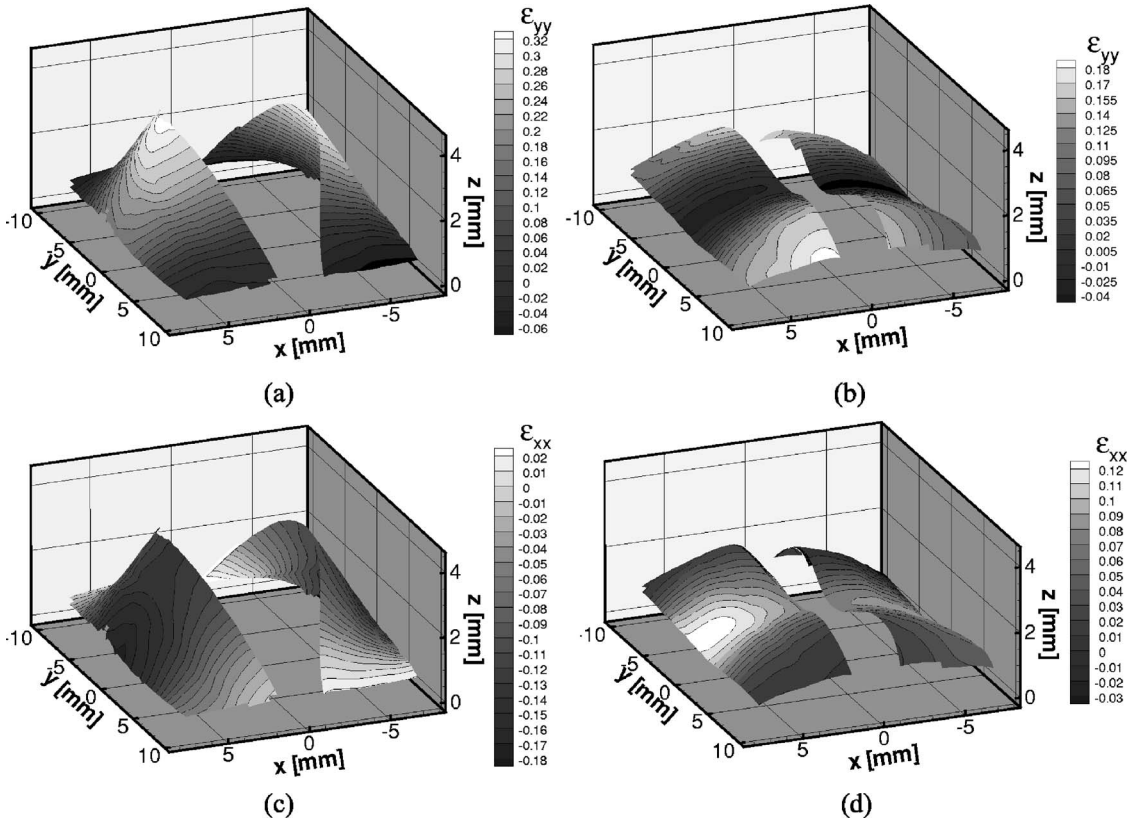


FIG. 8. Contour plots of strains on the deformed superior surface at $Q=690$ cm^3/s: (a), (b) Strain component in the anterior–posterior direction (ε_{yy}), and (c), (d) strain component in medial-lateral (ε_{xx}) direction, for open and closed states, respectively.

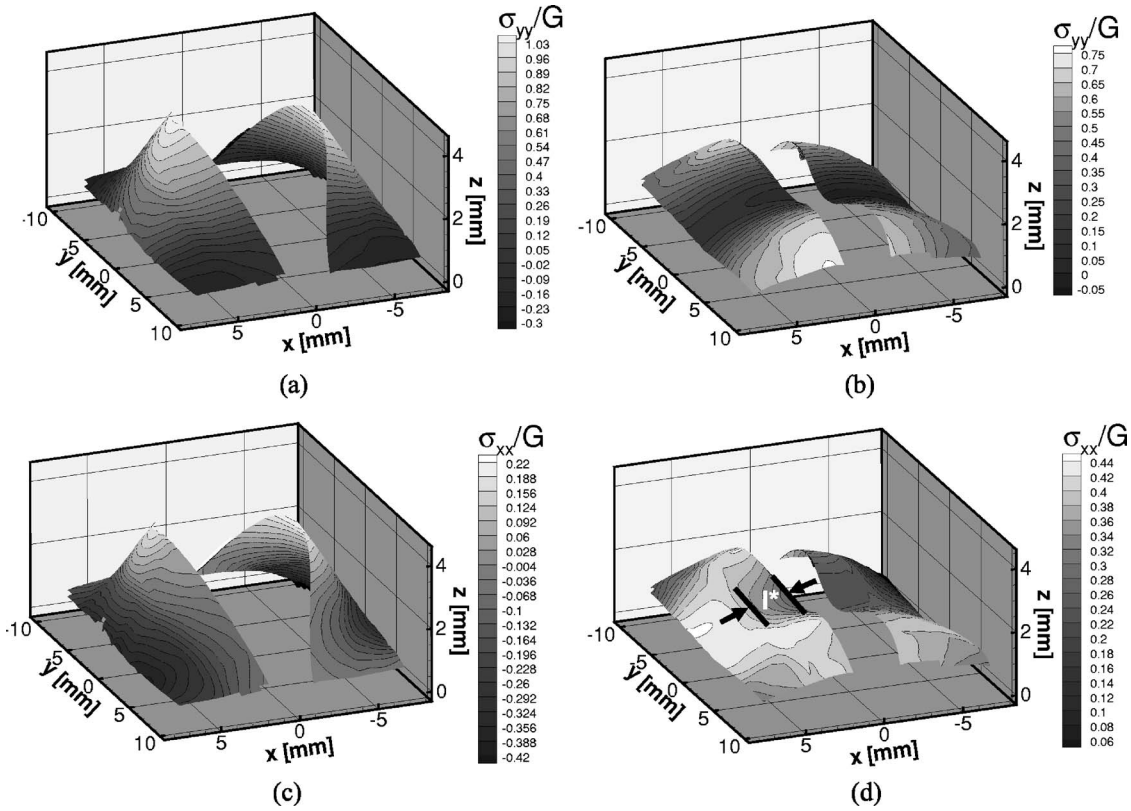


FIG. 9. Contour plots of normalized stress on the deformed superior surface at $Q=690$ cm^3/s; (a), (b) Stress component in the anterior-posterior direction (σ_{yy}), and (c), (d) Stress component in medial-lateral (σ_{xx}) direction, for open and closed states, respectively.

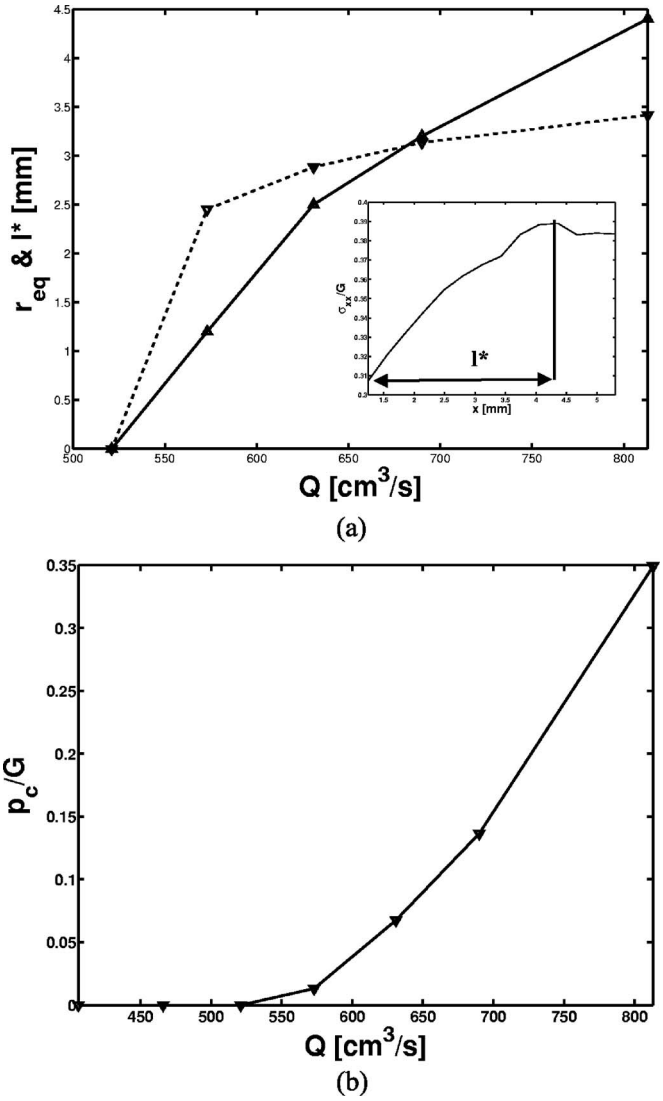


FIG. 10. (a) Equivalent radius, r_{eq} (---), and influence length, l^* (—), vs flow rate; insert shows the estimated influence length for Q=690 cm^3/s. (b) Estimated normalized contact pressure, p_c/G, vs flow rate, Q.

collision stress in terms of material properties, measured and extrapolated strains, and the geometry of the contact:

$$\frac{p_C}{G} = \frac{8}{\pi}(\hat{\varepsilon}_{xx} - \varepsilon_{xx}^{closed})\frac{l^*}{r_{eq}}. \quad (8)$$

The dependence of the equivalent contact radius r_{eq} on the flow rate is shown in Fig. 10(a) based on measurements of values of $2a$ from high speed images, e.g., Fig. 3(b). Estimates of the influence depth l^* were obtained from the distribution of the medial-lateral stress along the midline (y =const). l^* is the distance from the superior edge to the location on the superior surface where $\partial\sigma_{xx}/\partial x \approx 0$, see insert in Fig. 10(a). The resulting dependence of l^* on flow rate is given in Fig. 10(a). Figure 10(b) then depicts the resulting contact pressure on the medial surface in dependence of the flow rate. The contact pressure was found to vary nonlinearly with flow rate. The contact pressure was zero for flow rates below the collision onset Q_C. For greater flow rates, $Q > Q_C$, the contact pressure increased nonlinearly with flow rate. From the dependence of medial-lateral strains $\varepsilon_{xx}^{closed}$ on flow rate [Fig. 6(b)] and of the trends in r_{eq} and l^* [Fig. 10(a)], it can be concluded that the nonlinear response of the contact pressure is due to the saturation of the strain values due to collision.

IV. DISCUSSION AND CONCLUSION

The application of a 3D DIC method for the noncontact and relatively noninvasive (the method requires a paint speckle) determination of the mechanical state of a model vocal fold superior surface during self-oscillation was investigated. The method was successfully implemented and applied in a laboratory experiment using a physical model VF system. The measurements provided time-resolved, full field measurements of several mechanical states of possible interest in phonation studies, including the out-of-plane displacements, the glottal opening displacement, as well as the strain and strain rate fields. Indirect methods were investigated for the estimation of contact pressure from strain data on the superior surface. The use of high speed DIC was beneficial to determine the maximum strain rates directly. Stresses were calculated based on the assumption of incompressible material behavior. The results showed that collision reduced the medial-lateral stress near the region of contact on the superior surface of the model vocal folds. For collision to occur, the model VFs needed to be stretched in the medial-lateral direction in order to overcome the glottal gap. Superimposition of a compressive load due to contact actually reduced stress on the superior surface. Further work is needed to verify these trends, and the accuracy of the contact stress estimates from measured medial-lateral strains and contact radii obtained from high-speed images. The analysis may be biased by the fact that tissue damage, in general, occurs beneath the tissue and not on the superior surface. Therefore, estimates of the magnitude of penetration based on influence depth l^* may be incorrect. The data show that conditions leading to collision also lead to high strain rates, which may be a factor in tissue damage.

The physical and acoustical properties of the physical, synthetic VF model were deemed to adequately represent those found in humans. The vocal fold length L was similar to the upper limit for humans (Zhang *et al.*, 2006). The value of the elastic modulus, assuming incompressibility, was comparable to the lower bound on longitudinal elastic properties of the human VF cover (Zhang *et al.*, 2006; Goodyer *et al.*, 2006). Considering that the longitudinal elastic properties of the vocal fold tissue are considerably higher than that for transverse loading the material employed provides a reasonable homogenized approximation of actual vocal fold tissue stiffness. The frequency of oscillation at onset, f=88.9 Hz, was at the lower bound of the range of physiological values (Titze, 2000) due to the large vocal fold length. The slow increase of frequency with flow rate beyond onset was consistent with observations for human larynges (Titze, 1989). The characteristics of the physical model during deformation indicated a single orifice was present during the open state, while two smaller, distinct orifices were created around one-quarter span in the closed state. The overemphasis of the third longitudinal mode, along with the absence of a mucosal

wave on the superior and medial surfaces, is not very realistic. These and other similar trends observed in 3D finite element simulations (de Oliveira Rosa *et al.*, 2003) are common to many continuum synthetic vocal folds models. It is important to note that the results obtained with such models are only indicative of general trends, and should not be considered a substitute for data on live tissue. Hopefully, progress in synthetic models along with human subject studies will provide more definitive data in the near future.

The structural differences between the physical model and the human larynx make it difficult to directly relate the results of the present study to actual human phonation. The current model assumed an unstressed initial state, and thus did not account for posturing or pressed phonation. Recognizing these limitations, it is nevertheless interesting to relate the findings of the present study to previous observations made using live tissue models. The estimated influence depth of collision (for flow rates around $Q=690\ \mathrm{cm}^3/\mathrm{s}$) was similar to reported depths of tissue damage (Dikkers and Nikkels, 1995), as well as estimates from computer models (Gunter, 2003). The collision pressures measured in the present study are similar to those reported in hemilarynx experiments. For example Jiang and Titze (1994) found that for a subglottal pressure of around 1 kPa, the peak impact pressure ranged between 0.5 and 1.5 kPa. This compares with an estimated contact pressures around 0.7 kPa for a pressure of 1.2 kPa for the present model. Measurements using pressure sensors inserted between excised canine VFs (Jiang *et al.*, 2001) have yielded wide ranges of contact pressure; for subglottal pressures on the order of 0.9–1.5 kPa, contact pressure values ranged from 0.8–1.5 kPa. These values are slightly higher than those obtained in the present study. FEM models disregarding the flow pressure (Gunter, 2003) have reported compressive stresses orthogonal to the medial surface on the superior surface. Flow studies indicate that fluid pressure acts mainly on the inferior surface. The medial surface is a free surface before collision onset, the stresses orthogonal to that surface are small during the open phase. Upon collision, the contacting medial surfaces are subjected to a compressive load. This contrasts with the tensile stresses measured on the superior surface in the present study. This may be due to the neglect of the fluid pressure loading on the inferior surface in the FEM study or the possibility that both tension and compression exist simultaneously on the superior and medial surfaces during collision, respectively.

This study has demonstrated the potential of digital image correlation analysis for measuring the mechanical state of the superior surface of the vocal folds. Clinical applications are possible, provided methods to safely create a speckle pattern on the superior VF surface are developed. Possible limitations include the need to use only a two-dimensional DIC process, image distortion from endoscopes, glare, and uneven light distributions. The inability to obtain data at discontinuities such as near the glottal gap is also a serious limitation of DIC. Small speckle sizes are needed to partially overcome this problem by allowing for small subset sizes. The potential to assess mechanical states throughout the tissue must be assessed. One possible approach would be the enforcement of displacement data as time varying boundary conditions in finite element models. Optimization techniques may also be used for the indirect determination of material properties in situations where the strain field over the inferior, medial, and superior surfaces could be visible, for example in excised hemi-larynx studies.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC005788 from the National Institute for Deafness and other Communication Disorders (NIH-NIDCD). Thanks are expressed to Li-Jen Chen for his help with the construction of the model, and to Sravan Mantha for initiating this investigation. The excellent suggestions and comments of two anonymous reviewers are also gratefully acknowledged.

APPENDIX

Reduction of the deviatoric stress vector into a system of linear equations:

$$\underline{\sigma}^d = 2G\left(I_0 - \frac{1}{3}\underline{m}\underline{m}^T\right)\underline{\varepsilon}, \tag{A1}$$

$$\underline{\sigma}^d = 2G\left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \times [1\ 1\ 1\ 0\ 0\ 0]\right)\begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \gamma_{xy} \\ 0 \\ 0 \end{bmatrix}, \tag{A2}$$

$$\underline{\sigma}^d = 2G\left(\begin{bmatrix} 2/3 & -1/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & 2/3 & -1/3 & 0 & 0 & 0 \\ -1/3 & -1/3 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \end{bmatrix}\right) \times \begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \gamma_{xy} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma^d_{xx} \\ \sigma^d_{yy} \\ \sigma^d_{zz} \\ \tau_{xy} \\ 0 \\ 0 \end{bmatrix}, \tag{A3}$$

$$\frac{\sigma_{xx}^d}{G} = 2[(2/3)\varepsilon_{xx} - (1/3)\varepsilon_{yy} - (1/3)\varepsilon_{zz}], \tag{A4}$$

$$\frac{\sigma_{yy}^d}{G} = 2[-(1/3)\varepsilon_{xx} + (2/3)\varepsilon_{yy} - (1/3)\varepsilon_{zz}], \tag{A5}$$

$$\frac{\sigma_{zz}^d}{G} = 2[-(1/3)\varepsilon_{xx} - (1/3)\varepsilon_{yy} + (2/3)\varepsilon_{zz}]. \tag{A6}$$

Reducing the overall stress vector:

$$\underline{\sigma} = \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \tau_{xy} \\ 0 \\ 0 \end{bmatrix} = \underline{\sigma}^d + \sigma^h \underline{\mathrm{m}} \approx \underline{\sigma}^d - \sigma_{zz}^d \underline{\mathrm{m}} = \begin{bmatrix} \sigma_{xx}^d - \sigma_{zz}^d \\ \sigma_{yy}^d - \sigma_{zz}^d \\ 0 \\ \tau_{xy} \\ 0 \\ 0 \end{bmatrix}. \tag{A7}$$

By incompressibility:

$$\varepsilon_{zz} = -\varepsilon_{xx} - \varepsilon_{yy}. \tag{A8}$$

Inserting Eq. (A8) into Eqs. (A4)–(A6),

$$\frac{\sigma_{xx}^d}{G} = 2[(2/3)\varepsilon_{xx} - (1/3)\varepsilon_{yy} - (1/3)(-\varepsilon_{xx} - \varepsilon_{yy})] = 2\varepsilon_{xx}, \tag{A9}$$

$$\frac{\sigma_{yy}^d}{G} = 2[-(1/3)\varepsilon_{xx} + (2/3)\varepsilon_{yy} - (1/3)(-\varepsilon_{xx} - \varepsilon_{yy})] = 2\varepsilon_{yy}, \tag{A10}$$

$$\frac{\sigma_{zz}^d}{G} = 2[-(1/3)\varepsilon_{xx} - (1/3)\varepsilon_{yy} + (2/3)(-\varepsilon_{xx} - \varepsilon_{yy})] = -2\varepsilon_{xx} - 2\varepsilon_{yy} \tag{A11}$$

Inserting Eqs. (A9)–(A11) into Eq. (A7) gives the final stress state:

$$\frac{\sigma_{xx}}{G} \approx \frac{(\sigma_{xx}^d - \sigma_{zz}^d)}{G} = 2\varepsilon_{xx} - (-2\varepsilon_{xx} - 2\varepsilon_{yy}) = 4\varepsilon_{xx} + 2\varepsilon_{yy}, \tag{A12}$$

$$\frac{\sigma_{yy}}{G} \approx \frac{(\sigma_{yy}^d - \sigma_{zz}^d)}{G} = 2\varepsilon_{yy} - (-2\varepsilon_{xx} - 2\varepsilon_{yy}) = 4\varepsilon_{yy} + 2\varepsilon_{xx}. \tag{A13}$$

Berry, D. A., Montequin, D. W., and Tayama, N. (**2001**). "High-speed digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547.

Berry, D. A., Zhang, Z., and Neubauer, J. (**2006**). "Mechanisms of irregular vibration in a physical model of the vocal folds," J. Acoust. Soc. Am. **120**, EL36–EL42.

Chu, C. T., Ranson, W. F., Sutton, M. A., and Peters, W. H. (**1985**). "Applications of digital-image-correlation techniques to experimental mechanics," Exp. Mech. **25**, 232–244.

Dikkers, F. G., and Nikkels, P. G. J. (**1995**). "Benign lesions of the vocal folds: Histopathology and phonotrauma," Manage. Int. Rev. **104**, 698–703.

de Oliveira, Rosa M., Pereira, J. C., Grellet, M., and Alwan, A. (**2003**). "A contribution to simulating a three-dimensional larynx model using the finite element method," J. Acoust. Soc. Am. **114**, 2893–2905.

Flanagan, J., and Landgraf, L. (**1968**). "Self-oscillating source for vocal-tract synthesizers," IEEE Trans. Audio Electroacoust. **16**, 57–64.

Fulcher, L. P., Scherer, R. C., Zhai, G., and Zhu, Z. (**2006**). "Analytic representation of volume flow as a function of geometry and pressure in a static physical model of the glottis," J. Voice **20**, 489–512.

Gardner, G. M., Castracane, J., Conerty, M., and Parnes, S. M. (**1995**). "Electronic speckle pattern interferometry of the vibrating larynx," Ann. Otol. Rhinol. Laryngol. **104**, 5–12.

Goodyer, E., Muller, F., Bramer, B., Chauhan, D., and Hess, M. (**2006**). "In vivo measurement of the elastic properties of the human vocal fold," Eur. Arch. Otorhinolaryngol. **263**, 455–462.

Greenwood, J. A. (**2006**). "A simplified elliptic model of rough surface contact," Wear **261**, 191–200.

Gunter, H. E. (**2003**). "Mechanical stress in vocal fold tissue during voice production," Ph.D. thesis, Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA.

Helm, J. D., McNeill, S. R., and Sutton, M. A. (**1996**). "Improved three-dimensional image correlation for surface displacement measurement," Opt. Eng. (Bellingham) **35**, 1911–1920.

Horáček, J., Šidlof, P., and Švec, J. G. (**2005**). "Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces," J. Fluids Struct. **20**, 853–869.

Hung, P., and Voloshin, A. S. (**2003**). "In-plane strain measurement by digital image correlation," J. Braz. Soc. Mech. Sci. **25**, 215–221.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal folds," Bell Syst. Tech. J. **51**, 1233–1268.

Jiang, J. J., Shah, A. G., Hess, M. M., Verdolini, K., Banzali, F. M., Jr., and Hanson, D. G. (**2001**). "Vocal fold impact stress analysis," J. Voice **15**, 4–14.

Jiang, J. J., and Titze, I. R. (**1994**). "Measurement of vocal fold intraglottal pressure and impact stress," J. Voice **8**, 132–144.

Johnson, K. L. (**2003**). *Contact Mechanics* (Cambridge University Press. Cambridge, UK), Chap. 11.

Kerdok, A. E., Cotin, S. M., Ottensmeyer, M. P., Galea, A. M., Howe, R. D., and Dawson, S. L. (**2003**). "Truth cube: Establishing physical standards for real time soft tissue simulation," Med. Image Anal. **7**, 283–291.

Manneberg, G., Hertegard, S., and Liljencrantz, J. (**2001**). "Measurement of human vocal fold vibrations with laser triangulation," Opt. Eng. (Bellingham) **40**, 2041–2044.

Mantha, S., Mongeau, L., and Siegmund, T. (**2005a**). "Dynamic digital image correlation of a dynamic physical model of the vocal folds," Adv. Bioeng. **57**, 77–78.

Mantha, S., Siegmund, T., and Mongeau, L. (**2005**). "Estimation of strain fields in self oscillating physical glottis models using 3D digital image correlation," *Proceedings of the Fourth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Florence, Italy, 29–31 October.

Mongeau, L., Franchek, N., Coker, C., and Kubli, R. (**1997**). "Characteristics of a pulsating jet through a small modulated orifice, with application to voice production," J. Acoust. Soc. Am. **102**, 1121–1133.

Ouaknine, M., Garrel, R., and Giovanni, A. (**2003**). "Separate detection of vocal fold vibration by optoreflectometry: A study of biphonation on excised porcine larynges," Folia Phoniatr Logop **55**, 28–38.

Ouaknine, M., Garrel, R., and Giovanni, A. (**2003**). "Separate detection of vocal fold vibration by optoreflectometry: A study of biphonation on excised porcine larynges," Folia Phoniatr Logop **55**, 28–38.

Park, J. B., and Mongeau, L. (**2007**). "Instantaneous orifice discharge coefficient of a physical, driven model of the human larynx," J. Acoust. Soc. Am. **121**, 442–455.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angel of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Schreier, H. W., Garcia, D., and Sutton, M. A. (**2004**). "Advances in light microscope stereo vision," Exp. Mech. **44**, 278–288.

Siebert, T., Becker, T., Spiltthof, K., Neumann, I., and Krupka, R. (**2007**). "High-speed digital image correlation: Error estimations and applications," Opt. Eng. **46**, 051004-1–7.

Spencer, M. (**2007**). "Indirect determination of the strain and stress in physi-

cal models of the vocal folds using digital image correlation," Masters thesis, Purdue University West Lafayette, IN.
Stronge, W. L. (**2000**). *Impact Mechanics* (Cambridge University Press, Cambridge, UK), Chap. 6, pp. 116–127.
Sutton, M. A., Cheng, M., Peters, W. H., Chao, Y. J., and McNeil, S. R. (**1986**). "Application of an optimized digital image correlation method to planar deformation analysis," Image Vis. Comput. **4**, 143–150.
Sutton, M. A., Helm, J. D., and Boone, M. L. (**2001**). "Experimental study of crack growth in thin sheet 2024-T3 aluminum under tension-torsion loading," Int. J. Fract. **109**, 285–301.
Švec, J. G., Horáček, J., Sram, F., and Vesely, J. (**2000**). "Resonance properties of the vocal folds: in vivo laryngoscopic investigation of the externally excited laryngeal vibrations," J. Acoust. Soc. Am. **108**, 1397–1407.
Tao, C., Jiang, J. J., and Zhang, Y. (**2006**). "Simulation of vocal fold impact pressure with a self-oscillating finite-element model," J. Acoust. Soc. Am. **119**, 3987–3994.
Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.
Tigges, M., Wittenberg, T., Rosanowski, F., and Eysholdt, U. (**1996**). "High-speed imaging and image processing in voice disorders," Proc. SPIE **2927**, 209–216.
Titze, I. R. (**1989**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906.
Titze, I. R. (**1994**). "Mechanical stress in phonation," J. Voice **8**, 99–105.
Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Iowa City, IA).
Verdolini, K., Hess, M. M., Titze, I. R., Bierhals, W., and Gross, M. (**1999**) "Investigation of vocal fold impact stress in human subjects," J. Voice **13**, 184–202.
Zhang, K., Siegmund, T., and Chan, R. W. (**2006**). "A constitutive model of the human vocal fold cover for fundamental frequency regulation," J. Acoust. Soc. Am. **119**, 1050–1062.
Zienkiewicz, O. C., and Taylor, R. L. (**2000**). *The Finite Element Method* (Butterworth-Heinemann, Oxford, UK), Chap. 12.

Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo

Hugo Quené[a)]
Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands

(Received 11 October 2006; accepted 7 November 2007)

Speech tempo (articulation rate) varies both between and within speakers. The present study investigates several factors affecting tempo in a corpus of spoken Dutch, consisting of interviews with 160 high-school teachers. Speech tempo was observed for each phrase separately, and analyzed by means of multilevel modeling of the speaker's sex, age, country, and dialect region (between speakers) and length, sequential position of phrase, and autocorrelated tempo (within speakers). Results show that speech tempo in this corpus depends mainly on phrase length, due to anticipatory shortening, and on the speaker's country, with different speaking styles in The Netherlands (faster, less varied) and in Flanders (slower, more varied). Additional analyses showed that phrase length itself is shorter in The Netherlands than in Flanders, and decreases with speaker's age. Older speakers tend to vary their phrase length more (within speakers), perhaps due to their accumulated verbal proficiency. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2821762]

PACS number(s): 43.70.Fq, 43.70.Jt, 43.70.Bk, 43.71.Bp [BHS] Pages: 1104–1113

I. INTRODUCTION

Speech tempo may be defined as the rate at which phonetic events occur over time. It is often expressed in phonemes or syllables per second (e.g., Stetson, 1951), or inverted, as seconds per syllable [average syllable duration (ASD), e.g., Goldman-Eisler (1968), Crystal and House (1990)]. Humans do not produce speech at a constant tempo (e.g., Miller *et al.*, 1984b). Speech tempo is reportedly affected by various factors, some varying within speakers, e.g., length of phrase, and others varying between speakers, e.g., sex. The present study aims to investigate several of these factors affecting tempo, by means of a corpus of spontaneous Dutch produced by 160 speakers from The Netherlands and Flanders (Belgium). A secondary purpose is to demonstrate how multilevel models can be used to capture effects varying at multiple hierarchical levels: that of speakers, and of phrases within speakers.

Speech tempo is defined here as the articulation rate, excluding pause time. In the present analysis, speech tempo is computed by taking the duration of each interpausal phrase in the speech corpus, and dividing that duration by the number of canonical or intended syllables in that chunk (Koreman, 2006), yielding ASD for that phrase.

A previous study of speech tempo in Dutch (Verhoeven *et al.*, 2004), using the same Dutch corpus material as in the present study, found significant effects of four predictors varying between speakers, viz., speakers' age, sex, country, and region. Speakers' average tempo was reported to be faster for younger speakers (ages 21–40) than for older speakers (ages 45–59), and faster for men than for women (Verhoeven *et al.*, 2004). For speakers of Dutch from The Netherlands, four dialect regions were investigated (Verhoeven *et al.*, 2004). The Randstad region (Zuid-Holland) is considered the linguistic center of The Netherlands. The Mid-region (Utrecht, Gelderland) is a transition zone. The North (Groningen, Drenthe) and South (Dutch Limburg) regions have distinct regional dialects, although the "western" variety of Standard Dutch is widely used. For speakers of Dutch from Flanders (Vlaanderen), four dialect regions were also investigated. The Brabant region is considered the linguistic center of Flanders. East Flanders is regarded as a transition zone, whereas Belgian Limburg and West Flanders are regarded as more peripheral. Speakers' average tempo in the standard variety of the language was found to be related to their distance from the "linguistic center." For speakers from The Netherlands, average tempo was fastest in the Randstad region, intermediate in the Mid-region, and slowest in the North and South regions (Verhoeven *et al.*, 2004). For speakers from Flanders, regional differences were not significant. On average, speech tempo was considerably slower in Flanders than in The Netherlands.

This previous analysis of the Dutch corpus material, however, was limited to *between-speaker* variation in tempo. In practice, however, within-speaker changes in tempo are probably more relevant for speech communication. The present study therefore extends previous models of Dutch speech tempo, by including several *within-speaker* predictors in the model.

Longer phrases, containing more syllables, tend to be spoken at a faster rate, with shorter average syllable durations; this is known as "anticipatory shortening" (Nooteboom, 1972; Lindblom and Rapp, 1973; De Rooij, 1979; Nakatani *et al.*, 1981). A plausible model of speech tempo should therefore include phrase length as a predictor, to capture the considerable within-speaker variation (Miller *et al.*, 1984b). If speakers from different regions would differ in their average phrase length, for example, then such differences would yield artifactual differences in speech tempo,

[a)]Electronic mail: hugo.quene@let.uu.nl.

through the mechanism of anticipatory shortening. Hence, phrase length has to be accounted for when studying speech tempo in corpus materials.

The speech corpus under analysis consists of interviews with high-school teachers of Dutch language and literature. Each interview lasted about 15 min. During the interview, speakers may gradually speed up (e.g., due to arousal) or slow down (e.g., due to fatigue). Hence, the sequential position of a phrase constitutes a second within-speaker predictor of the speech tempo within that phrase.

There may be cross correlations in tempo among adjacent phrases within each interview, e.g., because speakers may tend to alternate slow and fast responses. In order to investigate such possible autocorrelations the average speech tempo of a previous phrase was also included as a predictor. This was done for lags of one to five phrases, yielding five autocorrelation predictors.

Unfortunately, the present corpus material does not allow investigations of two other within-speaker factors that may well be relevant, viz., emphasis and emotional involvement. Many textbooks in phonetics state that speakers vary their speaking rate, in anticipation of the time listeners will need to process their words. Hence, important or unpredictable portions are typically spoken at a relatively slower rate (Zwaardemaker and Eijkman, 1928; Nooteboom and Eefting, 1994), because "we speak…in order to be understood" (Jakobson and Waugh, 1979, p. 95). Both emphasis and emotional involvement are very difficult and costly to annotate, however, and such annotations are therefore not available for any large-scale corpus such as used in this study.

The primary goal of this study is to adequately model spontaneous speech tempo, and speakers' variation in tempo, using a large corpus of spontaneous speech. These variations are relevant in speech communication, because listeners use the input speech tempo as a scaling factor for other phonetic distinctions, such as the distinction between /ba/ vs /wa/ (e.g., Miller *et al.*, 1984a), voicing of consonants (e.g., Miller and Grosjean, 1981; Volaitis and Miller, 1992), and quantity of vowels (e.g., Traunmüller and Krull, 2003). If there is a lot of within-speaker tempo variation, then this might make a listener's task more difficult, because the listener has to readjust and rescale the phonetic distinctions more frequently. The secondary goal of this study is to illustrate how multilevel modeling may help us achieve the first goal, by simultaneously modeling both tempo and tempo *variations*, at multiple hierarchical levels.

II. METHOD

The Corpus of Spoken Dutch (Oostdijk, 2000) was used to investigate which factors contribute to (variation in) speaking rate. For this purpose, we concentrated on the subcorpus containing interviews with $N=160$ high-school teachers of Dutch in Flanders (the Dutch-speaking part of Belgium) and in The Netherlands. Interviewed speakers ("interviewees") were stratified by dialect region (four regions within The Netherlands, and four within Belgium), sex, and age group (below 40 versus over 45 years of age), with $n=5$ speakers in each cell. All speakers are assumed to speak a variety of Standard Dutch. The 80 interviews within each country were conducted by the same interviewer (Netherlands: male, age 26; Belgium: female, age 23). (Using a single interviewer across all interviews would have resulted in varying degrees of cross-cultural difference between conversation partners.) Similar topics were discussed across all interviews. Hence, language variety, conversation partner, and conversation topics were largely eliminated as confounding factors, and the speech samples were highly comparable among speakers. Nevertheless, any differences in results between The Netherlands and Belgium may also be ascribed to different interviewers (conversation partners) being used in the two countries. Note, however, that the sex difference among conversation partners is balanced in both countries, and that both interviewers have approximately the same age. Hence any differences observed in this corpus study are most likely due to cultural differences between the two countries, and effects of the two interviewers' properties are further ignored in this study. For more details about dialect regions, speaker selection, and recording procedure, see Van Hout *et al.* (1999), Verhoeven *et al.* (2004), and Adank *et al.* (2007).

For each interview, the orthographic transcripts were extracted from the annotations provided with the corpus, broken down by interpause chunks. The speaking time of each chunk or phrase was determined from the time marks in the transcript. The number of orthographic syllables in each phrase was determined by dictionary look-up of the orthographic words (with manual correction where necessary). Speech tempo is expressed here as ASD (Goldman-Eisler, 1968; Crystal and House, 1990). Interpause chunks or phrases, as determined by the original transcribers, constitute the smallest units of observation. Hence, the average syllable durations per phrase correspond to so-called articulation rates, from which pause time is excluded. Most of the short phrases (of one or two orthographic syllables) consisted of hesitation sounds, filled pauses, backchannel sounds, etc. These were excluded from the data set [after Crystal and House (1990)]. The proportion of excluded phrases per speaker ranged from 0.01 to 0.27 (median 0.08, interquartile range 0.08). The number of included phrases per speaker ranged from 139 to 629 (median 328, interquartile range 97), with corresponding speech periods ranging from 4 to 31 min of speech per speaker (median 13.6 min, interquartile range 4.7). In total, $N=52\,975$ phrases from 160 speakers were included in the analyses (38.5 h of speech).

Average syllable durations (per phrase) were modeled by means of multilevel analysis (Goldstein, 1995; Cnaan *et al.*, 1997; Kreft and De Leeuw, 1998; Snijders and Bosker, 1999; Pinheiro and Bates, 2000; Luke, 2004), with speakers and phrases within speakers as two nested random factors. This type of analysis has several important advantages over more conventional techniques such as repeated measures analysis of variance (ANOVA), or ordinary least-squares linear regression [see Quené and Van den Bergh (2004) for a longer review and tutorial]. First, it allows for multiple, nested random effects, such as speakers and phrases within speakers. Second, multilevel modeling does not require homogeneity of variance, or sphericity. Between-speaker variance s_u^2 and within-speaker variance s_e^2 are modeled explic-

itly, instead of assumed to be homogeneous. This property allows us to investigate whether these between-speaker and within-speaker variances are affected by any of the predictors under study. Third, multilevel modeling allows for incomplete designs, and for varying numbers of observations per cell. This property removes the necessity to aggregate multiple lower-level observation units (phrases) into a single value at the higher level (speakers), as was done by Verhoeven *et al.* (2004). The present multilevel analysis assumes that phrases are nested under speakers, i.e., that phrases are not repeated between or within speakers. Although a few phrases were repeated (e.g., *ja precies* "yes exactly"), the number of these repeated phrases in the corpus was sufficiently low (only 147 out of 52 975) to warrant this assumption.

Before multilevel modeling, the eight levels of the region factor were converted to eight binary (dummy) factors. Sex was also included as a binary dummy factor (0 female, 1 male). Age was not included as a discrete factor discriminating two speaker groups [like Verhoeven *et al.* (2004)], but as a linear predictor, centralized to the mean age of 43 before modeling (Snijders and Bosker, 1999; Kreft and De Leeuw, 1998; Luke, 2004). Phrase length was log-transformed and centralized to its mean log value; the sequential position of each phrase was also centralized.

In any type of statistical modeling, the aim is to obtain an optimal model that contains the lowest number of predictors but explains the highest amount of variance of the dependent variable. In multilevel modeling, this is complicated somewhat because the optimal model can be different for the fixed part, and for the random parts at each level. Therefore one may find different predictors in the various parts of the models. For each model reported in the following, the fixed part contains estimated regression coefficients (β), and the random parts contain estimated amounts of variance between speakers (σ_u^2) and between phrases within speakers (σ_e^2), with standard errors for the former estimates, and confidence intervals for the latter estimates. The optimal model was selected by means of comparisons among candidate models with and without relevant effects in their random parts, using likelihood ratio tests (Pinheiro and Bates, 2000). Of the multiple candidate models for a particular set of predictors, only the optimal one will be reported in the following, for the sake of clarity.

The resulting estimates can be used for testing hypotheses. For binary predictors, e.g., sex, and for continuous predictors, e.g., age, the H_0 states that the estimated coefficient equals zero. Such hypotheses may be tested by using the Wald criterion that the estimated coefficient is significant if the coefficient exceeds 1.96 times its standard error [at $\alpha=0.05$; Hox (1995), Snijders and Bosker (1999)]. Other factors are modeled with multiple binary dummy predictors, e.g., region consisting of one binary predictor for each of the four regions. The main effect of the region is investigated by means of a planned contrast among the estimated coefficients for these binary predictors, with H_0 stating that this contrast equals zero. The amount of variance associated with such a contrast is then tested by means of a χ^2 distribution (Winer, 1971; Goldstein, 1995). Random estimates are evaluated by means of Markov chain Monte Carlo sampling; this is a Bayesian technique yielding estimates of the posterior distributions of all parameters, from which the 95% confidence of the random estimate may be extracted (Browne, 2005).

III. RESULTS

A. Speech tempo

The first model fitted was deliberately similar to the between-speaker ANOVA reported by Verhoeven *et al.* (2004) for the same corpus. This preliminary model given in Eq. (1) included only the between-speaker factors country, region, sex, and age (centralized) as fixed predictors. Random effects (in parentheses) were modeled at two hierarchical levels, viz., that of speakers (u_{0j}) and that of phrases within speakers (e_{ij}). In other words, articulation rates were not aggregated at the speaker level in the present study,

$$\begin{aligned} y_{ij} = {} & \mathrm{reg.NR}[\gamma_{\mathrm{NR}\ 00}] + \mathrm{reg.NM}[\gamma_{\mathrm{NM}\ 00}] \\ & + \mathrm{reg.NN}[\gamma_{\mathrm{NN}\ 00}] + \mathrm{reg.NS}[\gamma_{\mathrm{NS}\ 00}] \\ & + \mathrm{reg.FB}[\gamma_{\mathrm{FB}\ 00}] + \mathrm{reg.FE}[\gamma_{\mathrm{FE}\ 00}] \\ & + \mathrm{reg.FL}[\gamma_{\mathrm{FL}\ 00}] + \mathrm{reg.FW}[\gamma_{\mathrm{FW}\ 00}] \\ & + \mathrm{sex.Male}[\gamma_{\mathrm{sex}\ 00}] + \mathrm{age}[\gamma_{\mathrm{age}\ 00}] + (u_{0j} + e_{ij}). \end{aligned} \quad (1)$$

None of the within-speaker predictors were included, and random variances were assumed to be homogeneous, i.e., the predictors were not included in the random part of the model. The results of this preliminary model are listed in the left-hand part of Table I.

Results for this preliminary model (1) confirm previous analyses of speakers' average tempo in this corpus (Verhoeven *et al.*, 2004). First, comparisons of the regional means show that speakers from Flanders produced significantly longer syllables (i.e., slower tempo) than did those from The Netherlands [$\chi^2(1)=201.3$, $p<0.001$]. Second, within The Netherlands, speakers from the Randstad region (the linguistic center of The Netherlands) produced significantly shorter syllables (i.e., faster tempo) than did those from the other regions [$\chi^2(3)=12.2$, $p=0.007$]. Within Flanders, the corresponding contrasts between Brabant and the other regions are marginally significant [$\chi^2(3)=7.5$, $p=0.058$]. Third, male speakers produced significantly shorter syllables (i.e., faster tempo) than female speakers ($Z=-3.70$, $p<0.001$). Fourth, older speakers produced significantly longer syllables (slower tempo) than younger speakers ($Z=4.24$, $p<0.001$). For each additional year of age, ASD increases by 0.75 ms. With a grand mean ASD of 236 ms, the tempo difference between speakers aged 25 and 65 is $(40\times0.75)/236$ or about 13%. This age effect is well above the just noticeable difference (JND) for speech tempo of about 5% (Quené, 2007). The between-speaker variance is comparable to that reported in previous studies. Speakers' average syllable duration ranged from 177 to 333 ms (mean 238 ms, s.d. 36 ms). Tsao *et al.* (2006) compared the 15 slowest and the 15 fastestspeakers, in a sample of 100 speakers, and reported vowel durations of the slowest speakers that were

TABLE I. Estimated parameters of multilevel modeling of syllable durations (in ms). Estimates of fixed parameters are given with standard error (in parentheses); estimates of random parameters are given with 95% confidence intervals obtained from Markov chain Monte Carlo sampling (in parentheses).

	Model (1)		Model (2)	
Fixed				
reg.N.Randstad	204.9	(5.5)	193.1	(3.5)
reg.N.Mid	224.1	(5.5)	201.1	(3.6)
reg.N.North	225.2	(5.5)	193.5	(3.5)
reg.N.South	227.1	(5.5)	200.3	(3.5)
reg.F.Brab	279.8	(5.6)	262.9	(3.6)
reg.F.East	261.9	(5.5)	245.7	(3.6)
reg.F.Limb	277.7	(5.5)	263.2	(3.6)
reg.F.West	270.1	(5.5)	262.3	(3.6)
sex.Male	−13.5	(3.7)	−9.6	(2.1)
age[a]	0.75	(0.18)	0.27	(0.10)
length[b]			−88.6	(2.0)
position[c]			−0.004	(0.001)
lag.1			0.025	(0.002)
lag.2			0.008	(0.002)
lag.3			0.006	(0.002)
lag.4			0.006	(0.002)
lag.5			0.005	(0.002)
Random				
$\sigma^2_{u_{0j}}$ [d]	506.1	(388.1,636.9)	529.9	(421.8,668.3)
$\sigma^2_{u_{\text{length }0j}}$			595.1	(472.1,744.5)
$\sigma_{u_{0j}}\sigma_{u_{\text{length }0j}}$			−469.9	(−602.8,−364.1)
$\sigma^2_{e_{ij}}$ [e]	7564.6	(7473.1,7656.1)		
$\sigma^2_{e_{\text{N}ij}}$			1396	(1365.0,1429.3)
$\sigma^2_{e_{\text{F}ij}}$			3440	(3370.4,3511.7)
$\sigma^2_{e_{\text{length }ij}}$			3424	(3312.3,3524.5)
$\sigma_{e_{\text{N}ij}}\sigma_{e_{\text{length }ij}}$			−1917	(1963.5,1870.5)
$\sigma_{e_{\text{F}ij}}\sigma_{e_{\text{length }ij}}$			−3044	(3113.0,2967.4)
Deviance	623 952		546 270	

[a]Speaker's age in years, centralized, in ms/year.
[b]Length of phrase in syllables, log-transformed and centralized, in ms/log(syllables).
[c]Sequential number of phrase within interview, centralized, in ms/number.
[d]$\sigma^2_{u_{0j}}$ denotes the variance between the j higher-level units (speakers).
[e]$\sigma^2_{e_{ij}}$ denotes the variance between the i lower-level units (phrases) within the j higher-level units (speakers).

about 1.3 times as long as those of the fastest speakers. For the present sample of 160 speakers, the comparable slowest 24 speakers yield a mean ASD of 186 ms, for the fastest 24 this is 297 ms. The present ratio of about 1.6 between slowest and fastest speakers' ASD is somewhat larger than the ratio 1.3 reported for vowel durations. (Comparing only speakers within each country yields ratios of 1.4 for both The Netherlands and Flanders). Finally, the preliminary model (1) confirms that within-speaker variance is indeed far larger than between-speaker variance in speech tempo (Miller *et al.*, 1984b), here by more than one order of magnitude. The subsequent model was developed to investigate this within-speaker variance further.

The preliminary model was extended by including several within-speaker predictors: *phrase length* (in syllables, converted to log units, and centralized to the mean log length), *sequential position* of each phrase within its interview (also centralized), and autocorrelation predictors with lags 1–5. (No multicollinearity was observed among the latter five predictors, with $r<0.2$ for all pairs of predictors, and showing low eigenvalues for their principal components.)

Contrary to ANOVA models, multilevel models do not require the assumption that random variances (between speakers and within speakers) are homogeneous. Instead, such variances are modeled explicitly, which allows us to investigate the effects of the predictors on these variance components (Luke 2004; Snijders and Bosker, 1999; Pinheiro and Bates, 2000; Quené and Van den Bergh, 2004). The optimal model, specified in Eq. (2), contains effects of phrase length in its random parts (in parentheses) at both levels, and of country at the level of phrases within speakers. Speakers (higher-level units) vary in their average speech tempo. In addition, speakers may vary in their individual slopes of the phrase length effect. Between-speaker variance components are pooled over the speaker groups. Phrases within speakers (lower-level units) vary in their average speech tempo, and this variance within speakers may be different (nonhomogeneous) for speakers from The Netherlands and from Flanders. In addition, phrases within speakers vary in their

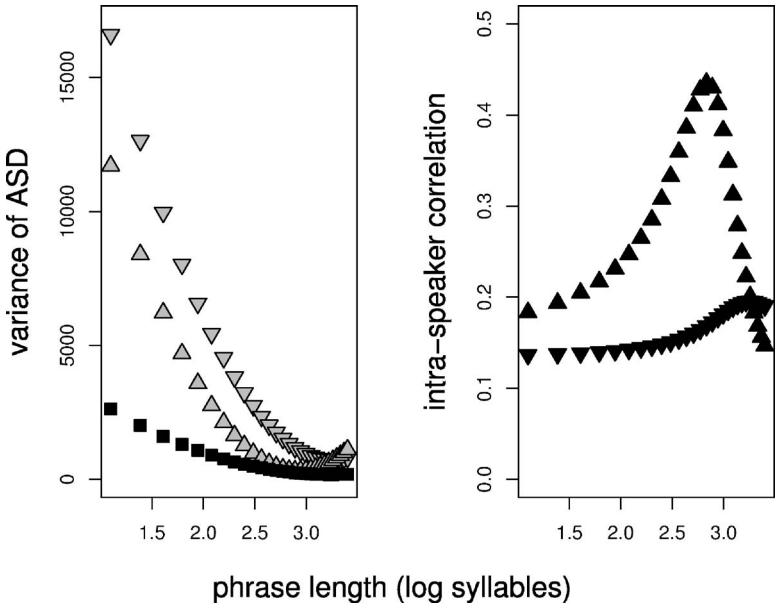


FIG. 1. Left panel: Variance estimates broken down by phrase length (log syllables), for between-speaker (dark squares) and within-speaker (light triangles) variances in ASD. Right panel: Intraspeaker correlations $[\sigma_u^2/(\sigma_u^2+\sigma_e^2)]$ broken down by phrase length (log syllables). Both panels show separate patterns for speakers from The Netherlands (triangles) and Flanders (inverted triangles).

slopes of the phrase length effect, and again this variance in slope may be different for the two speaker groups.

$$
\begin{aligned}
y_{ij} = {} & \text{reg.NR}[\gamma_{\text{NR }00}] + \text{reg.NM}[\gamma_{\text{NM }00}] \\
& + \text{reg.NN}[\gamma_{\text{NN }00}] + \text{reg.NS}[\gamma_{\text{NS }00}] \\
& + \text{reg.FB}[\gamma_{\text{FB }00}] + \text{reg.FE}[\gamma_{\text{FE }00}] \\
& + \text{reg.FL}[\gamma_{\text{FL }00}] + \text{reg.FW}[\gamma_{\text{FW }00}] \\
& + \text{sex.Male}[\gamma_{\text{sex }00}] + \text{age}[\gamma_{\text{age }00}] \\
& + \text{length}[\gamma_{\text{length }00} + (u_{\text{length }0j}) + (e_{\text{length }ij})] \\
& + \text{position}[\gamma_{\text{position }00}] + \text{lag.1}[\gamma_{\text{lag.1 }00}] \\
& + \text{lag.2}[\gamma_{\text{lag.2 }00}] + \text{lag.3}[\gamma_{\text{lag.3 }00}] \\
& + \text{lag.4}[\gamma_{\text{lag.4 }00}] + \text{lag.5}[\gamma_{\text{lag.5 }00}] \\
& + (u_{0j} + \text{N}[e_{\text{N }ij}] + \text{F}[e_{\text{F }ij}]). \qquad (2)
\end{aligned}
$$

Results for this model are listed in the right-hand part of Table I.

First, results for the *fixed* part confirm that phrase length indeed has a large and highly significant effect on speaking rate ($Z=44.3$, $p<0.001$), as known from previous research [Nooteboom (1972); Lindblom and Rapp (1973); De Rooij (1979); Nakatani *et al.* (1981)]. Speakers produce longer phrases with shorter average syllable duration, hence with faster speech tempo. In addition, sequential position has a small effect on speaking rate. Speakers tend to speed up by a very small amount (yielding somewhat shorter syllable durations) during the interview.

Second, one would expect that the effects of the between-speaker predictors (sex, age, country, and region) will be reduced in magnitude in model (2), where within-speaker variation is taken into account. The effect of sex ($Z=-4.7$, $p<0.001$) is indeed smaller than in the preliminary model, and it is below the 5% JND for speech tempo (Quené, 2007). The effect of age also decreases ($Z=2.71$, $p=0.003$). The reduction of the sex and age effects, and the significant phrase length effect, together suggest that the significant effects of sex and age observed in previous analyses may have been inflated by between-speaker effects on phrase length. This issue will be further discussed in the following. The tempo difference between speakers from The Netherlands (mean ASD 213 ms) and Flanders (mean ASD 263 ms) is again highly significant [$\chi^2(1)=886.1$, $p<0.001$]. Tempo was faster for speakers from the Randstad region (the linguistic center of The Netherlands) than for speakers from other regions of The Netherlands, although the difference is not significant [$\chi^2(3)=6.6$, $p=0.083$]. Regional differences within Flanders are significant in the current model (2) [$\chi^2(3)=26.0$, $p<0.001$], but speakers from the Brabant region (the linguistic center of Flanders) did not produce the fastest speaking rates.

Third, the positive autocorrelation coefficients suggest that speaking rate is weakly correlated among subsequent phrases. Phrases that are spoken faster (e.g., because of their pragmatic content) tend to be followed by somewhat faster phrases. These autocorrelation effects weaken even more as distance increases.

The *random* part of model (2) shows several interesting effects, regarding the effect of phrase length and the country effect (differences between The Netherlands and Flanders). These effects are illustrated in Fig. 1 for phrase lengths from 3 to 30 syllables, for which data are available from both countries (this includes 98% of phrases). First, within-speaker variances are smaller for speakers from The Netherlands (triangles pointing North) than for those from Flanders (triangles pointing South), as shown in the left-hand panel. As before, the within-speaker variance is considerably greater than the between-speaker variation in average or habitual speaking rate.

Second, phrase length affects the variances between speakers (dark symbols). As phrase length increases, individual speakers tend to converge to the same (fast) speaking rate, with decreasing variation between speakers. This decreasing variance presumably reflects universal phonetic constraints. In order to produce many syllables in one breath, speakers need to attain a fast tempo. For these long phrases, with associated fast tempo, speakers' average tempo is limited more by universal physiological and phonetic constraints (which reduce between-speaker variance) than by speakers' individual properties (which enlarge variance between speakers).

Even more striking than the decrement in between-speaker variance, however, is the similar decrement in within-speaker variance in ASD (Fig. 1, left-hand panel, triangles). As phrase length increases, phrases tend to converge to the speakers' average (fast) tempo for phrases of that particular length, with decreasing variation between phrases within speakers. The variance between phrases is presumably due to the semantic and pragmatic properties of each phrase (emphasis, attitude, etc.), as well as to its phonetic properties (phonetic complexity of the constituent syllables, phonological vowel length, etc.).

A large difference is observed between speakers from The Netherlands (triangles pointing North) and Flanders (triangles pointing South) in their within-speaker variances. Speakers from Flanders vary their tempo more (around the speaker's average) than speakers from The Netherlands do. This difference is probably not due to universal phonetic constraints, which presumably apply equally to speakers from The Netherlands and Flanders. Speakers apparently have some individual freedom in how strongly they vary tempo between phrases. The pattern in Fig. 1 suggests a cultural difference in within-speaker variations in tempo. The pattern also suggests that the (unknown) semantic-pragmatic effects are smaller for longer phrases.

In multilevel modeling, the intraspeaker correlation expresses the relative "uniqueness" of the higher-level units (speakers). This intraspeaker correlation is computed from the amounts of variances between and within speakers. The intraspeaker correlation is high if most variance is found between speakers (i.e., observations within an individual speaker are highly correlated) and only a small amount of variance is observed within speakers between phrases. The right-hand panel of Fig. 1 shows this intraspeaker correlation for the two speaker groups, as a function of phrase length.

For shorter phrases (up to about 20 syllables), the speakers from The Netherlands have a higher individual "uniqueness" than those from Flanders, mainly due to the lower within-speaker variances of the former, which makes their tempo more predictable. Speakers from The Netherlands adhere more to their own habitual average tempo than those from Flanders, yielding a higher intraspeaker correlation for the former. Stated otherwise, speakers from Flanders produce more expressive variation in tempo, which makes their tempo less predictable from the individual speaker's average, yielding a lower intraspeaker correlation.

For longer phrases, however, which are more challenging, the individual "uniqueness" of the speakers from The Netherlands decreases suddenly. All speakers from this group produce these phrases at the same tempo, with little variation between speakers, and with little (semantic-pragmatic) expressive variation within speakers. The remaining variance is presumably mainly due to universal phonetic constraints of the produced phrases. The pattern is different for speakers from Flanders, who still produce expressive variation even for these long and challenging phrases (as shown by their larger within-speaker variances). Speakers from Flanders also have to let go of their expressive variation, as phrase length increases, but they do so only for the longest phrases, of about 27 syllables or more. This ties in with their overall lower speech tempo. Because their average speech tempo is slower, speakers from Flanders still have some room for tempo variation even for very long phrases. Speakers from The Netherlands, by contrast, by speaking faster on average, are closer to their phonetic limitations, and have little room for variation in very long phrases.

In summary, these results suggest that speech tempo is slower, and has more expressive variation, for speakers from Flanders as compared to those from The Netherlands. Results also confirm that speech tempo increases with phrase length. Interestingly, average speech tempo is *not* affected by a speaker's age, if tempo measurements are corrected for phrase length, as in model (2). Without such corrections, however, age appeared to yield a significant main effect on speech tempo [Verhoeven *et al.* (2004) and preliminary model (1)]. The most likely explanation for this discrepancy between the models is that phrase length, as an intermediate variable controlling speech tempo, is itself affected by the speaker's age. For the sake of argument, let us assume that a speaker's mean phrase length is influenced by his or her age. In turn, phrase length affects speech tempo, so that speakers differing in (mean) phrase length would yield artifactual differences in speech tempo, if phrase length is not taken into account. If phrase length is included as a predictor for speech tempo, however, then the between-speaker differences in tempo would vanish, as was indeed observed in model (2).

This tentative explanation predicts that phrase length is indeed affected by the speaker's age (and perhaps also by other between-speaker predictors: sex, country, and region), in a way that is congruent with the observed differences in speech tempo. This was investigated by multilevel modeling of phrase length as the dependent variable.

B. Phrase length

Phrase length was also modeled by means of multilevel analysis, with speakers and phrases within speakers as two nested random factors. The optimal model, specified in Eq. (3), contains all remaining predictors in the fixed part, as

well as age in the random part at the phrases-within-speakers level.

$$
\begin{aligned}
y_{ij} = {} & \text{reg.NR}[\gamma_{\text{NR }00}] + \text{reg.NM}[\gamma_{\text{NM }00}] \\
& + \text{reg.NN}[\gamma_{\text{NN }00}] + \text{reg.NS}[\gamma_{\text{NS }00}] \\
& + \text{reg.FB}[\gamma_{\text{FB }00}] + \text{reg.FE}[\gamma_{\text{FE }00}] \\
& + \text{reg.FL}[\gamma_{\text{FL }00}] + \text{reg.FW}[\gamma_{\text{FW }00}] \\
& + \text{sex.Male}[\gamma_{\text{sex }00}] + \text{age}[\gamma_{\text{age }00} + (u_{\text{age }0j}) \\
& + (e_{\text{age }ij})] + \text{position}[\gamma_{\text{position }00}] + \text{lag.1}[\gamma_{\text{lag.1 }00}] \\
& + \text{lag.2}[\gamma_{\text{lag.2 }00}] + \text{lag.3}[\gamma_{\text{lag.3 }00}] \\
& + \text{lag.4}[\gamma_{\text{lag.4 }00}] + \text{lag.5}[\gamma_{\text{lag.5 }00}] \\
& + (u_{0j} + \text{N}[e_{\text{N }ij}] + \text{F}[e_{\text{F }ij}]). \qquad (3)
\end{aligned}
$$

Resulting estimates for this model are listed in Table II.

Results for the *fixed* part of this model suggests that speakers from The Netherlands produce shorter phrases than those from Flanders, yielding a significant main effect of country on phrase length [$\chi^2(1)=56.0$, $p<0.001$]. Second, the phrase length is similar for speakers from different regions within the Netherlands [$\chi^2(3)=3.10$, n.s.] and within Flanders [$\chi^2(3)=5.54$, n.s.], yielding a nonsignificant main effect of region within country [$\chi^2(6)=8.64$, n.s.]. Third, phrases from female speakers are equally long as those from male speakers, as indicated by the insignificant coefficient for the sex factor. Fourth, the main effect of age is significant ($Z=2.08$, $p=0.019$), which indicates a tendency for older speakers to produce shorter phrases than younger speakers do. The log of phrase length decreases by 0.0027 for each 1 year increment of age.

The *random* part of model (3) shows no effects of speaker's age on the between-speaker variance in phrase length. [If the random part is simplified to a single term σ_u^2 at the higher level, ignoring age as a between-speaker predictor, then that simpler model fits the data equally well. Model (3) is preferred here, however, because its parametrization corresponds more closely to the above mentioned model (2) of the tempo data, allowing comparable analyses.]

The within-speaker effects of age and of country (differences between The Netherlands and Flanders) are illustrated in Fig. 2. First, within-speaker variances are again smaller for speakers from the Netherlands (triangles pointing North) than for those from Flanders (triangles pointing South), as shown in the left-hand panel. Not surprisingly, within-speaker variance in phrase length is considerably greater than the between-speaker variation (squares).

Second, the effect of age on within-speaker variance is small, but highly significant. [If the within-speaker effect of age is removed from model (3), then the log-likelihood increases to 79509.3, $\chi^2(3)=19.5$, $p<0.001$]. Figure 2 illustrates this age effect. The left-hand panel shows the amount of between-speaker variance (squares) and within-speaker variance (triangles). Within-speaker variance in phrase length decreases with age for young adult speakers, and then gradually increases with age for older adult speakers. The turning point seems to lie around age 35 for Flemish speakers, and around age 45 for speakers from The Netherlands. For the speakers over 45, this age effect on within-speaker variance is stronger for those from Flanders than from The Netherlands, as indicated by the steeper slope of the former. The right-hand panel of Fig. 2 shows the amount of intra-speaker correlation, i.e., speaker's "uniqueness" or individuality, across the age range. As before, the lower within-speaker variation for the speakers from The Netherlands corresponds with higher intraspeaker correlations for this group.

TABLE II. Estimated parameters of multilevel modeling of phrase length (in syllables, log-transformed and centralized). Estimates of fixed parameters are given with standard error (in parentheses); estimates of random parameters are given with 95% confidence intervals obtained from Markov chain Monte Carlo sampling (in parentheses).

	Model (3)	
Fixed		
reg.N.Randstad	−0.012	(0.041)
reg.N.Mid	−0.049	(0.041)
reg.N.North	−0.0002	(0.041)
reg.N.South	−0.081	(0.041)
reg.F.Brab	0.195	(0.043)
reg.F.East	0.094	(0.042)
reg.F.Limb	0.153	(0.042)
reg.F.West	0.204	(0.043)
sex.Male	−0.017	(0.026)
age[a]	−0.0027	(0.0013)
position[b]	0.000	(0.000)
lag.1	0.000	(0.000)
lag.2	0.000	(0.000)
lag.3	0.000	(0.000)
lag.4	0.000	(0.000)
lag.5	0.000	(0.000)
Random		
$\sigma^2_{u_{0j}}$ [c]	0.027	(0.021,0.034)
$\sigma^2_{u_{\text{age }0j}}$	0.000	n/a
$\sigma_{u_{0j}}\sigma_{u_{\text{age }0j}}$ [d]	0.000	n/a
$\sigma^2_{e_{\text{N }ij}}$	0.209	(0.204,0.214)
$\sigma^2_{e_{\text{F }ij}}$	0.335	(0.328,0.343)
$\sigma^2_{e_{\text{age }ij}}$	0.00005	(0.00002,0.00008)
$\sigma_{e_{\text{N}ij}}\sigma_{e_{\text{age }ij}}$	−0.00008	(−0.00024,0.00009)
$\sigma_{e_{\text{F}ij}}\sigma_{e_{\text{age }ij}}$	0.00059	(0.00031,0.00091)
Deviance	79 489.8	

[a]Speaker's age in years, centralized, in ms/year.
[b]Sequential number of phrase within interview, centralized, in ms/number.
[c]$\sigma^2_{u_{0j}}$ denotes the variance between the j higher-level units (speakers).
[d]$\sigma^2_{e_{ij}}$ denotes the variance between the i lower-level units (phrase) within the j higher-level units (speakers).

IV. GENERAL DISCUSSION AND CONCLUSION

The first aim of this study was to model between-speaker and within-speaker effects on speech tempo, and on tempo variations. The results for the full model (2) in Table I show two robust effects in its fixed part. First, tempo differs considerably between The Netherlands (mean 213 ms) and Flanders (mean 263 ms). Second, tempo is strongly affected by the length of the phrase, with anticipatory shortening as the connecting mechanism. There are also small effects of speaker's region, sex, and age. Tempo variation within

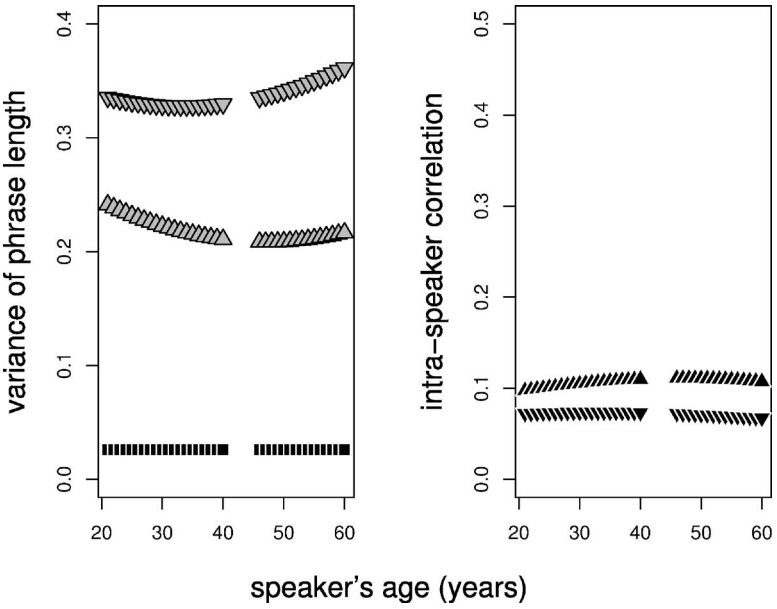


FIG. 2. Left panel: Variance estimates broken down by speaker's age (in years), for between-speaker (dark squares) and within-speaker (light triangles) variances in phrase length (which was log-transformed and centralized, see the text). Right panel: Intraspeaker correlations $[\sigma_u^2/(\sigma_u^2+\sigma_e^2)]$ broken down by speaker's age (in years). Both panels show separate patterns for speakers from The Netherlands (triangles) and Flanders (inverted triangles).

speakers is considerably larger than between speakers, but less so for speakers from The Netherlands than from Flanders. The faster tempo of speakers from The Netherlands leaves them less room for expressive variation, i.e., for within-speaker variations in tempo, presumably because these speakers are closer to their phonetic limitations. If they would speak slower, they might run out of breath before the end of the phrase; if they would speak faster, they might run into articulatory and perceptual difficulties.

Speech tempo in a phrase strongly depends of the length (in log syllables) of that phrase. This phenomenon is known as anticipatory shortening (Nooteboom, 1972; Lindblom and Rapp, 1973; De Rooij, 1979; Nakatani *et al.*, 1981): speakers shorten their syllables if they anticipate more syllables within a phrase. This finding also supports previous findings for American English read speech (Crystal and House, 1990). In the latter study, ASD was predicted in single-level fashion by seven phrase-internal predictors (e.g., proportion of stressed syllables, number of phones, etc.). This yielded an $R^2=0.579$. Since all six speakers read the same text, ASDs could be correlated between phrases spoken by pairs of speakers, yielding correlations between 0.66 and 0.88. ASD depends for a large part on the contents of a phrase [Crystal and House (1990), p.111], although it is also sensitive to other factors. In the present study, spontaneous speech instead of read speech was investigated. The spontaneous nature will probably lead to greater tempo variability within speakers, i.e., to a lower R^2 than observed for read speech (Crystal and House, 1990), because of the larger variation in pragmatic and semantic properties among phrases.

Previous analyses of speakers' aggregated tempi have suggested that dialect region, sex, and age are significant predictors of this between-speaker variation [Verhoeven *et al.* (2004) and preliminary model (1)]. The present study, however, nuances these findings. First, the differences among regions within each country may be significant, but these differences are not robust. The different rank orders of the regions' average tempi in models (1) and (2) and Verhoeven *et al.* (2004) constitute a warning against overinterpretation of these regional differences. Second, the sex difference may be significant, with males speaking somewhat faster than females, but this small difference is below the JND for speech tempo. Third, speech tempo is only weakly affected by speaker's age, if phrase length is also taken into account.

One possible explanation for these different findings in the present study is that the previously reported effects of age (and perhaps of sex, too) may have been indirect consequences of systematic variation in phrase length. If older speakers produce relatively shorter phrases than younger speakers, then this difference in phrase length would explain the observed age effect in speech tempo if phrase length is ignored, as well as the absence of an age effect in tempo if phrase length is not ignored. This predicted pattern was indeed observed for one between-speaker predictor, viz., speaker's age. In other words, between-speaker effects of age are mainly attributed to between-speaker differences in phrase length, with anticipatory shortening as the causal mechanism.

Phrase length turns out to vary with the speaker's age, in two ways. First, as predicted, older speakers produce shorter phrases than younger speakers do, as shown by the negative regression coefficient for the age predictor. Second, the amount of within-speaker *variation* in phrase length varies with the speaker's age (in years). Speakers over 45 tend to vary their phrase length more as the speaker's age increases. From age 45 to 60, within-speaker variance of (log-transformed) phrase length increases by about 4% (The Netherlands) to 8% (Flanders). Although modest in scale, the effect of age on within-speaker variance in phrase length is highly significant. The age grading effects may be explained by two mechanisms. Older speakers may successfully vary phrase length for communicative purposes, after decades of experience in expressing themselves as teachers. On the other hand, older teachers may increasingly suffer from cognitive constraints, both in retrieving words from their mental lexicon (e.g., Burke and Shafto, 2004) and in sentence construction (e.g., Kemper *et al.*, 2003). These cognitive constraints may have hampered the older speakers in this study

more than the younger speakers, forcing the former to produce shorter phrases occasionally, and yielding more variation in phrase length. More research is needed to further investigate such possible explanations.

Phrase length was also predicted to vary with speaker's sex, but this was not observed in the present corpus study. Hence the somewhat faster tempo observed for males cannot be explained phonetically by their longer phrases (through anticipatory shortening), and cannot be explained by physical sex differences (e.g., in mass of articulators), which tend to be in the opposite direction. The small but significant differences in tempo are therefore most likely coupled to gender differences in the speakers' social dominance and status. This tentative explanation is supported by the tendency for male interviewees to produce more syllables during the whole interview (average 4210 syllables, $s=1093$) than female interviewees do [average 3926 syllables, $s=922$; $t(158)=1.777$, $p=0.077$]. Male interviewees are more talkative than female ones in the present corpus. Similar gender differences in talking behavior have been reported for a large corpus of telephone conversations, where male speakers produced more words than female speakers [mean 926 words versus 867 words, respectively, in mixed-sex conversations; Liberman (2006)], as well as for formal meetings, where male participants talk longer, and interrupt more often, than female participants (Holmes, 1995). This gender difference is also reported in a recent meta-analytical review (Leaper and Ayres 2007; cf. Mehl *et al.*, 2007). In addition to producing more words and syllables, and interrupting more often, male speakers may also express their social dominance by speaking somewhat faster than female speakers.

Phrase length also differs between speakers from The Netherlands and from Flanders, but not in the direction one would predict from the observed tempo difference between the countries. Speakers from The Netherlands produce significantly *shorter* phrases, spoken at a significantly faster tempo, than speakers from Flanders do. At this international level, then, the relation between phrase length and tempo is reversed. In addition, within-speaker variation in phrase length is again smaller in The Netherlands than in Flanders. These findings point toward a cultural difference in speaking styles between The Netherlands (allegro: shorter, faster, less varied) and Flanders (andante: longer, slower, more varied). Now that speech corpora are becoming available for other languages with multiple linguistic centers (French, Portuguese, English, etc.), it would be interesting to further investigate these intercultural differences in speaking styles, and their consequences for intercultural speech perception.

The present study also aimed at illustrating how multilevel modeling may help us understand variation at multiple hierarchical levels simultaneously (Quené and Van den Bergh, 2004). It allows us to model not only average effects, but also variances around these effects. This allowed us to observe, for example, that speakers from The Netherlands produce less within-speaker variation in their speech tempo than speakers from Flanders, yielding higher intraspeaker correlations for the former (Fig. 1). It also allowed us to observe that a speaker's age affects not only his or her *average* phrase length, but also the within-speaker *variation* in phrase length (Fig. 2). At present, multilevel modeling is the best statistical technique available for drawing such multiple inferences simultaneously. Moreover, it requires fewer assumptions, and allows for missing data. These properties make multilevel modeling perfectly suited for understanding data from semispontaneous speech corpora as used in the present study.

In the final multilevel model of speech tempo (2), most within-speaker variance may be attributed to phrase length and other predictors. But there still remains some unexplained within-speaker variance (of 1396 and 3440 variance units, for The Netherlands and Flanders, respectively). This suggests that other unknown factors, *not* related to phrase length or any other predictor, also control a speaker's expressive variations in speech tempo. In particular, emphasis and emotional involvement are known to affect speech tempo, as discussed earlier. These semantic, pragmatic, and affective factors cannot be investigated with the present corpus material. Most of the within-speaker variance in tempo, however, can be ascribed to the length of a phrase, due to anticipatory shortening. From the perspective of a listener, who has to use speech tempo as a scaling factor during speech perception, this is good news: most variation in tempo is predictable from an easily observed property of the speech stimulus.

In conclusion, the present study has yielded several robust estimates of variations in speech tempo, based on a large corpus of spoken Dutch. Between-speaker variations in habitual speech tempo are mainly due to the speaker's country background. Most variations in speaking rate occur within-speakers, however, and they are mainly related to the length of a phrase, due to anticipatory shortening. Speakers from The Netherlands produce faster speech, and with less expressive variation, than speakers from Flanders. Interestingly, phrase length itself decreases with speaker's age; this may have inflated previously reported effects of speaker's age on speech tempo. In addition, older speakers tend to produce more individual variation in phrase length; this may be due to a decrement in cognitive performance, or to an increment in their expressive verbal variation.

ACKNOWLEDGMENTS

Preliminary versions of this work were presented in a Festschrift *On Speech and Language: Studies for Sieb G. Nooteboom* (Utrecht: Netherlands Graduate School of Linguistics LOT, 2004), and at the Ninth European Conference on Speech Communication and Technology (InterSpeech), Lisbon, Portugal, 4–8 September 2005. My sincere thanks are due to Hans Van de Velde for his assistance in the corpus analysis, to Griet Depoorter (TST-Centrale) for help with speaker identification codes, to Huub van den Bergh for statistical guidance, and to Esther Janse, Sieb Nooteboom, Willemijn Heeren, Matthias Mehl, and two anonymous reviewers for helpful comments and suggestions.

Adank, P., Van Hout, R., and Van de Velde, H. (**2007**). "An acoustic description of the vowels of northern and southern Standard Dutch. II. Regional varieties," J. Acoust. Soc. Am. **121**, 1130–1141.
Browne, W. (**2005**). *MCMC Estimation in MLwiN* (Centre of Multilevel Modeling, Bristol, UK.).
Burke, D. M., and Shafto, M. A. (**2004**). "Aging and language production," Curr. Dir. Psychol. Sci. **13**, 21–24.
Cnaan, A., Laird, N. M., and Slasor, P. (**1997**). "Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data," Stat. Med. **16**, 2349–2380.
Crystal, T. H., and House, A. S. (**1990**). "Articulation rate and the duration of syllables and stress groups in connected speech," J. Acoust. Soc. Am. **88**, 101–112.
De Rooij, J. (**1979**). "Speech punctuation: An acoustic and perceptual study of some aspects of speech prosody in Dutch," Ph.D. thesis, Rijksuniversiteit Utrecht, Utrecht, The Netherlands.
Goldman-Eisler, F. (**1968**). *Psycholinguistics: Experiments in Spontaneous Speech* (Academic, London).
Goldstein, H. (**1995**). *Multilevel Statistical Models*, 2nd ed. (Arnold, London).
Holmes, J. (**1995**). *Women, Men and Politeness* (Longman, London).
Hox, J. (**1995**). *Applied Multilevel Analysis* (TT Publicaties, Amsterdam).
Jakobson, R., and Waugh, L. R. (**1979**). *The Sound Shape of Language* (Harvester, Brighon).
Kemper, S., Herman, R., and Lian, C. (**2003**). "Age differences in sentence production," J. Gerontol. B Psychol. Sci. Soc. Sci. **58B**, 260–268.
Koreman, J. (**2006**). "Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech," J. Acoust. Soc. Am. **119**, 582–596.
Kreft, I. G., and De Leeuw, J. (**1998**). *Introducing Multilevel Modeling* (Sage, London).
Leaper, C., and Ayres, M. M. (**2007**). "A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliate speech, and assertive speech," Personality and Social Psychology Review **11**, 328–363.
Liberman, M. (**2006**). "Gabby guys: The effect size," Language Log Vol. **3607**, www.languagelog.org, last accessed 10 October 2006.
Lindblom, B., and Rapp, K. (**1973**). "Some temporal regularities of spoken Swedish," Papers in Linguistics from the University of Stockholm, Vol. **21**, pp. 1–59.
Luke, D. A. (**2004**). *Multilevel Modeling* (Sage, Thousand Oaks, CA).
Mehl, M. R., Vazire, S., Ranirez-Esparza, N., Slatcher, R. B., and Pennebaker, J. W. (**2007**). "Are women really more talkative than men?" Science **317**, 82.
Miller, J., Aibel, I., and Green, K. (**1984a**). "On the nature of rate-dependent processing during phonetic perception," Percept. Psychophys. **35**, 5–15.
Miller, J., and Grosjean, F. (**1981**). "How the components of speaking rate influence perception of phonetic segments," J. Exp. Psychol. Hum. Percept. Perform. **7**, 208–215.
Miller, J., Grosjean, F., and Lomanto, C. (**1984b**). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications," Phonetica **41**, 215–225.
Nakatani, L. H., O'Connor, K. D., and Aston, C. H. (**1981**). "Prosodic aspects of American English speech rhythm," Phonetica **38**, 84–105.
Nooteboom, S. (**1972**). "Production and perception of vowel duration: A study of durational properties of vowels in Dutch," Ph.D. thesis, Rijksuniversiteit Utrecht, Utrecht, The Netherlands.
Nooteboom, S., and Eefting, W. (**1994**). "Evidence for the adaptive nature of speech on the phrase level and below," Phonetica **51**, 92–98.
Oostdijk, N. (**2000**). "Het corpus Gesproken Nederlands," (The Spoken Dutch Corpus), Nederlandse Taalkunde **5**, 280–284.
Pinheiro, J. C., and Bates, D. M. (**2000**). *Mixed-Effects Models in S and S-Plus*, Statistics and Computing (Springer, New York).
Quené, H. (**2007**). "On the just noticeable difference for tempo in spontaneous speech," J. Phonetics **35**, 353–362.
Quené, H., and Van den Bergh, H. (**2004**). "On multi-level modeling of data from repeated measures designs: A tutorial," Speech Commun. **43**, 103–121.
Snijders, T., and Bosker, R. (**1999**). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Sage, London).
Stetson, R. (**1951**). *Motor Phonetics: A Study of Speech Movements in Action*, 2nd ed. (North Holland, Amsterdam).
Traunmüller, H., and Krull, D. (**2003**). "The effect of local speaking rate on the perception of quantity in Estonian," Phonetica **60**, 187–207.
Tsao, Y.-C., Weismer, G., and Iqbal, K. (**2006**). "The effect of intertalker speech rate variation on acoustic vowel space," J. Acoust. Soc. Am. **119**, 1074–1082.
Van Hout, R., De Schutter, G., De Crom, E., Huinck, W., Kloots, H., and Van de Velde, H. (**1999**). "De uitspraak van het Standaard-Nederlands: Variatie en varianten in Vlaanderen en Nederland" [The pronunciation of Standard Dutch: Variation and variants in Flanders and the Netherlands], in *Artikelen van de Derde Sociolinguïstische Conferentie*, edited by E. Huls and B. Weltens (Eburon, Delft, The Netherlands), pp. 183–196.
Verhoeven, J., De Pauw, G., and Kloots, H. (**2004**). "Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands," Lang Speech **47**, 297–308.
Volaitis, L., and Miller, J. (**1992**). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," J. Acoust. Soc. Am. **92**, 723–735.
Winer, B. (**1971**). *Statistical Principles in Experimental Design*, 2nd ed. (McGraw-Hill, New York).
Zwaardemaker, H., and Eijkman, L. (**1928**). *Leerboek der Phonetiek: Inzonderheid met betrekking tot het Standaard-Nederlandsch* [Textbook of Phonetics: Particularly in relation to Standard Dutch] (Bohn, Haarlem, The Netherlands).

Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners

Kazumi Maniwa[a] and Allard Jongman
Department of Linguistics, The University of Kansas, Lawrence, Kansas 66044

Travis Wade
Posit Science Corporation, 225 Bush St. 7th floor, San Francisco, California, 94104

(Received 17 April 2007; accepted 12 November 2007)

Speakers may adapt the phonetic details of their productions when they anticipate perceptual difficulty or comprehension failure on the part of a listener. Previous research suggests that a speaking style known as clear speech is more intelligible overall than casual, conversational speech for a variety of listener populations. However, it is unknown whether clear speech improves the intelligibility of fricative consonants specifically, or how its effects on fricative perception might differ depending on listener population. The primary goal of this study was to determine whether clear speech enhances fricative intelligibility for normal-hearing listeners and listeners with simulated impairment. Two experiments measured babble signal-to-noise ratio thresholds for fricative minimal pair distinctions for 14 normal-hearing listeners and 14 listeners with simulated sloping, recruiting impairment. Results indicated that clear speech helped both groups overall. However, for impaired listeners, reliable clear speech intelligibility advantages were not found for non-sibilant pairs. Correlation analyses comparing acoustic and perceptual data indicated that a shift of energy concentration toward higher frequency regions and greater source strength contributed to the clear speech effect for normal-hearing listeners. Correlations between acoustic and perceptual data were less consistent for listeners with simulated impairment, and suggested that lower-frequency information may play a role.
[DOI: 10.1121/1.2821966]

PACS number(s): 43.71.Es, 43.71.Ky, 43.71.Gv, 43.70.Mn [PEI] Pages: 1114–1125

I. INTRODUCTION

Fricative consonants, especially non-sibilants, present considerable identification difficulty for hearing-impaired listeners and for normal-hearing listeners under adverse conditions (Boothroyd, 1984; Dubno and Levitt, 1981; Dubno *et al.*, 1982; Miller and Nicely, 1955; Owens, 1978; Owens *et al.*, 1972; Sher and Owens, 1974; Singh and Black, 1966; Soli and Arabie, 1979; Wang and Bilger, 1973). This study was designed to measure whether, and how, speakers may be able to alleviate this difficulty by deliberately producing fricatives more clearly.

A. Clear speech intelligibility advantage

Many language users spontaneously adapt the phonetic details of their speech on-line in response to social and communicative demands, adopting an intelligibility-enhancing speaking style when they anticipate or sense perceptual difficulty or comprehension failure on the part of a listener (due to, e.g., background noise, reverberation, hearing impairment, lack of linguistic/world knowledge). "Clear speech" has been elicited in laboratory settings (e.g., Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Ferguson and Kewley-Port, 2002; Gagné *et al.*, 1994, 1995, 2002; Helfer, 1997, 1998; Iverson and Bradlow, 2002; Krause and Braida, 2002; Liu *et al.*, 2004; Payton *et al.*, 1994; Picheny *et al.*, 1985; Schum, 1996; Uchanski *et al.*, 1996), and shown to result in intelligibility advantages relative to "conversational" speech ranging from 7 to 38 percentage points. Clearly spoken sentences benefit young normal-hearing listeners in noise and/or reverberation (Bradlow and Bent, 2002; Gagné *et al.*, 1995; Krause and Braida, 2002; Payton *et al.*, 1994; Uchanski *et al.*, 1996) and with simulated hearing loss or cochlear implants (Iverson and Bradlow, 2002; Liu *et al.*, 2004), hearing-impaired listeners in quiet (Picheny *et al.*, 1985; Uchanski *et al.*, 1996) and in noise or reverberation (Payton *et al.*, 1994; Schum, 1996), cochlear-implant users (Iverson and Bradlow, 2002;Liu *et al.*, 2004), elderly listeners with or without hearing loss (Helfer, 1998; Schum, 1996), children with or without learning disabilities (Bradlow *et al.*, 2003) and (perhaps to a lesser extent) non-native listeners (Bradlow and Bent, 2002).

Recent results from Ferguson and Kewley-Port (2002) question the robustness of the "clear speech effect" and suggest that hyperarticulation strategies may interact in complicated ways with different types of signal degradation. While Ferguson and Kewley-Port found intelligibility benefits for clearly produced vowels with young, normal-hearing listeners, they observed *negative* clear-speech intelligibility benefits (better recognition of conversational tokens) with elderly hearing-impaired listeners, at least for one talker's productions. This pattern was mostly due to front vowels. A hallmark of clear speech is a greater concentration of energy

[a]Author to whom correspondence should be addressed. Electronic mail: kazumi.maniwa@uni-konstanz.de

in higher frequencies, in terms of both overall spectral distributions and individual formant frequencies (e.g., Krause and Braida, 2002; Picheny *et al.*, 1985). Since F2 values for front vowels fell in a frequency region where these listeners had sloping hearing loss (above 2000 Hz), clear vowels' higher F2 resonances, on average, fell in regions of greater impairment than those of conversational vowels.

It is of course unclear whether the patterns observed for this talker are unique to him or whether they are typical of the production, and perception by hearing-impaired or older listeners, of clear front vowels. The present study was designed to determine whether clear speech advantages occur for another class of sounds with a preponderance of high-frequency energy, fricatives, over a range of talkers and for young normal- hearing listeners and listeners with simulated high-frequency hearing loss.

B. Talker-related acoustic correlates of clear speech intelligibility

A secondary goal of this study was to determine which aspects of clear fricative production influence intelligibility. Previous investigations of the intelligibility of clear and conversational speech that have included more than a single talker have revealed considerable differences in the magnitude of the clear speech effect across talkers (e.g., Bradlow *et al.*, 2003; Chen, 1980; Ferguson, 2002, 2004; Gagné *et al.*, 1994, 1995; Schum, 1996). A few studies have attempted to identify talker-specific acoustic-phonetic parameters that may be responsible for the clear speech effect, by relating intelligibility differences to acoustic differences in clear and conversational speech. The talker in Bradlow *et al.* (2003) who showed the greater intelligibility advantage for clear speech substantially decreased her speaking rate with increased frequency and duration of pauses. Ferguson (2002) compared ten vowel measurements (five steady-state metrics, four dynamic metrics, and duration) across speakers and found that "big benefit" talkers showed the greatest increases in front vowel F2, Fl range, and the overall size of the vowel space.

The present study included an extended analysis of this type. Intelligibility was tested using a database of 8800 clear and conversational fricative productions by 20 talkers (10 M, 10 F), for which several spectral, temporal, and amplitudinal measurements have been reported (Maniwa *et al.*, submitted). These fricatives were all produced in vowel-consonant-vowel (VCV) (/a/-fricative-/a/) contexts. Based on features known to contribute to the perception of fricatives (described in the next section), the following measurements were made for these sounds: the frequency of the peak in the discrete Fourier transform (DFT); the mean, standard deviation, skewness, and kurtosis of the (DFT) spectral distribution; F2 onset transitions; the slope of the power spectrum below and above the (expected) peak location; the mean fundamental frequency (f0) of the adjacent vowels; root-mean-square (rms) frication amplitude relative to the surrounding vowels; *frequency-specific relative amplitude* (FSRA), i.e., the amplitude of the frication relative to the surrounding vowels in the F3 region for sibilants and the F5 region for non-sibilants; fricative harmonic-to-noise ratio (HNR), energy below 500 Hz, and duration. Very briefly, this analysis revealed several overall acoustic-phonetic modifications in the production of clear fricatives. Some of these effects were straightforwardly predictable based on previous findings (e.g. longer duration; energy at higher frequencies shown by higher peak, mean, and F2 frequencies; lower skewness indicating more positive spectral tilt; and steeper spectral slopes suggesting more defined peaks), and some were more surprising (especially lower relative amplitude). In most cases there were also Style × Fricative interactions indicating that these effects differed depending on the fricative; these effects were usually in a direction such that the acoustic distance between neighboring sounds was maximized in clear speech. For all measures, there was a wide range of variability across talkers in the extent to which a modification was implemented. In the present study, correlation analyses of acoustic and intelligibility measures across talkers were performed to assess the contributions of different acoustic modifications to intelligibility.

C. Cues to English fricative identity in different listener groups

Acoustic cues that have been reported to affect perception of English fricative place of articulation for listeners with normal hearing include frication duration, spectrum, and amplitude, as well as adjacent formant transitions and vowel quality. Experiments using natural (Harris, 1958; Zeng and Turner, 1990), synthetic (Heinz and Stevens, 1961; Zeng and Turner, 1990), and hybrid (Nittrouer, 1992, 2002; Nittrouer and Miller, 1997a and 1997b) speech suggest that spectral cues are important for distinguishing sibilants, while formant transition cues may help to distinguish nonsibilants (Harris, 1958; Heinz and Stevens, 1961; Nittrouer, 2002) and take on more weight when spectral cues are ambiguous (Hedrick, 1997; Hedrick and Carney, 1997; Hedrick and Ohde, 1993; Hedrick and Younger, 2003; Whalen, 1981). Overall noise duration and amplitude seem to have less perceptual significance (Behrens and Blumstein, 1988; Hedrick, 1997; Hedrick and Carney, 1997; Hedrick and Ohde, 1993; Hughes and Halle, 1956; Jongman, 1989; cf. Guerlekiean, 1981; McCasland, 1979a and 1979b), but manipulation of frication amplitude in particular frequency regions does influence listeners' perception of place of articulation for /s/-/ʃ/ and /s/-/θ/ contrasts (Hedrick, 1997; Hedrick and Carney, 1997;Hedrick and Ohde, 1993; Hedrick and Younger, 2003; Stevens, 1985). Fewer studies have investigated which acoustic components serve to distinguish voiced and voiceless fricatives. It appears that noise duration, the amplitude and duration of glottal vibration at the edge of the fricative, and the extent of Fl transitions interact in determining listener judgments of voicing for intervocalic fricatives (Stevens *et al.*, 1992).

Listeners do not seem to process fricative acoustic cues independently, but integrate information obtained from multiple dimensions; furthermore, the perceptual weights assigned to different acoustic properties depend on contexts and listeners. Adult listeners with normal hearing seem to make more use of spectral cues for place of articulation information (Heinz and Stevens, 1961; Harris, 1958; Hedrick and Ohde, 1993; Hughes and Halle, 1956; Nittrouer, 1992;

Nittrouer and Miller, l997a and 1997b; Nittrouer, 2002; Zeng and Turner, 1990), and temporal information for the voicing distinction (Cole and Cooper, 1975; Raphael, 1972; Soli, 1982). Hearing-impaired listeners may have difficulty integrating amplitude and spectral cues, and may generally place less weight on formant transitions than listeners with normal hearing (Hedrick, 1997; Hedrick and Younger, 2003; Zeng and Turner, 1990). In addition, listeners with sloping hearing loss commonly have elevated thresholds, and reduced dynamic range, in regions relevant to fricative perception (e.g., Dubno *et al.*, 1982; Owens *et al.*, 1972; Sher and Owens, 1974). It is likely, then, that clear speech alternations involving fricative spectra may have different results depending on the listener population. To address this possibility, this study examined the perception of clear and conversational fricatives by normal-hearing listeners (Experiment 1) and listeners with simulated hearing impairment (Experiment 2).

D. Hypotheses

Two experiments were performed to address three questions. First, are clearly produced fricatives more intelligible than conversational fricatives for listeners with normal hearing in degraded conditions? Based on previous findings, we hypothesized that they would be, although the effects might vary depending on fricatives (e.g. Ferguson and Kewley-Port, 2002). Second, what acoustic modifications are related to intelligibility? It was hypothesized that not all strategies employed by talkers serve to improve fricative identification, although it was difficult to predict which modifications would be most effective given previous conflicting results. Third, do clear-speech intelligibility differences differ based on listener population, in particular for listeners with sloping hearing losses? We expected that hearing loss might interact with clear-speech strategies, perhaps resulting in reduced benefit where high-frequency information was critical.

II. EXPERIMENT 1: EFFECT OF CLEAR SPEECH FOR FRICATIVE RECOGNITION BY LISTENERS WITH NORMAL HEARING

A. Method

1. Participants

Fourteen normal-hearing listeners (8 F, 6 M) aged 19–32 were recruited from the University of California, Berkeley. Participants were native speakers of American English, without noticeable regional dialects. Participants reported normal hearing and no history of speech or language disorders. Listeners were paid for their participation in the experiment.

2. Materials

As discussed in Sec. I B, intelligibility was assessed using a previously described corpus of VCV stimuli (Maniwa *et al.*, submitted). Briefly, conversational and clear tokens were elicited using an interactive program that ostensibly attempted to identify the sequence of fricatives produced by a speaker. The program made frequent, systematic errors involving voicing and place alternations, after which the speaker repeated a sound more clearly, as if trying to disambiguate the production for an elderly or hearing-impaired listener. All stimuli were normalized to the same long-term (word-level) rms amplitude and presented at 60 dB sound pressure level using MATLAB (The Math Works, Inc., 2000). Test stimuli were delivered in a background of 12-talker (6 F, 6 M) babble recorded at a sampling rate of 44.1 kHz. A total of 60 s of babble was created for the purposes of the experiment; for each stimulus, a segment of babble was selected from a random location within this 60 s sample. The duration of this segment exceeded that of the test item by a total of 600 ms, with the test stimulus centered temporally in the babble. There were 5 and 100 ms linear on-off ramps for the target stimulus and the noise, respectively.

3. Procedures and apparatus

The perception test employed a two-alternative forced-choice identification task. The eight fricatives were divided into eight minimal pairs, depending on place of articulation and voicing: /f/-/θ/, /v/-/ð/, /s/-/ʃ/, /z/-/ʒ/, /f/-/v/, /θ/-/ð/, /s/-/z/, and /ʃ/-/ʒ/. Each pair was tested separately for clear and conversational styles, for a total of 16 sub tests. Sub-test order was randomized across subjects in a single 1 h session. Subjects listened to stimuli presented via Koss headphones in sound-attenuated rooms, seated in front of a computer monitor and mouse. On each trial, test VCV and babble waveforms were scaled based on the selected signal-to-noise ratio (snr) (described below) and the constant target stimulus level, combined additively, and presented diotically to the subjects, who were prompted to identify the fricative from a minimal pair by using the mouse to click one of two letter combinations on the computer screen. Response alternatives were written: "ff," "th," "ss," "sh," "vv," "dh," "zz," and "zh." Listeners were first oriented to the spelling of response alternatives and the test procedure, and a ten-trial block of fricative tokens at a high snr (+10 dB) was run with feedback before each sub-test.

The goal of each sub test was to determine the snr threshold at which a distinction could be made with 75% accuracy. In each test, two 40-trial adaptive tracks were initiated at +3 dB and −3 dB snr and interleaved at random over the 80-trial block. Signal-to-noise ratio values for each track were selected using a Bayesian adaptive algorithm (ZEST; e.g., King-Smith *et al.*, 1994). The final threshold estimate was simply taken as the average (in dB) of the snr values for each track on the final (40th) trial. While this approach may have resulted in less precise measurements of thresholds that were further from the initial guesses (since termination was not based on confidence criteria) it was considered more important that participants were exposed to equal numbers of stimuli from each contrast pair; ±20 dB were chosen as absolute maximum and minimum allowable snr values. Individual test tokens were selected randomly from the productions of the 20 speakers, so that speakers and productions would, on average, be represented with equal frequency.

4. Data analysis

The clear speech intelligibility effect was tested using a repeated measures analysis of variance (ANOVA) with two within-subject factors (Style; two levels, Pair; eight levels) and threshold (dB snr) as the dependent variable. In order to assess the effect of pair type more thoroughly, another repeated measures ANOVA was calculated with three within-subject factors. One of the factors was Style. The second factor, Sibilance, depended on whether the pair consisted of sibilant fricatives or non-sibilant fricatives, and the third factor, Distinction, was labeled depending on whether the pair involved a place or voicing distinction. Pairwise comparisons for significant within-subject factors were done using Bonferroni corrected 95% confidence intervals.

In addition, as a first step in determining which acoustic modifications were related to intelligibility, correlation analyses were carried out across the 20 speakers included in the experiment, relating differences in their production strategies to differences in their clear-speech benefit. First, for each speaker, a single clear-minus-conversational difference value, averaged over all fricatives and productions, was calculated for each of the 14 acoustic measures reported in the Maniwa *et al.* (submitted) study: DFT peak location (1), the first four spectral moments moments (M1–M4; 2–5), F2 onset transitions (6), spectral slopes below (7) and above (8) typical peak locations (SlpBef, SlpAft), averaged f0 of adjacent vowels (9), normalized rms amplitude (rmsamp, 10), frequency-specific relative amplitude (FSRA) (11), HNR (12), energy below 500 Hz (13) and fricative duration (14). For (1)–(5), (7)–(8), (10), and (13), analyses considered 40 ms Hamming windowed segments at five locations: centered over the fricative onset, 25, 50, and 75% points, and offset (window (W) 1–5). For (6), acoustic values were derived at fricative onset and offset and each vowel midpoint from an analysis (Wl–4). For (9), f0 was averaged across the vowels preceding and following the target. For (11), (12) and (14), the values were obtained over the entire fricative. In the present analyses, 50-order linear predictive coding (LPC) peaks (at the same five locations) were included as well, and f0 was considered separately preceding and following the fricative. Thus, the total number of acoustic values considered was 59. Since many of these variables were closely related and correlated strongly, principal component analysis (PCA) was used to transform the data (equated for mean and variance) into a smaller number of more independent dimensions, which were also compared with talkers' clear speech intelligibility benefits.

Next, a similar overall clear-minus-conversational *intelligibility* difference had to be estimated for each speaker. This was less straightforward, since the adaptive procedure ensured that overall accuracy (at least toward the end of sub tests) was about the same across fricative pairs and speaking styles. However, since different trials within sub tests involved different speakers and productions, absolute difficulty was not necessarily exactly equal for all stimuli with a given snr. This lack of homogeneity (which is inevitable when using natural productions) probably added some noise to the threshold estimation procedure. However, we were able to exploit it in order to measure, in parallel, differences in the clear speech benefit across talkers. First, we verified that over the 32 total adaptive tracks that each listener in Experiment 1 heard, tokens from different speakers occurred, on average, with equal frequency and at equal signal-to-noise ratios. Then we simply took the clear-minus-conversational difference in accuracy (% correct), averaged across listeners, sub tests, and snr values, for each speaker as that speaker's approximate clear speech intelligibility advantage. While listener, sub test, and snr certainly all contributed to mean accuracy, we assumed that these contributions would essentially amount to random variability across speakers (serving only to make our measure of intelligibility advantage more conservative), and therefore no corrections were made based on these variables.

Of course, this comparison was limited in the types of acoustic-perceptual relationships it could detect. As reported by Maniwa *et al.* (submitted), clear fricatives were characterized not only by overall differences in acoustic measures depending on speaking style, but by numerous and complex Style × Fricative interactions. Since correlation analysis capable of capturing these higher-order acoustic differences was not feasible given the constraints of the perception experiments described here (individual speakers were not represented well enough within subtests to ensure equalized average snr), we did not consider these patterns in the present study.

B. Results and discussion

1. Fricative intelligibility for listeners with normal hearing

Figure 1 shows mean snr thresholds as a function of fricative pair and speaking style. The Style × Pair ANOVA showed an effect of Style [$F(1,13)=149.5$, $p>0.001$], with 3.1 dB lower thresholds for clear speech than for conversational speech, indicating that clearly produced fricatives are more intelligible than casually produced fricatives for listeners with normal hearing in degraded listening conditions. The Pair effect was also significant [$F(7,91)=113.8$, $p<0.001$]; across speaking styles, thresholds were lowest for the voiceless sibilant place of articulation contrast /s/-/ʃ/, followed by /s/-/z/ and /ʃ/-/ʒ/. Non-sibilant place of articulation pairs /f/-/θ/ and /v/-/ð/ were the most difficult, in accordance with previous studies (e.g. Jongman *et al.*, 2000; Miller and Nicely, 1955; Wang and Bilger, 1973). The Style × Pair interaction was marginally significant [$F(7,91)=2.1$, $p=0.051$], probably due to pairs /v/-/ð/ and /f/-/v/. *Post-hoc* comparisons revealed that the "clear speech effect" did not reach significance for these two pairs; all other pairs showed significant clear speech advantages.

The Style × Sibilance × Distinction Type ANOVA revealed a main effect of Sibilance [$F(1,27)=370.9$, $p<.001$] with lower thresholds for sibilants than for non-sibilants. The main effect of Distinction Type was also significant [$F(1,27)=103.7$, $p<0.001$] with lower thresholds for voicing distinctions relative to place of articulation distinctions. A Style × Sibilance interaction [$F(1,27)=10.33$, $p<0.01$] showed that while both sibilants and non-sibilants were more intelligible in clear speech, the effect was larger for sibilant

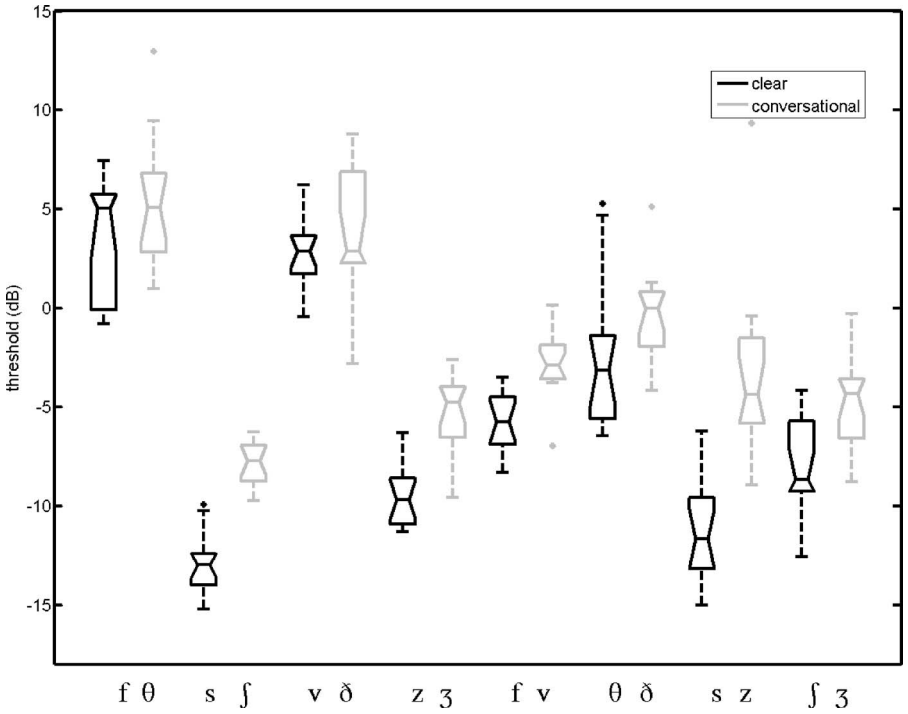


FIG. 1. Signal-to-noise ratio (snr) thresholds (dB) as a function of style and fricative pair in Experiment 1. Boxplots show the median, upper, and lower quartile, and outlier data (asterisks).

pairs. The Style × Distinction Type interaction was not significant [$F < 1$]; clear speech resulted in similar benefits for place and voicing distinctions. There was no Style × Sibilance × Distinction Type interaction. In accordance with previous findings (e.g. Jongman *et al.*, 2000; Miller and Nicely, 1955; Wang and Bilger, 1973), sibilant pairs and voicing distinction pairs were easier to identify relative to non-sibilant and place of articulation pairs, respectively, regardless of speaking style.

2. Talker-related acoustic-phonetic correlates of clear intelligibility advantage

On average, individual talkers appeared in 336 (std. 18.8) clear and 336 (20.01) conversational trials. Due to the adaptive procedure and the initial threshold guesses of ±3 dB across styles and pairs, talkers appeared at −4.91 dB (std. 0.28) and −2.61 dB (0.22) snr values, and were responded to with 81.9% (std. 4.58) and 77.3% (5.0) accuracy in clear and conversational conditions, respectively. Averaged across listeners, contrasts, and snr values, the clear-minus-conversational difference in accuracy (% correct) varied considerably across speakers, from −4% to +11% (mean 4.6%, std. 3.9%), at least partly as a result of differences in the clear speech strategies that these talkers employed (i.e., this difference did not correlate well [$p = 0.34$] with clear-minus-conversational snr differences). As described above, then, individual speakers' previously reported average style-related differences in production were compared with their style- related intelligibility differences in an effort to relate clear speech benefits to specific acoustic modifications. Table I summarizes the results of Pearson's correlations between each individual acoustic measure and the clear-minus conversational threshold difference. Positive correlations were obtained between intelligibility advantages and acoustic modifications in DFT and LPC peak location, spectral moment 1, and the slope before the peak, at most window locations. These results suggest that a shift of spectral energy to higher frequency regions and greater source strength (Jesus and Shadle, 2002) in clear fricatives—resulting in higher peak locations, higher frequency content on average, and more defined peaks—are most closely related to the overall intelligibility enhancement.

Principal components analysis of acoustic measures supported this observation. Figure 2 shows the contributions of individual acoustic measures to the first two components. The first component accounted for 41% of the variability; acoustic variables with the highest-magnitude coefficients for this component were those related to source strength and energy at higher frequencies (higher peaks and mean frequencies, lower skewness at central window locations). Talker scores for the first component correlated significantly with their clear speech benefit ($r = 0.45$; $p = 0.047$). The next two-components accounted for 14% and 11% of the variability, respectively. Intensity measures seemed to contribute most to the second component, and slope after peak locations most to the third; neither correlated significantly with the clear speech benefit ($p > 0.5$).

Since perception of place and voicing distinctions probably involve different acoustic cues, this analysis was repeated separately for the four place distinction subtests and the four voicing sub tests in the experiment. Comparison of these analyses suggested that most of the effects mentioned above were due to place of articulation distinctions. As shown in Table I, considering only place distinctions, strong positive correlations between clear speech intelligibility and acoustic differences were found for peak locations and slope before the peak, whereas negative correlations were seen for M3. Similarly, the first acoustic principal component correlated strongly with the clear speech benefit in place of articulation (POA) distinctions ($r = 0.61$, $p = 0.004$), but none of the

TABLE I. Correlation coefficients (Pearson's r) showing the relation between the clear-minus-conversational differences in acoustic measures and the clear-minus-conversational differences in the intelligibility (percent identification correctness) in Experiments 1 and 2. Significant values, $p<0.001$, $p<0.01$, and $p<0.05$ are starred as ***, **, and *, respectively. Moderate values, $p<0.1$ are marked as., and no effect was given N. Negative correlation was marked as ξ.

	Experiment 1			Experiment		
	Overall	Place	Voicing	Overall	Sibilant	Nonsibilant
Durs	0.3	0.11	0.25	−0.09	−0.02	−0.14
F2W1	−0.25	−0.21	−0.12	0.32	0.25	0.14
F2W2	0.34	0.45*	0.07	−0.28	−0.2	−0.19
F2W3	0.16	0.1	0.15	−0.33	−0.14	−0.43
F2W4	0.24	−0.08	0.34	0.05	−0.1	0.1
DFTpkWl	0.24	0.23	0.12	0.18	−0.11	0.25
DFTpkW2	0.53*	0.55*	0.21	0.1	−0.31	0.26
DFTpkW3	0.47*	0.56**	0.11	−0.03	−0.39	0.29
DFTpkW4	0.38	0.64**	−0.06	−0.03	−0.43	0.35
DFTpkW5	−0.12	−0.37	0.16	−0.08	−0.17	−0.06
HNR	−0.04	−0.19	0.08	−0.21	0.14	−0.3
Int500Wl	0.01	0.27	0.22	−0.02	0.12	−0.14
Int500W2	−0.32	−0.37	−0.07	−0.01	−0.07	−0.08
Int500W3	−0.32	−0.42	−0.02	−0.16	−0.09	−0.3
Int500W4	−0.35	−0.39	−0.09	−0.16	−0.12	−0.25
Int500W5	0.26	0.32	0.06	0.43	0.41	0.44
M1W1	0.32	0.38	0.1	0.16	−0.13	0.24
M1W2	0.51*	0.41	0.29	0.17	0.07	0.27
M1W3	0.43	0.39	0.21	0.22	−0.05	0.39
M1W4	0.41	0.46*	0.12	0.14	−0.13	0.33
M1W5	0.17	−0.03	0.24	−0.01	−0.44	−0.02
M2W1	0.36	0.52*	0.04	0.02	−0.12	0.08
M2W2	0.25	0.26	0.1	0.06	−0.34	−0.16
M2W3	0.2	0.16	0.12	−0.06	−0.34	0.09
M2W4	0.14	0.2	0.02	−0.23	−0.51*	0.01
M2W5	0.19	−0.12	0.34	−0.19	−0.43	−0.25
M3W1	−0.36	−0.56**	−0.01	−0.15	−0.01	−0.16
M3W2	−0.46*	−0.56*	−0.11	−0.12	0.1	−0.14
M3W3	−0.36	−0.48*	−0.04	−0.09	0.34	−0.21
M3W4	−0.28	−0.54*	0.09	0.04	0.44	−0.14
M3W5	−0.38	−0.26	−0.26	−0.14	0.24	−0.08
M4W1	−0.31	−0.6**	−0.09	−0.11	−0.05	−0.06
M4W2	−0.31	−0.56*	−0.07	−0.01	0.24	0.02
M4W3	−0.24	−0.47*	−0.09	0.08	0.33	0.02
M4W4	−0.17	−0.49*	−0.19	0.21	0.55*	0.07
M4W5	−0.39	−0.28	−0.25	−0.11	0.26	−0.02
follF0	0.24	0.66**	−0.23	0.39	0.15	0.49*
prevF0	−0.15	−0.1	−0.13	−0.11	−0.13	0.18
FSRA	0.04	−0.02	0.1	−0.09	−0.03	0.11
rmsampW1	0	−0.34	0.28	−0.18	0.1	−0.34
rmsampW2	−0.12	−0.33	0.15	−0.13	0.13	−0.27
rmsampW3	−0.06	−0.25	0.23	−0.21	0.14	−0.41
rmsampW4	−0.05	−0.24	0.17	−0.11	0.27	−0.32
rmsampW5	0.34	0.34	0.16	0.33	0.42	0.26
SlpAftW1	−0.42	−0.12	−0.4	−0.21	−0.03	−0.25
SlpAftW2	−0.29	−0.36	−0.04	−0.12	−0.09	−0.16
SlpAftW3	−0.16	−0.33	0.09	−0.01	0.22	−0.13
SlpAftW4	−0.02	−0.24	0.2	−0.1	0.27	−0.2
SlpAftW5	−0.15	−0.29	0.09	−0.27	−0.11	−0.42
SlpBefW1	0.42	0.54*	0.11	0.09	0.05	−0.06
SlpBefW2	0.39	0.54*	0.04	0	0.17	−0.12
SlpBefW3	0.39	0.62**	−0.01	0.01	0.28	−0.24
SlpBefW4	0.45*	0.62**	0.06	0.02	0.23	−0.15
SlpBefW5	0.29	0.45*	−0.01	0.12	0.34	−0.14
LPCPeakW1	0.3	035	0.11	0.15	0.01	0.17

TABLE I. (*Continued.*)

	Experiment 1			Experiment		
	Overall	Place	Voicing	Overall	Sibilant	Nonsibilant
LPCPeakW2	0.57**	0.65**	0.16	0.16	−0.34	0.31
LPCPeakW3	0.49*	0.54*	0.15	0.09	−0.45*	037
LPCPeakW4	0.44*	0.63**	0.03	0.02	−0.46*	0.41
LPCPeakW5	−0.02	−0.17	0.13	−0.18	−0.18	0.05

other components. These results clearly suggest that shifts toward higher frequency regions, and greater source strength, are likely to contribute to the better recognition of place of articulation for fricatives. In contrast, no significant correlations were observed between any acoustic measures—or principal components—and intelligibility benefits for voicing distinctions

III. EXPERIMENT II: EFFECTS OF CLEAR SPEECH FOR FRICATIVE RECOGNITION BY LISTENERS WITH SIMULATED HEARING IMPAIRMENT

A. Simulation method

1. Rationale

Experiment 1 results suggest that intelligibility advantages for place-of-articulation distinctions are related to spectral changes in clear speech; higher peak locations, higher mean frequency, lower skewness (more positive spectral tilt), and steeper spectral slopes before peak locations contributed to higher correct identification scores in clear speech. Given these apparent relationships, it is important to ask whether the clear fricative advantages would hold for listeners who have impaired hearing at higher frequencies. Listeners with sloping hearing losses have considerable difficulty recognizing sounds that have important acoustic information in higher frequency regions, such as fricatives (Dubno *et al.*, 1982; Owens *et al.*, 1972; Sher and Owens, 1974). These difficulties may at least partially derive from suprathreshold abnormalities in the perceptual analysis of the speech signal, including reduced dynamic range (related to loudness recruitment) (e.g., Villchur, 1974), reduced frequency selectivity (e.g., Glasberg and Moore, 1989; Moore and Peters, 1992), and impaired temporal resolution (e.g., Fitzgibbons and Wightman, 1982; Glasberg *et al.*, 1987; Glasberg and Moore, 1992). It is difficult to determine which aspects of auditory processing contribute most to degraded speech reception, since elevation of absolute thresholds is usually correlated with a variety of suprathreshold changes that have similar effects. A common strategy for controlling these confounding factors is to process sounds to simulate the effects of one specific aspect of hearing impairment, and to allow listeners with normal hearing to experience selected perceptual effects of hearing impairments. In this experiment, we were particularly interested in the influence of high-frequency threshold elevation on recognition of fricative sounds, since important fricative information occurs in fre-

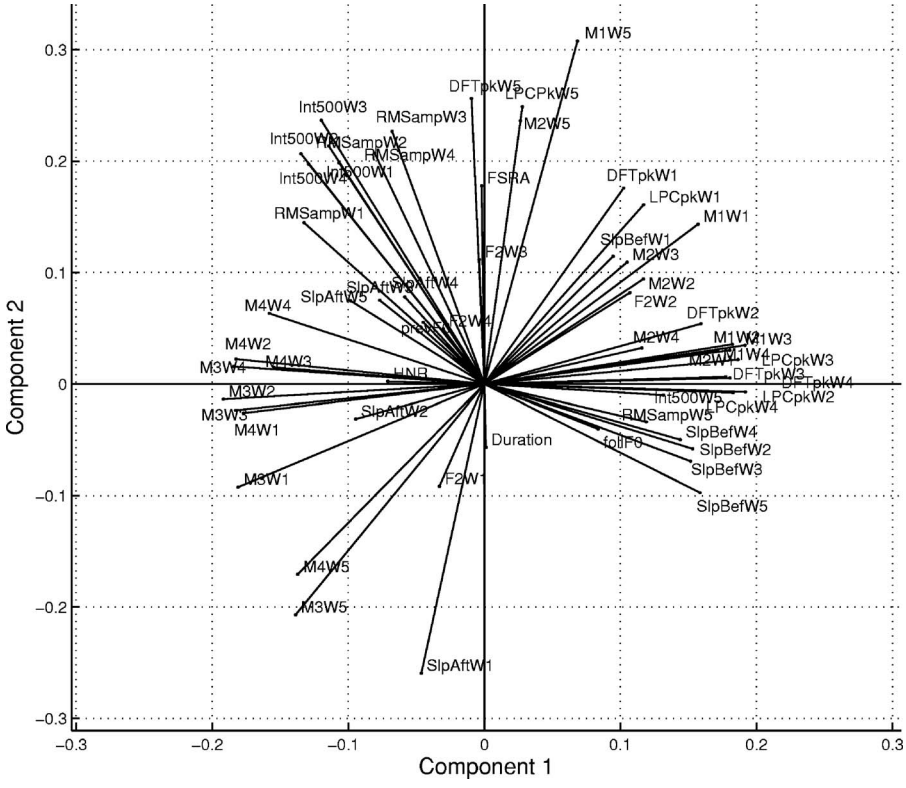


FIG. 2. Coefficients of individual measures (see text and Table I for abbreviations) for the first two components resulting from principal components analysis of acoustic data.

quency ranges where many impaired listeners have elevated thresholds, and since Experiment 1 suggests that this may be increasingly so for clear fricatives. It is possible that listeners with sloping hearing loss cannot make use of enhanced acoustic-phonetic information since it is less audible to them. To assess how this aspect of hearing impairment would affect the perception of clear fricative sounds, we repeated the perception experiment using stimuli processed to simulate sloping hearing loss.

2. Implementation

Sloping, recruiting hearing loss was simulated in a manner similar to that described by Moore and Glasberg (1993), with some modifications due to a higher sampling rate (44.1 kHz) and the fact that all processing was done on-line during the experiment. Following the combination of signal and noise components, stimuli were separated into 24 equivalent rectangular bandwidth (ERB)-spaced bands, from 100 Hz to 22.05 kHz, using fourth order gammatone filters (Slaney, 1998). For each band, a smoothed envelope (E) was derived by low-pass filtering the full-wave rectified waveform at 100 Hz (fourth order Butterworth filter, implemented in both forward and reverse directions to minimize phase distortions). The temporal fine structure for the band was then extracted by dividing the original waveform by this envelope. Loss simulation was accomplished by raising the envelope to a power related to the slope of the loudness growth function:

$$E_p = E^N,$$

where N is frequency dependent. Following Moore and Glasberg (1993), N was a constant 1.5 at bands up to 900 Hz, increased linearly to 3.0 at 4500 Hz, and remained at this value for all higher bands. Finally, the modified stimulus was obtained by multiplying E_p by the fine structure and summing the resulting band-limited waveforms. All processing was performed in MATLAB. Processing on average took ~2 s; this resulted in an inter-trial interval that a few participants found slightly annoying but generally not distracting.

B. Experiment method

1. Participants

Fourteen normal-hearing listeners (9 F, 5 M) aged between 19 and 33 were recruited from the University of California, Berkeley community. Participants were native speakers of American English, without noticeable regional dialects. Participants reported normal hearing and no history of speech or language disorders. Listeners were paid for their participation.

2. Materials

Test stimuli were identical to those of Experiment 1 except that (1) speech/babble stimuli were processed as described above, and that (2) only the four place-of-articulation pairs /f/-/θ/, /v/-/ð/, /s/-/ʃ/, and /z/-/ʒ/ were tested, since these were the contrasts for which increased high-frequency content seemed to benefit normal-hearing listeners.

3. Procedures and apparatus

The procedure, task, presentation method, and adaptive procedure were identical to those of Experiment 1, except that since only four pairs were tested there was no break during the experiment. Testing took about 50 min.

4. Data analysis

As in Experiment 1, a repeated measure analysis of variance (ANOVA) with two within-subject factors (Style; 2 levels, Pair; 4 levels) and thresholds (dB snr) as dependent variable was performed. Acoustic measures and principal components were similarly compared across talkers with the clear speech intelligibility advantage.

C. Results and discussion

1. Fricative intelligibility for listeners with simulated hearing loss

Figure 3 shows snr thresholds as a function of pair type for clear and conversational fricative identification. For all place pairs except /f/-/θ/, clear speech showed lower snr thresholds relative to conversational speech. The Style × Pair ANOVA showed an effect of Style [$F(1,13)=13.9$, $p<.01$] with 2.5 dB lower thresholds for clear speech. There was also a Pair effect [$F(3,39)=149.5$, $p<0.001$], mostly derived from lower thresholds for sibilant pairs relative to non-sibilant pairs. The Style × Pair interaction was also significant [$F(3,39)=6.0$, $p<0.01$]. Pairwise comparisons showed significant differences in thresholds as a function of style for /s/-/ʃ/ and /z/-/ʒ/ pairs, but not for non-sibilant pairs. In fact, for /f/-/θ/, clear speech resulted in higher (n.s.) thresholds compared to conversational speech. These results thus differed from Experiment 1 results in that (1) thresholds were on average much higher, (2) there was no clear speech effect for /f/-/θ/, and (3) the /z/-/ʒ/ pair showed the biggest clear speech effect, followed by /s/-/ʃ/, /v/-/ð/, and /f/-/θ/; in Experiment 1 the order was /z/-/ʒ/, /f/-/θ/,/s/-/ʃ/ and /v/-/ð/. On the other hand, the relative overall difficulty of fricative pairs was similar to Exp. 1; across speaking styles, the pair /s/-/ʃ/ resulted in the lowest thresholds, followed by /z/-/ʒ/, /v/-/ð/, and /f/-/θ/.

To determine how the loss simulation influenced the perception of fricatives in interaction with speaking style and contrastive pair, a three-way mixed model ANOVA was performed with two within-subject factors (Style, Pair) and listener group as a between-subject factor (two levels; Exp. 1 and Exp. 2). Since the four voicing distinction pairs were not included in Experiment 2, only the four place-of-articulation distinction pairs from Experiment 1 were considered. This analysis showed a main effect of Group [$F(1,26)=26.4$, $p<0.001$] with considerably (4.47 dB) higher thresholds for listeners with simulated hearing. A main effect of Style [$F(1,26)=49.0$, $p<0.001$] indicated, again, an overall clear speech advantage across listener groups. There was no Style × Group interaction, suggesting that, on average, listeners with normal hearing and listeners with simulated impairment enjoyed comparable significant benefits from clear speech. The main-effect of Pair was significant [$F(3,78)=212.8$, $p<0.001$] but not the Pair × Group interaction,

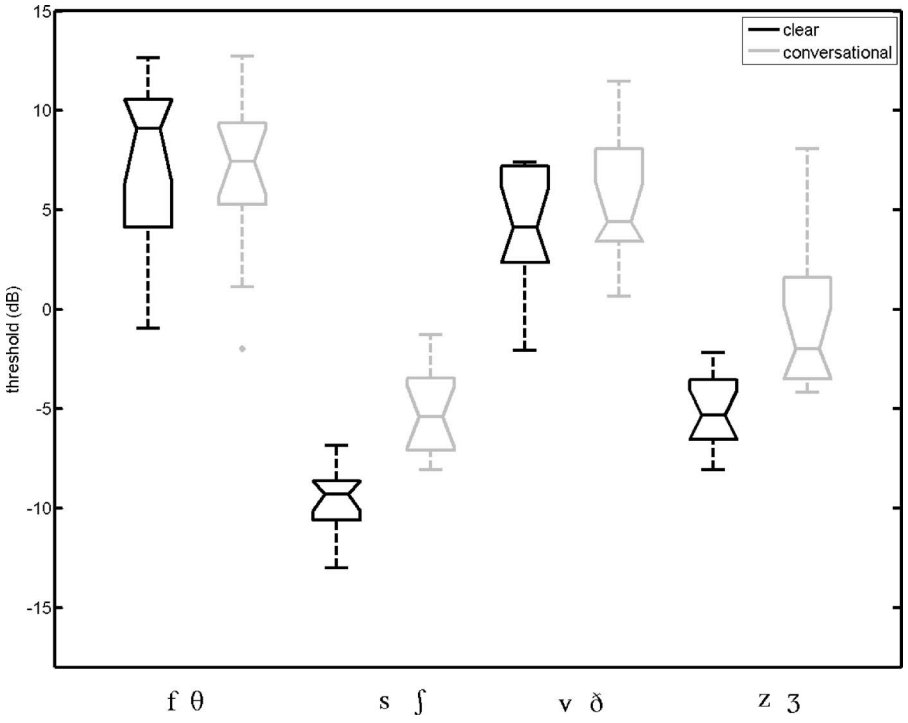


FIG. 3. Signal-to-noise ratio (snr) thresholds (dB) as a function of style and fricative pair in Experiment 2.

reflecting the common difficulty hierarchy mentioned above. Again, pairwise comparisons indicated that all four pairs were significantly different from each other, and that the effect was most notably derived from differences between sibilant and non-sibilant pairs. A Style × Pair interaction $[F(3,78)=212.8,\ p<0.01]$ indicated that, across listener groups, the clear speech effect differed depending on the fricative pair. The Style × Pair × Group interaction was significant $[F(3,78)=2.9,\ p<0.05]$; *post-hoc* tests suggested that the interaction was related to an increase in the magnitude of the clear effect for sibilants, and a *decrease* in the effect for non-sibilants, in the simulated impairment condition. This finding is illustrated in Fig. 4, which shows the clear speech effect as a function of pair and listening condition. It seems likely that, since non-sibilants are characterized by the highest peak and F2 values with a diffuse spread of energy below 10 kHz, important spectral cues for these sounds are less audible/available to listeners with sloping hearing loss the higher they are transposed. Sibilants, on the other hand, have both higher relative amplitudes and more potential cues (esp. palatoalveolar peak frequencies) involving energy in lower regions. These cues would be better preserved in stimuli with simulated sloping loss.

2. Acoustic correlates of intelligibility benefit for listeners with simulated hearing impairment

In Experiment 2, individual talkers appeared on average in 168 (std. 12.4) clear and 168 (11.63) conversational trials. Again, averaged across listeners, contrasts, and snr values, the clear-minus-conversational difference in accuracy (% correct) varied considerably across speakers, from −6% to +18% (mean 3.9%, std. 6.6%). As discussed in Experiment 1, individual speakers' previously reported average style-

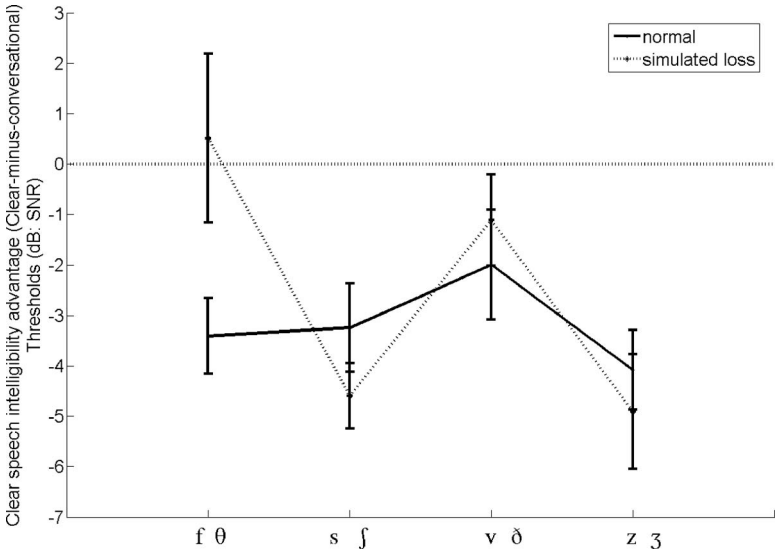


FIG. 4. Clear speech intelligibility advantage (clear-minus-conversational thresholds) in dB snr for listeners with normal hearing and listeners with simulated hearing impairment as a function of fricative pair.

related differences in production were compared with their style-related intelligibility differences in a first effort to relate clear speech benefits to specific acoustic modifications.

The results of individual measure correlations are shown in Table I. Overall, correlations were much less consistent than for Experiment 1; in particular, conspicuously absent were the positive correlations with spectral measures indicating shifts to higher frequency regions that were seen for place contrasts in Exp. 1. Since the perception of sibilant and non-sibilant pairs was affected differentially by the impairment, another set of correlation analyses compared intelligibility differences across speakers with acoustic differences separately for each class of sounds. While this comparison was considerably less well powered than the others described above, the results were potentially interesting and are also included in Table I. For sibilant pairs, *positive* correlations were seen between intelligibility advantages and spectral moment 3, and *negative* correlations with peak locations, at several window locations. For non-sibilant pairs, correlations were weaker and less straightforward. No significant correlations (all $p > 0.3$) were seen between clear speech advantages, either overall or considering sibilants and non-sibilants separately, with any of the acoustic principal components discussed above. Interestingly, the (nonsignificant) correlation between the first component (related to high-frequency energy) and the benefit for sibilants was negative.

IV. DISCUSSION

A. Overall clear fricative intelligibility

In two experiments, lower snr identification thresholds for place of articulation identification were seen for clear relative to conversational fricatives, indicating that, on average, clearly produced fricatives are more intelligible for both young normal-hearing listeners and listeners with simulated sloping, recruiting hearing impairment. In addition, clear speech was beneficial to normal-hearing listeners for voicing distinctions. However, these effects were not as uniform and robust across fricatives and listener groups as might have been expected. In Experiment 1, sibilant fricatives were easier to identify than non-sibilants for normal-hearing listeners overall, and clear speech provided slightly greater intelligibility benefits for sibilants than non-sibilants. Experiment 2 showed that these trends were exaggerated for simulated hearing-impaired listeners. In particular, a clear speech effect was seen *only* for sibilants, and clear speech may have even hurt intelligibility for voiceless non-sibilants, the worst-recognized sounds. These results are consistent with the notion (e.g., Ferguson and Kewley-Port, 2002) that the perceptual effects of clear speech acoustic modifications may be population dependent, and may interact in complex ways with different types of hearing impairment. As discussed below, they probably derive from differences in the audibility and weighting of acoustic cues across fricatives and listening conditions.

B. Acoustic and talker-related correlates of clear speech intelligibility effect

Comparison of individual speakers' estimated clear-speech intelligibility advantages with their previously reported (Maniwa *et al.*, submitted) clear-speech acoustic modifications revealed correlations that may be informative as to the acoustic sources of the "clear speech effect" in fricatives. Specifically, for place-of-articulation distinctions, strong positive correlations were found between acoustic and perceptual clear-vs-conversational differences for spectral measures, especially at central locations, including peak locations, M1, and spectral slope before peak locations. In addition, there were negative correlations between intelligibility improvement and increases in M3 and M4, and for intensity below 500 Hz. These results indicate that, overall, greater source strength (produced by higher volume velocity) in clear speech, resulting in spectral distributions with higher-frequency, more defined peaks, and more positive tilt, contributed to the intelligibility enhancement for place distinctions. Of course, it is more likely that these "global" changes in conjunction with higher-order patterns specific to individual fricatives and contexts actually led to the intelligibility effects that were seen. The experiments described here could not address this possibility, since within individual subtests snr values were not sufficiently equalized across speakers to make more specific comparisons. Probably for this reason, no strong correlations were seen relating acoustic measures and voicing intelligibility. In particular, acoustic results suggested that phonetic distance in terms of the voicing distinction was often "enhanced" in clear speech by increasing (or decreasing to a lesser degree) values for one class of fricatives while decreasing (or increasing to a lesser degree) values for the class. For example, intensity below 500 Hz decreased much less, and HNR significantly increased, for voiced fricatives whereas these values significantly decreased for voiceless fricatives. Similarly, noise duration and f0 increased for both voiceless and voiced fricatives in clear speech, but to a much greater extent for voiceless fricatives. These differences in clear speech manipulations, and their perceptual effects, would have mostly been obscured by the analysis described here.

Our previous study (Maniwa *et al.*, submitted) also indicates that voiceless non-sibilants have, in addition to very low amplitudes, very high peak frequencies (higher than /s/), mean frequency, and F2, across speaking styles, and that these values are even higher in clear speech. This was probably a cause of the lack of clear speech benefits for (especially voiceless) non-sibilants, since the simulated impairment targeted higher frequencies (and low amplitudes). Sibilants, on the other hand, were characterized by more and lower energy, in some cases (esp. palato-alveolars) even more so in clear speech, so more potential cues for these sounds were preserved in the loss simulation. As a result of these differences, for listeners with simulated hearing impairment few overall correlations between acoustic and intelligibility differences in clear speech were apparent in Experiment 2. For identification of sibilant pairs specifically, contrary to Experiment 1 results, there were some *negative* correlations between acoustic changes in peak frequencies

and enhanced intelligibility (and marginal *positive* correlation between M3 and intelligibility advantage). This suggests that the lower the spectral information moved for palato-alveolar fricatives in clear speech, the more intelligible these sounds were, because this information was better preserved in the impairment simulation. Fewer and less consistent patterns could be seen to relate non-sibilant acoustic modifications to intelligibility. In other words, elevated thresholds and loudness recruitment influenced listeners' cue weighting for the perception of fricative sounds.

There were no Style × Gender interactions in either experiment, indicating that female and male talkers did not differ in terms of the effectiveness of their clear speech acoustic modifications for intelligibility (cf. Bradlow *et al.*, 2003).

C. Conclusion

This study showed that clear speech enhanced the intelligibility of fricatives for both listeners with normal hearing and listeners with simulated hearing impairment. However, the effect was fricative and population dependent; notably, compared to normal-hearing listeners, impaired listeners showed reduced clear speech effects for non-sibilant place of articulation distinctions. Likewise, apparent acoustic correlates of the clear speech benefit differed across populations. For normal-hearing listeners, intelligibility benefits seemed to correlate with moves toward higher frequency regions for important cues; these patterns were generally not seen for impaired listeners, and may even have been reversed for some sounds. These results are straightforwardly explained based on audibility of cues at different levels and frequencies. We leave for future study a more thorough investigation of potential higher-order acoustic correlates of the clear speech effect in fricatives; this could be accomplished straightforwardly by using the results of the adaptive design described here to inform blocked-design experiments that are optimally controlled (and powered) for the distribution of fricatives, styles, and snr values across speakers and tokens. It will also be necessary to measure perception by actual hearing-impaired listeners in order to characterize the population-based differences we observed more quantitatively.

ACKNOWLEDGMENT

Portions of this research were conducted as part of K.M.'s doctoral dissertation under the supervision of A.J.

Behrens, S. J., and Blumstein, S. E. (**1988**). "On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants," J. Acoust. Soc. Am. **84**, 861–867.

Boothroyd, A. (**1984**). "Auditory perception of speech contrasts by subjects with sensorineural hearing loss," J. Speech Hear. Res. **27**, 134–144.

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Chen, F. R. (**1980**). "Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level," Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Cole, R. A., and Cooper, W. E. (**1975**). "Perception of voicing in English affricates and fricatives," J. Acoust. Soc. Am. **58**, 1280–1287.

Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**, 249–261.

Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**). "Evaluation of hearing-impaired listeners using a Nonsense-syllable Test. II. Syllable recognition and consonant confusion patterns," J. Speech Hear. Res. **25**, 141–148.

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.

Ferguson, S. H. (**2002**). "Vowels in clear and conversational speech: Talker differences in acoustic features and intelligibility for normal-hearing listeners," Doctoral dissertation, Indiana University, Bloomington, IN.

Ferguson, S. H., 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.

Fitzgibbons, P. J., and Wightman, F. L. (**1982**). "Gap detection in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **72**, 761–765.

Gagné, J. P., Masterson, V. M., Munhall, K. G., Bilida, N., and Querengesser, C. (**1994**). "Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech," J. Acad. Rehabil. Audiol. **27**, 135–158.

Gagné, J. P., Querengesser, C., Folkeard, P., Munhall, K. G., and Masterson, V. M. (**1995**). "Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech," Volta Rev. **97**, 33–51.

Gagné, J. P., Rochette, A.-J., and Charest, M. (**2002**). "Auditory, visual and audiovisual clear speech," Speech Commun. **37**, 213–230.

Glasberg, B. R., and Moore, B. C. J. (**1992**). "Effects of envelope fluctuations on gap detection," Hear. Res. **64**, 81–92.

Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Glasberg, B. R., Moore, B. C. J., and Bacon, S. P. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1546–1556.

Guerlekian, J. A. (**1981**). "Recognition of the Spanish fricatives /s/ and /f/," J. Acoust. Soc. Am. **70**, 1624–1627.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech **1**, 1–7.

Hedrick, M. S. (**1997**). "Effect of acoustic cues on labeling fricatives and affricates," J. Speech Lang. Hear. Res. **40**, 925–938.

Hedrick, M. S., and Carney, A. E. (**1997**). "Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants," J. Speech Lang. Hear. Res. **40**, 1445–1457.

Hedrick, M. S., and Ohde, R. N. (**1993**). "Effect of relative amplitude of frication on perception of place of articulation," J. Acoust. Soc. Am. **94**, 2005–2026.

Hedrick, M. S., and Younger, M. S. (**2003**). "Labeling of /s/ and /ʃ/ by listeners with normal and impaired hearing, revisited," J. Speech Lang. Hear. Res. **46**, 636–648.

Heinz, J. M., and Stevens, K. N. (**1961**). "On the properties of voiceless fricative consonants," J. Acoust. Soc. Am. **33**, 589–596.

Helfer, K. (**1997**). "Auditory and auditory-visual perception of clear and conversational speech," J. Speech Lang. Hear. Res. **40**, 432–443.

Helfer, K. (**1998**). "Auditory and auditory-visual recognition of clear and conversational speech by older adults," J. Am. Acad. Audiol **9**, 234–242.

Hughes, G. W., and Halle, M. (**1956**). "Spectral properties of fricative consonants," J. Acoust. Soc. Am. **28**, 303–310.

Iverson, P., and Bradlow, A. R. (**2002**). "The recognition of clear speech by adult cochlear implant users," in *Temporal Integration in the Perception of Speech*, edited by S. Hawkins and N. Nguyen (Cambridge: Center for Research in the Arts, Aix-en-Provence, France, Social Sciences, and Humanities), p. 78.

Jesus, L. M. T., and Shadle, C. H. (**2002**). "A parametric study of the spectral characteristics of European Portuguese fricatives," J. Phonetics **30**, 437–464.

Jongman, A. (**1989**). "Duration of frication noise required for identification of English fricatives," J. Acoust. Soc. Am. **85**, 1718–1725.

Jongman, A., Wang, Y., and Sereno, J. (**2000**). "Acoustic and perceptual properties of English fricatives," *Proceedings of the International Conference on Spoken Language Processing*, Beijing, China, **II**, 511–514.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., and Supowit,

A. (**1994**). "Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation," Vision Res. **34**, 885–912.
Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.
Liu, S., Del Rio, E. Bradlow, A. R., and Zeng, F.-G. (**2004**). "Clear speech perception in acoustic and electric hearing," J. Acoust. Soc. Am. **116**, 2373–2383.
Maniwa, M., Jongman, A., and Wade, T. (**2007**). "Acoustic and perceptual properties of clearly produced fricatives," Unpublished doctor's dissertation, The University of Kansas, Lawrence, KS.
Mathworks, Inc., The. (**2000**). "MATLAB, The language of technical computing, version 7.0.0.19920."
McCasland, G. P. (**1979a**). "Noise intensity and spectrum cues for spoken fricatives," J. Acoust. Soc. Am. **65**, S78–S79.
McCasland, G. P. (**1979b**). "Noise intensity cues for spoken fricatives," J. Acoust. Soc. Am. **66**, S88.
Miller, G. A., and Nicely, P. A. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J., and Glasberg, B. R. (**1993**). "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech," J. Acoust. Soc. Am. **94**, 2050–2062.
Moore, B. C. J., and Peters, R. W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.
Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 351–382.
Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.
Nittrouer, S., and Miller, M. E. (**1997a**). "Predicting developmental shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**, 2253–2266.
Nittrouer, S., and Miller, M. E. (**1997b**). "Developmental weighting shifts for noise components of fricative-vowel syllables," J. Acoust. Soc. Am. **102**, 572–580.
Owens, E. (**1978**). "Consonant errors and remediation in sensorineural hearing loss," J. Speech Hear Disord. **43**, 331–347.
Owens, E., Benedict, M., and Schubert, E. D. (**1972**). "Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairments," J. Speech Hear. Res. **15**, 308–322.
Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.
Raphael, L. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.
Schum, D. (**1996**). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol **7**, 212–218.
Sher, A. E., and Owens, E. (**1974**). "Consonant confusions associated with hearing loss above 2000 Hz," J. Speech Hear. Res. **17**, 669–681.
Singh, S., and Black, J. W. (**1966**). "Study of twenty-six intervocalic consonants as spoken and recognized by four language groups," J. Acoust. Soc. Am. **39**, 372–387.
Slaney, M. (**1998**). "Auditory Toolbox, version 2.0," http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/. last viewed 25 September 2007.
Soli, S. D. (**1982**). "Structure and duration of vowels together specify fricative voicing," J. Acoust. Soc. Am. **72**, 366–378.
Soli, S. D., and Arabie, P. (**1979**). "Auditory versus phonetic accounts of observed confusions between consonant phonemes," J. Acoust. Soc. Am. **66**, 46–58.
Stevens, K. N. (**1985**). "Evidence for the role of acoustic boundaries in the perception of speech sounds," in *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, edited by V. Fromkin (Academic, New York), pp. 243–255.
Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (**1992**). "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters," J. Acoust. Soc. Am. **91**, 2979–3000.
Uchanski, R. S., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.
Villchur, E. (**1974**). "Simulation of the effect of recruitment on loudness relationships in speech," J. Acoust. Soc. Am. **56**, 1601–1611.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.
Whalen, D. H. (**1981**). "Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary," J. Acoust. Soc. Am. **69**, 275–282.
Zeng, F.-G., and Turner, C. W. (**1990**). "Recognition of voiceless fricatives by normal and hearing-impaired subjects," J. Speech Hear. Res. **33**, 440–449.

Perceptual learning of spectrally degraded speech and environmental sounds

Jeremy L. Loebach[a)]
Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405

David B. Pisoni[b)]
Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405 and DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana 46202

(Received 15 March 2007; revised 5 November 2007; accepted 20 November 2007)

Adaptation to the acoustic world following cochlear implantation does not typically include formal training or extensive audiological rehabilitation. Can cochlear implant (CI) users benefit from formal training, and if so, what type of training is best? This study used a pre-/posttest design to evaluate the efficacy of training and generalization of perceptual learning in normal hearing subjects listening to CI simulations (eight-channel sinewave vocoder). Five groups of subjects were trained on words (simple/complex), sentences (meaningful/anomalous), or environmental sounds, and then were tested using an open-set identification task. Subjects were trained on only one set of materials but were tested on all stimuli. All groups showed significant improvement due to training, which successfully generalized to some, but not all stimulus materials. For easier tasks, all types of training generalized equally well. For more difficult tasks, training specificity was observed. Training on speech did not generalize to the recognition of environmental sounds; however, explicit training on environmental sounds successfully generalized to speech. These data demonstrate that the perceptual learning of degraded speech is highly context dependent and the type of training and the specific stimulus materials that a subject experiences during perceptual learning has a substantial impact on generalization to new materials. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2823453]

PACS number(s): 43.71.Es, 43.72.Gy, 43.66.Ts, 43.71.Sy [MSS] Pages: 1126–1139

I. INTRODUCTION

Despite recent advances in cochlear implant technology, a large amount of variability in outcome and benefit is consistently reported among cochlear implant (CI) users that cannot be accounted for by differences in etiology, onset and duration of deafness, age at implantation and physiological factors (NIH, 1995). Given that there are no standardized rehabilitation programs consistently implemented after implantation, the experiences of CI users may differ from the start, placing them at fundamentally different baseline levels and contributing to the variability in the outcome measures. Could the standardization of training protocols establish a more stable baseline, and account for a portion of this variability? Moreover, what type of training is most effective, and yields the most robust levels of generalization to new materials? The present study was designed to assess the effectiveness of training when adapting to stimuli that have been processed by a cochlear implant simulation. Subjects were trained on speech (simple or complex single words; meaningful or anomalous sentences) or environmental sounds and compared on-open set recognition at posttest and during generalization to novel stimuli to quantify how explicit training affects the perceptual learning of stimuli processed by a CI simulation.

Cochlear implantation can provide sufficient acoustic input to a deaf individual to allow the establishment of some form of hearing (NIH, 1995). Whereas early implants provided the hope of recovering some auditory ability, most recipients of modern cochlear implants have the expectation that they will recover oral communication skills, including the ability to talk on the telephone (Shannon, 2005). In the worst case, patients are expected to regain some awareness of sound (Clark, 2002), including the detection and recognition of environmental sounds, however the degree to which CI users can actually recognize and identify environmental sounds is largely unknown (see, however, Reed and Delhorne, 2005).

Research using acoustic simulations of cochlear implants has met with great success (Shannon *et al.*, 1995). Vocoders simulate the limited number of spectral channels available in the electrode array by dividing the acoustic signal using a series of band-pass filters but preserve the temporal profile of electrical stimulation by modulating the noise bands with the amplitude envelope. Previous studies using the vocoder have demonstrated that successful speech recognition can occur in response to severely spectrally degraded stimuli; only a limited number of spectral channels are re-

a)Author to whom correspondence should be addressed. Electronic mail: jlloebac@indiana.edu
b)Electronic mail: pisoni@indiana.edu

quired so long as the temporal information in the envelope is preserved (Shannon *et al.*, 1995; Dorman *et al.*, 1997). Moreover, the performance of CI users on consonants and vowels was similar to that of normal hearing subjects listening to six-channel vocoded stimuli, demonstrating that the vocoder can successfully simulate the output of a CI in order to elicit equivalent levels of performance (Dorman and Loizou, 1998).

Although studies using vocoders have focused primarily on the identification of the linguistic content of the materials, the real world is composed of many other complex auditory events that are transmitted via the acoustic signal (Gaver, 1993). Compared to speech, considerably less is known about the perception of environmental sounds, both in the clear and processed by vocoders. Although there may be some commonalities between the perceptual systems required for the identification of speech and environmental sounds, the degree to which they operate independently is unknown. At a surface level, it appears that environmental sounds may be encoded in a similar manner to speech, in that the stimulus specific form may be preserved in addition to the more abstract symbolic lexical form (Lachs *et al.*, 2003) as demonstrated by cross-modal priming of environmental sounds (Chiu and Schacter, 1995; Chiu, 2000). More recently, the perception of both speech and environmental sounds has been shown to rely on a common auditory ability for familiar sound recognition (Kidd *et al.*, 2007). How efficiently a subject can locate stored information about an auditory stimulus, the problem solving strategies they engage in to constrain possible response options, and the ability to focus attention on the most important spectral and temporal information in the signal were all found to be common factors for the identification of environmental sounds and speech in noise (Kidd *et al.*, 2007). Moreover, the authors proposed a general auditory ability (g), which governs a listener's ability to process and perceive auditory information, and may be a necessary component for speech and language processing, as well as for identifying environmental sounds (Kidd *et al.*, 2007).

Additionally, a series of recent experiments by Gygi and colleagues demonstrated that the most important acoustic information for the recognition of environmental sounds overlaps with the information for speech (1200–2400 Hz, Gygi *et al.*, 2004). When processed with a vocoder, the results were much like those reported for speech: recognition accuracy increased with the number of channels (Gygi *et al.*, 2004). Moreover, the stimuli that showed the greatest improvement from training were those that had broader harmonic structure and spectral detail (Gygi *et al.*, 2004). Other work suggests that that the effects of processing environmental sounds using a noise vocoder may not be as straightforward as it is for speech (Shafiro, 2004). Closed-set recognition of environmental sounds improved as the number of channels increased, however, the improvement was stimulus dependent: environmental sounds that rely more on spectral information showed increases in accuracy with the addition of more spectral channels, whereas those that rely on temporal information showed decreases (Shafiro, 2004). Thus, it appears that some environmental sounds may show an altogether different pattern of spectrotemporal dependence as compared to speech signals.

Few studies have examined the perception of environmental sounds by CI users. Although CI users do improve in their ability to recognize environmental sounds following implantation, they do so at a slower rate than is typically observed for speech (Tye-Murray *et al.*, 1992). In a recent study, Reed and Delhorne (2005) demonstrated that CI users were significantly better at identifying environmental sounds (79% correct) than words (39% correct), and that subjects who were better at word identification were also better at environmental stimulus identification. Significant variability was observed across subjects, however, arising from differences in exposure to environmental sounds in their daily environment (Reed and Delhorne, 2005), and it is possible that additional exposure and explicit training could further increase performance.

One common theme throughout the studies using the vocoder is the issue of perceptual learning. Although subjects can accurately identify speech processed by a vocoder, a period of learning and adjustment is frequently required (Shannon *et al.*, 1995; Dorman *et al.*, 1997; Dorman and Loizou, 1998). Although some type of auditory training is necessary when adapting to acoustic simulations of cochlear implants, few studies have explicitly examined the effects of training and feedback on the adaptation to CI simulations. Moreover, the training conditions that maximize perceptual learning and promote robust generalization and transfer to new materials are not well understood. In a series of recent experiments, Davis and colleagues investigated the effects of training and feedback during adaptation to six-channel noise vocoded sentences (Davis *et al.*, 2005). When subjects were merely exposed to the stimuli, open set identification increased significantly across 30 sentences (a gain of 11%), which can be attributed solely to perceptual attunement to the synthesis conditions, since subjects received no feedback whatsoever. When subjects were provided with unprocessed auditory feedback, significant increases in recognition accuracy were also observed (Davis *et al.*, 2005). Although these results are encouraging, the typical CI user will not have access to the unprocessed version of a stimulus, making this type of feedback not clinically viable. Adding orthographic feedback proved to be just as effective as presentation of the original unprocessed acoustic version, producing robust gains over exposure alone (Davis *et al.*, 2005).

Another issue with perceptual learning lies in which stimulus materials are most effective during training. Davis and colleagues (2005) trained subjects to identify meaningful sentences, semantically anomalous sentences (sentences where the function words are intact, but the content words are unrelated), nonword sentences, or Jabberwocky sentences (anomalous sentences where the content words are replaced by nonwords). Although all groups showed significant improvement over the training interval, those who were trained on meaningful and anomalous sentences performed identically to one another, and significantly better than those trained on non-word and Jabberwocky sentences. These findings suggest that access to the syntactic structure of the ma-

terials may be required to elicit effective levels of perceptual learning, but when the content words are replaced with nonwords the task of determining the syntactic structure becomes exceedingly difficult (Davis *et al.*, 2005).

As an initial investigation into the effectiveness of training and perceptual learning, the results reported by Davis and colleagues raise several important questions. Although they demonstrated that feedback has a significant impact on performance, the type of feedback they used would not necessarily be relevant for the typical CI user. An individual with electric hearing never has an opportunity to hear an unprocessed version of the stimulus. The finding that the subjects who received orthographic feedback paired with the vocoded version of the sentence performed just as well as those who received the unprocessed version, suggests that such feedback could be beneficial to individuals with cochlear implants. In addition, subjects who did not receive explicit feedback showed significantly lower levels of performance overall, but still achieved similar gains due to training. Due to methodological constraints, however, neither pre- to posttest gains in performance nor generalization to new materials could be assessed.

In another study, Burkholder (2005) demonstrated that the use of feedback consisting of the correct orthographic form of the sentence paired with the repetition of the vocoded stimulus produced significantly greater pre- to posttest gains than receiving the unprocessed version alone. Moreover, subjects who were trained on the anomalous sentences showed identical pre- to posttest gains to subjects trained on meaningful sentences, but showed significantly greater benefits during generalization to new materials including environmental sounds (Burkholder, 2005). These data suggest that access to the syntactic structure of the sentence without relying on sentence meaning may provide a greater benefit, presumably because the listener is forced to reallocate attention to the acoustic–phonetic structure of the signal and rely on bottom-up processes for recognition. Evidence showing that training successfully generalized to the identification of environmental sounds, underscores this point.

Several questions remain unanswered, however. Burkholder (2005) only assessed the generalization of training with speech to environmental sounds, but not the reverse. If subjects were truly relying on the acoustic structure of the stimuli, one would predict that training on environmental sounds should successfully generalize to speech. This possibility, however, has not been experimentally addressed. In addition, no baseline identification data were collected for the environmental sounds, so it is unclear if the subjects were performing significantly better at identifying the environmental sounds than with no training at all. Moreover, although training with meaningful sentences appears to generalize to novel sentences, it is unknown whether this training generalizes to single words. Anomalous sentences can be conceptualized as a series of unrelated words connected by a permissible syntactic structure. If this is the case, then training on single word identification should generalize to anomalous sentences, and training on anomalous sentences should generalize to single words. In addition, recent studies have shown that training on simple consonant vowel (CV) and consonant vowel consonant clusters (CVC) produces only modest gains in performance on sentence identification (Fu *et al.* 2006). It is unclear whether the converse is true; that is, would training on sentences, both high and low in context, generalize to single words and CVCs?

The purpose of the present study was to examine the effect of training on the recognition of speech and environmental sounds processed by a sinewave vocoder. Specifically, we assessed the perceptual learning of CVCs, words, meaningful sentences, anomalous sentences, and complex nonspeech environmental sounds using a pre-/posttest design, and then compared the generalization to different materials. As little is known on how exposure affects open-set recognition of environmental sounds, we collected baseline data from subjects who were exposed to but not explicitly trained on environmental sounds as a control. For sentences, mere exposure without feedback leads to gains of 11% over time (Davis *et al.*, 2005), suggesting that for both sentences and words, any gain from training that exceeds 10% is likely due to explicit training. Additionally, Burkholder (2005) demonstrated that subjects gain only 10% across three phases of training with semantically anomalous sentences when provided with unprocessed auditory feedback. Given that subjects were receiving feedback, it is expected that mere exposure to anomalous sentences would produce gains that would be far less than 10%.

II. METHODS

A. Subjects

One hundred fifty-five normal-hearing, healthy young adults from the Indiana University community participated in the study (107 female, 47 male, and 1 transgender). The age of the subjects ranged between 18 and 60, with a mean age of 22.38 years. All subjects reported that English was the first language that they learned in infancy. Most subjects ($n=141$) were monolingual, but 14 reported being fluent bi- ($n=12$) or trilingually ($n=2$). Subjects were given credit in their introductory psychology course for their participation ($n=52$), or were paid at the rate of $10 per hour ($n=103$). Of these 155 subjects, five were excluded from the final data analysis, leaving 25 unique subjects in each training condition. One subject was excluded after reporting that they could not understand the stimuli as speech and one due to a program malfunction. After the experiment, one subject revealed that they were not a native English speaker, and so their data were excluded. Three subjects were excluded after the decision was made that they were not on task: one subject left many spaces intentionally blank and made frequent spelling errors that rendered the data impossible to score; one only transcribed the first few words of each sentence and it was thought that they had a possible memory deficit; and the final subject typed only gibberish (random keystrokes) rather than making a meaningful response to the stimuli.

B. Stimuli

Stimulus materials came from five different corpora that consisted of digital wav files of meaningful words, meaningful sentences, anomalous sentences, and environmental sounds.

1. Modified rhyme test

The modified rhyme test (MRT) corpus consisted of 300 words organized into 50 lists, where each list contains six rhymed variations on a common syllable (House *et al.*, 1965). Within each list, the word initial or word final consonant is systematically varied to produce six rhyming items (e.g., "bat," "bad," "back," "bass," "ban," and "bath"). Stimuli consisted of 90 CVC words drawn from the MRT list, and their associated wav file recordings that were obtained from the PB/MRT Word Multi-Talker Speech Database in the Speech Research Laboratory at Indiana University Bloomington. Forty-two of the words were produced by a female talker, and the remaining forty-eight by a male talker.

2. Phonetically balanced words

The phonetically balanced (PB) corpus consisted of 20 lists of 50 monosyllabic words whose phonemic composition approximates the statistical occurrence in American English (e.g., "bought," "cloud," "wish," and "scythe") (Egan, 1948). Stimuli consisted of 90 unique words drawn from lists 1–3 of the PB corpus so that no overlaps occurred with those selected from the MRT corpus. Wav file recordings were obtained from the PB/MRT Word Multi-Talker Speech Database in the Speech Research Laboratory at Indiana University Bloomington. Half of the stimuli were produced by a male talker, and the other half by a female talker.

3. Harvard/IEEE sentences

The Harvard/IEEE sentence database consisted of 72 lists of 10 meaningful sentences (IEEE, 1969). These phonetically balanced (relative to American English) sentences contained five keywords embedded in a semantically rich meaningful sentence (e.g., "Her purse was full of useless trash," "The colt reared and threw the tall rider"). Stimuli consisted of 25 sentences drawn from lists 1–10 of the Harvard/IEEE sentence database and their associated wav file recordings obtained from the speech corpus originally created by Karl and Pisoni (1994) A female talker produced 14 sentences and a male talker produced the remaining 11. Selection of these two talkers was based on their production of speech that was highly intelligible (90% correct keyword accuracy across the 100 sentences) as demonstrated by previous research (Bradlow *et al.*, 1996).

4. Anomalous Harvard/IEEE sentences

Semantically anomalous sentences preserve the canonical syntactic structure of English, but have no meaning. Herman and Pisoni (2000) used the Harvard/IEEE sentence materials to create phonetically balanced semantically anomalous sentences. The keywords from the 100 sentences in lists 11–20 of the Harvard sentence corpus were coded according to semantic category (noun, verb, adjective, adverb) and replaced with words from equivalent semantic categories from lists 21–70 (Herman and Pisoni, 2000). This operation created sentences that have legal syntactic structure in American English, but were semantically anomalous (e.g., "Trout is straight and also writes brass," "The deep buckle walked the old crowd"), thus precluding subjects from using typical sentence context to identify the keywords. Stimuli consisted of 25 anomalous sentences drawn from the anomalous Harvard/IEEE sentence corpus of Herman and Pisoni (Herman and Pisoni, 2000) and their associated wav file recordings. A female talker produced 13 of the sentences, whereas a male talker produced the remaining 12 sentences.

5. Environmental sounds

The environmental sound database of Marcell and colleagues consists of acoustic signals recorded from a wide variety of acoustic environments developed for use in neuropsychological evaluation and confrontation naming studies (Marcell *et al.*, 2000). The corpus consists of 120 sounds from various acoustic events spanning a wide variety of categories: sounds produced by vehicles (e.g., automobile, airplane, motorcycle), animals (bird, dog, cow), insects (mosquito, crickets), nonspeech sounds produced by humans (snoring, crying, coughing), musical instruments (piano, trumpet, flute), tools (hammer, vehicles), liquids (water boiling, rain), among others. These sounds have been normed in a group of neurologically normal subjects on a variety of subjective (e.g., familiarity, complexity, pleasantness and duration) and perceptual measures (e.g., naming accuracy and naming response latency) (Marcell *et al.*, 2000). Stimuli consisted of ninety environmental sounds and their associated wav file recordings obtained from a digital database published by the authors on the Internet (http://ww.cofc.edu/~marcellm/confront.htm) (Marcell *et al.*, 2000). Stimulus selection from a variety of acoustic categories provided a wide representation of sound types and familiarity ratings.

C. Synthesis

Stimuli were processed using Tiger CIS (http://www.tigerspeech.com/) to simulate an eight-channel cochlear implant using the CIS processing strategy. Stimulus processing involved two phases, an analysis phase, which divides the signal into frequency bands and derives the amplitude envelope from each band, and a synthesis phase, which replaces the frequency content of each band with a sinusoid that is modulated with the appropriate amplitude envelope. Analysis used band-pass filters to divide the stimuli into eight spectral channels between 200 and 7000 Hz in steps with corner frequencies based on the Greenwood function (24 dB/octave slope). Envelope detection used a low pass filter with an upper cutoff at 400 Hz with a 24 dB/octave slope. Following the synthesis phase, the modulated sinusoids were combined and saved as 22 kHz 16 bit windows PCM wav files. Normalization of the wav files to a standard amplitude (65 dB rms) using a leveling program (LEVEL V2.0.3 Tice and Carrell, 1998) ensured that

TABLE I. Block design of the experiment identifying the materials presented during each phase. Although all subjects were presented with the same materials, the order in which they were presented varied by block according to the training condition to which subjects were assigned (MRT: modified rhyme test words; PB: phonetically balanced words; HS: Harvard/IEEE sentences: AS: anomalous sentences; and ENV: environmental sounds).

Training	Block 1 Pretest	Block 2 Training	Block 3 Gen. 1	Block 4 Posttest	Block 5 Gen. 2	Block 6 Gen. 3	Block 7 Gen. 4
MRT	MRT	MRT	AS	MRT	HS	PM	ENV
PB	PB	PB	AS	PB	HS	MRT	ENV
AS	AS	AS	MRT	AS	HS	PB	ENV
HS	HS	HS	MRT	HS	AS	PB	ENV
ENV	ENV	ENV	AS	ENV	MRT	HS	PB

stimuli were equal in intensity across all materials, and that no peak clipping occurred.

D. Procedures

All methods and materials were approved by the Human Subjects Committee and Institutional Review Board at Indiana University. All subjects indicated their informed consent before beginning the experiment. A short subject information form asked for basic background information and inquired as to any prior hearing, speech, or language problems.

Data collection used a custom script written for PSYSCRIPT, and implemented on four Apple PowerMac G4 (512 Mb RAM) computers and four 15-in. color Sony LCD monitors (1024×768 pixels, 75 Hz refresh). Audio signals were presented over four sets of Beyer Dynamic DT-100 headphones, calibrated with a voltmeter to a 1000 Hz tone at 70 dB SPL using a voltage/intensity conversion table for the headphones. Sound intensity was fixed within PsyScript in order to guarantee consistent sound presentation across subjects.

Multiple booths in the testing room accommodated up to four subjects at the same time. Subjects were informed that the stimuli they would hear were processed by a computer and that while they may have difficulty understanding them at first, they would quickly adapt and that the purpose of the present study was to train them to understand the stimuli. On-screen instructions preceded each block to orient the subject to the materials and requirements of the upcoming task. Before the presentation of each audio signal, a fixation cross appeared at the center of the screen for 500 ms alerted the subject as to the upcoming trial and was followed by the presentation of the stimulus. Following stimulus offset, a dialog box appeared on the screen prompting subjects to type in what they heard. There were no time limits for responding; subjects performed at their own pace and were encouraged to rest between each trial as needed. The experimental session lasted on average 45 min. All subjects received written and verbal debriefing after the experiment.

1. Training

Each training condition consisted of seven blocks. Stimuli were prerandomized and organized into separate lists for presentation in each training condition. Although the stimuli used in each block varied as a function of training materials, the same basic block design was consistent throughout all conditions (Table I).

In order to establish a baseline level of performance before training, a pretest was conducted in which subjects transcribed stimuli from the appropriate training category, but did not receive any feedback. During training, subjects received feedback in the form of the repetition of the processed auditory stimulus paired with the written form of the stimulus on the computer screen irrespective of whether their previous response was correct (the transcription of the word or sentence, or the descriptive label of the environmental sounds). During the posttest, subjects heard a selection of old materials from the pretest and training, as well as new materials from the same category. The posttest materials were selected to assess the effects of explicit training (using training materials), familiarity without explicit training (pretest materials) and novelty (previously unheard materials). The remaining three blocks assessed the generalization of training to other types of materials.

a. MRT word training. During the pretest, listeners were presented with 20 MRT words. Training consisted of 50 novel MRT words. An intervening generalization block occurred in block 3 to prevent habituation to the stimuli, and consisted of 25 anomalous sentences. The posttest in block 4 presented a total of 60 MRT words, 20 of which were drawn from the pretest materials, 20 from training, and 20 were novel stimuli with which subjects had no previous experience in the experiment. The remaining three blocks tested generalization to 25 meaningful sentences (block 5), 50 PB words (block 6), and 60 environmental sounds (block 7).

b. PB word training. PB training utilized an identical design to the MRT training, except the pretest, training and posttest materials consisted of PB words and block 6 tested generalization to 50 novel MRT words.

c. Anomalous sentence training. In order to allow for the relative effect of words transcribed across sentences, fewer sentences were selected. The pretest block consisted of four anomalous sentences (20 key words); the training block consisted of ten novel anomalous sentences (50 key words). Block 3 was an intervening block, consisting of 50 MRT words. The posttest in block 4 utilized 12 anomalous sentences, four selected from the pretest, four from the posttest, and four novel sentences (60 keywords). The remaining three blocks tested the effects of generalization to new materials. Block 5 consisted of 25 meaningful sentences, block 6 of 50 PB words and block 7 of 60 environmental sounds.

d. Harvard/IEEE sentence training. Harvard/IEEE sentence training utilized an identical design to the anomalous sentence training, except that the pretest, training and post-test materials consisted of meaningful sentences, and block 6 tested generalization to 25 anomalous sentences.

e. Environmental sound training. Like the MRT and PB training, training on environmental sounds began with a pretest consisting of 20 environmental sounds and training consisting of 50 novel environmental sounds. Subjects were asked to respond by describing what event or object produced the sound and were given a typical example that did not occur in the materials. Subjects were told that if they heard something that sounded like music they should try to indicate what musical instrument produced the sound, rather than simply identifying it as music. An intervening block occurred in block 3 in order to prevent habituation to the stimuli and consisted of 25 anomalous sentences. The posttest in block 4 presented a total of 60 environmental sounds, 20 of which were drawn from the pretest materials, 20 from training, and 20 were novel stimuli with which subjects had no previous experience in the experiment. The remaining three blocks consisted of generalization to 50 MRT words (block 5), 25 Harvard/IEEE sentences (block 6), and 50 novel PB words (block 7).

A separate group of 25 subjects served as controls in the environmental stimulus training condition. These subjects were exposed to the same processed training stimuli as the experimental group, but did not receive any feedback. Pre- to posttest comparisons were carried out to determine whether subjects showed any gains from being exposed to the stimuli and processing conditions during training, and were compared to subjects in the experimental group to assess the effectiveness of explicit training on environmental sounds. Pre- and posttest scores were also compared to environmental stimulus performance of the other training groups (HS, AS, MRT, and PB) to further assess whether training on speech provided any benefit on the recognition of environmental sounds.

2. Analysis and scoring

A supervised spellchecker in Microsoft EXCEL was used to correct the more obvious spelling errors and standardized responses across subjects by recoding homophones into a standard spelling. An automated macro searched for target/response matches using a preordained target list, the result of which was hand checked by a research assistant. Responses that were morphologically related to the target (pluralization of nouns or conjugation of verbs) were scored as incorrect. PB and MRT words were scored based on whether the entire word was correct, whereas the anomalous and meaningful sentences were scored for keywords correct (five keywords per sentence).

Environmental sounds were checked using a similar procedure. Scoring rules were modified slightly from those originally used by Marcell and colleagues (2000) given the nature of the signal degradation. Responses were scored as correct if the subject identified the target agent (e.g., cow), the sound the agent made if it did not have multiple possible agents (e.g., moo), or the linking of the two (e.g., cow mooing) (Gaver, 1993). Failure to specify agent, or incorrectly specifying agent was scored as an incorrect response. Correct identification of musical instruments required accurate identification of the instrument. The generic response of "music" was scored as incorrect, given that the instructions explicitly told subjects that this was not a valid response option.

For each training condition, responses were averaged across subjects for each block. Within-subjects analyses compared performance across blocks of a given training condition. Paired samples t-tests were used to assess the effects of training by comparing pre- and posttest performance. Posttest scores were balanced by only including the responses to the materials on which subjects were not explicitly trained, to avoid biasing the findings: none of the posttest values reported in the text, tables, or figures contains responses to the stimuli used during training. The differences in performance on the source of the posttest materials (items from pretest, training, and novel lists) were assessed with a one-way analysis of variance (ANOVA) and post hoc Tukey tests which were corrected for multiple comparisons using the Bonferroni correction. Other paired t-tests were conducted to assess the effects of context (anomalous sentences versus Harvard/IEEE sentences) and complexity (PB words versus MRT words). A correlational analysis examined the relationship between performance across blocks to assess whether performance on one set of materials was correlated with performance on another. Between subjects comparisons assessed the effect of training on materials across training conditions using one-way ANOVA and post-hoc Tukey tests.

III. RESULTS

A. Pre-/posttest comparisons

All types of training produced significant pre- to posttest gains in performance (Table II). Moreover, all speech materials showed a benefit from training that was greater than 10%, indicating that explicit training produced changes that exceeded the effects of mere exposure reported by Davis and colleagues (2005).

For the MRT stimuli, subjects increased significantly from 5.8% correct at pretest to 37.5% after training [$t(1,24)=13.576$, $p<0.001$]. A one-way ANOVA comparing the source of the posttest materials demonstrated that subjects performed significantly better on materials on which they were explicitly trained (58% correct) than on materials with which they were familiarized but not explicitly trained (40.4% correct) and novel materials (34.6% correct) [$F(2,72)=18.967$, $p<0.001$]. Post-hoc Tukey tests revealed that subjects performed significantly better on posttest stimuli drawn from the training list (both $p<0.001$) demonstrating a significant effect of feedback, and indicating good retention of training by subjects. However, subject performance did not differ on post-test materials drawn from the pretest and novel lists ($p=0.313$).

Subjects trained on the PB words started out better than those subjects trained on the MRT words, also showing significant increases from pre- (23.4% correct) to posttest (46.2% correct) [$t(1,24)=7.134$, $p<0.001$]. Examination of the posttest materials revealed that subjects performed best on stimuli on which they were explicitly trained (55.4% correct), followed by novel PB words (48.20% correct) and words on which they were previously exposed, but not ex-

TABLE II. Performance at pretest and at posttest across training groups. The individual lists used in the posttest are decomposed into materials from the pretest list, training list and novel list. The final posttest score does not contain the scores from the training list to avoid a confound with feedback.

Training	Pretest	Pretest list	Training list	Novel list	Posttest
MRT	M=0.06	M=0.40	M=0.58	M=0.35	M=0.37
	S.D.=0.06	S.D.=0.13	S.D.=0.16	S.D.=0.12	S.D.=0.11
PB	M=0.23	M=0.44	M=0.55	M=0.48	M=0.46
	S.D.=0.11	S.D.=0.09	S.D.=0.15	S.D.=0.12	S.D.=0.09
AS	M=0.34	M=0.62	M=0.78	M=0.62	M=0.62
	S.D.=0.14	S.D.=0.12	S.D.=0.13	S.D.=0.13	S.D.=0.10
HS	M=0.40	M=0.72	M=0.97	M=0.56	M=0.64
	S.D.=0.20	S.D.=0.10	S.D.=0.07	S.D.=0.12	S.D.=0.08
ENV	M=0.38	M=0.40	M=0.50	M=0.53	M=0.46
	S.D.=0.10	S.D.=0.14	S.D.=0.11	S.D.=0.11	S.D.=0.11

plicitly trained (44.20% correct) [$F(2,72)=5.484$, $p=0.006$]. Post-hoc Tukey tests revealed that subjects performed significantly better on the post-test materials drawn from the training list ($p=0.005$) than on materials from the pre-test list, but no difference was observed for materials drawn from the novel list ($p=0.097$). More importantly, subject performance did not differ between materials drawn from the pretest and novel list ($p=0.477$).

For the anomalous sentences, performance was good at pretest (33.6% correct) but increased significantly following training [61.7% correct, $t(1,24)=11.713$, $p<0.001$]. Examination of the posttest materials revealed a significant main effect of source [$F(2,72)=14.115$, $p<0.001$], and post-hoc Tukey tests confirmed that subjects performed significantly better on the posttest materials drawn from the training list (78.2% correct) than on materials from either the pretest (61.6% correct, $p<0.001$) or novel list (61.8% correct, $p<0.001$). No differences in performance were observed on the materials from the pre-test and novel lists ($p=0.998$).

Performance on the Harvard/IEEE sentences also increased significantly from pre- (40% correct) to posttest [63.9% correct, $t(1,24)=7.041$, $p<0.001$]. A one way ANOVA revealed a significant main effect of source [$F(2,72)=114.043$, $p<0.001$] indicating that subjects performed significantly better on materials from the training list (97% correct) than on those from the pretest (71.6% correct) and novel (56.2% correct) lists (all $p<0.001$). Subjects also performed significantly better on the post-test materials drawn from the pre-test list as compared to the novel list ($p<0.001$). This is likely due to the high contextual salience of the sentences, because this pattern was not observed for the anomalous sentence training group.

Performance on the environmental sounds also showed a significant benefit from explicit training (Fig. 1). Subjects in the experimental group showed significant improvement between pre- (38.2% correct) and posttest [46.4% correct, $t(1,24)=2.804$, $p=0.01$]. An analysis of the posttest materials revealed a significant main effect of source [$F(2,72)=8.717$, $p<0.001$]. Subjects performed best on stimuli from the novel list (53.2% correct), followed by materials from the training list (50% correct) and pretest (39.6% correct). Subjects performed significantly better on posttest materials from both novel and training lists than on materials from the pretest list ($p=0.009$ and $p<0.001$, respectively) but did not differ from one another ($p=0.617$).

Subjects in the environmental sound control group did not show improvement between pre- (39% correct) and posttest [37% correct, $t(1,24)=1.056$, $p=0.301$]. A univariate ANOVA comparing performance on the pre- and posttest blocks across the control and experimental groups revealed a significant main effect of training [$F(3,96)=5.154$, $p=0.002$]. Subjects in both groups performed equally well during the pretest phases ($p=0.991$), but at posttest subjects in the experimental group (46.4% correct) performed significantly better than subjects in the control group (37% correct, $p=0.003$). These data indicate that subjects who received explicit training on environmental sounds show gains in performance above and beyond the gains obtained from merely being exposed to the stimuli.

B. Correlations

Correlations of the performance across blocks revealed several interesting results (Table III). In general, the specific

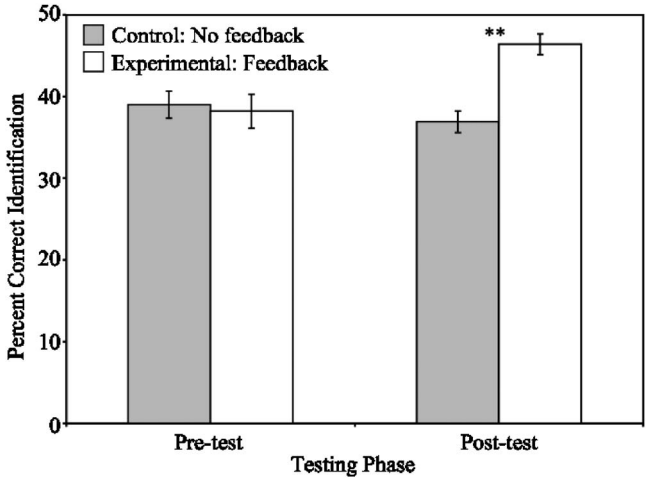


FIG. 1. Bar graph displaying perceptual accuracy scores at identifying environmental sounds across the experimental and control groups. The type of training that a subject received is indicated on the x axis and coded by colored bars (control: gray; experimental: white). Posttest scores only contain the responses to stimuli on which subjects did not receive explicit training (see the text). Asterisks indicate statistically significant differences between groups on pre- or posttest performance ($p<0.01$).

TABLE III. Correlations between the various stimuli presented at posttest and during generalization for each training group (see Table I for abbreviations). Rows are blocked by training condition and the specific materials used during posttest are indicated by italicized font (posttest values only contain the subset of materials on which subjects did not receive explicit feedback). Values along the diagonal indicate the percent correct recognition scores for the stimuli. Only statistically significant correlations are displayed ($^{*}p<0.05$; $^{**}p<0.01$, $^{***}p<0.001$).

Training		MRT	PB	AS	HS	ENV
MRT	*MRT*	37%	$r=0.77^{**}$	$r=0.57^{*}$	$r=0.67^{**}$	n.s.
	PB	⋯	43%	$r=0.55^{*}$	$r=0.65^{**}$	n.s.
	AS	⋯	⋯	48%	$r=0.90^{**}$	n.s.
	HS	⋯	⋯	⋯	67%	n.s.
	ENV	⋯	⋯	⋯	⋯	38%
PB	MRT	43%	$r=0.66^{***}$	$r=0.51^{**}$	n.s.	$r=0.55^{**}$
	PB	⋯	46%	$r=0.53^{**}$	n.s.	n.s.
	AS	⋯	⋯	48%	n.s.	n.s.
	HS	⋯	⋯	⋯	66%	n.s.
	ENV	⋯	⋯	⋯	⋯	35%
AS	MRT	31%	n.s.	n.s.	n.s.	n.s.
	PB	⋯	44%	n.s.	n.s.	$r=0.45^{*}$
	AS	⋯	⋯	62%	$r=0.63^{***}$	n.s.
	HS	⋯	⋯	⋯	67%	n.s.
	ENV	⋯	⋯	⋯	⋯	35%
HS	MRT	27%	$r=0.59^{**}$	$r=0.62^{***}$	$r=0.40^{*}$	n.s.
	PB	⋯	44%	$r=0.73^{***}$	$r=0.51^{**}$	$r=0.57^{**}$
	AS	⋯	⋯	59%	$r=0.55^{**}$	n.s.
	HS	⋯	⋯	⋯	64%	n.s.
	ENV	⋯	⋯	⋯	⋯	39%
ENV	MRT	33%	n.s.	n.s.	n.s.	n.s.
	PB	⋯	43%	n.s.	n.s.	n.s.
	AS	⋯	⋯	49%	$r=0.69^{***}$	n.s.
	HS	⋯	⋯	⋯	68%	n.s.
	ENV	⋯	⋯	⋯	⋯	46%

type of materials that subjects received in training was most strongly correlated with materials of a similar class (words with words, sentences with sentences).

For the MRT words, performance at posttest was most strongly correlated with performance on the PB words ($r=0.766$, $p<0.001$) followed by Harvard/IEEE sentences, and anomalous sentences, but not environmental sounds. Moreover, similar relationships were observed for the PB words and meaningful and anomalous sentences. It is interesting to note that performance on isolated words was most strongly correlated with performance on other words, followed by meaningful and anomalous sentences, and that sentences were most strongly correlated with other sentences.

Performance on the PB word posttest was most strongly correlated with performance on the MRT words followed by Harvard/IEEE sentences, but was not correlated with anomalous sentences or environmental sounds. MRT performance was also correlated with performance on anomalous sentences and environmental sounds. As observed in the MRT training group, performance on words (PB or MRT) was most strongly correlated with performance on other words and performance on sentences was most strongly correlated with performance on other sentences.

Performance on the anomalous sentence posttest was only correlated with performance on Harvard/IEEE sentences ($r=0.63$, $p=0.001$). The only other significant correlation observed was between PB words and environmental sounds ($r=0.446$, $p=0.025$). All other correlations were not significant.

Performance on the Harvard/IEEE sentence posttest was most strongly correlated with performance on anomalous sentences, followed by PB and MRT words. Anomalous sentences were most strongly correlated with performance on PB and MRT words.

Performance on the environmental stimulus post-test was not significantly correlated with any of the other materials, but as observed earlier, Harvard/IEEE sentences were significantly correlated with anomalous sentences.

C. Across Group Comparisons

To assess the effect of training on the source materials, the recognition accuracy scores for a given set of materials were compared across training conditions (Fig. 2). Comparison with the pretest data assessed whether the type of training a subject received had an effect on performance. Comparison with the posttest scores assessed whether the type of training significantly affected performance, and whether training on a specific set of materials produces better and more robust generalization than another.

Examination of the performance on the MRT words across training materials [Fig. 2(a)] with a one-way ANOVA revealed a significant main effect of training materials [$F(5,144)=37.495$, $p<0.001$]. Post-hoc Tukey tests revealed that subjects performed significantly better than the pretest regardless of the type of material that they were trained upon (all $p<0.001$). This is not surprising, given the poor baseline performance (5.8% correct). Although any type

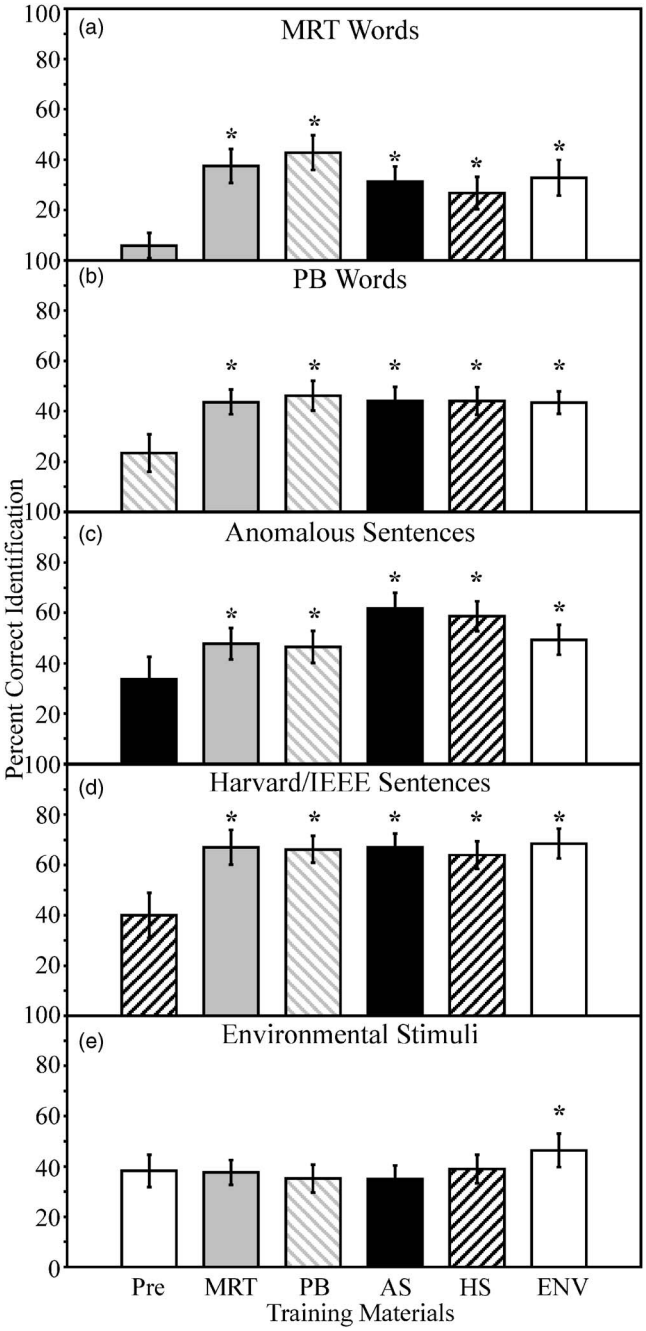


FIG. 2. Bar graph displaying perceptual accuracy scores at identifying MRT words (A), PB words (B), anomalous sentences (C), Harvard/IEEE sentences (D), and environmental sounds (E) as a function of training. The type of training that a subject received is indicated on the *x* axis and coded by colored bars (MRT: gray; PB: gray striped; AS: black; HS: black striped; ENV: white). Pretest scores correspond to the baseline for each type of stimuli. Posttest scores only contain the responses to stimuli on which subjects did not receive explicit training (see the text). Asterisks indicate statistically significant differences from the baseline for each group ($p<0.05$).

of training produced a benefit, subjects trained on single words regardless of their origin (MRT and PB) performed identically ($p=0.477$) and significantly better than subjects trained on any other materials (all $p<0.01$). Training on anomalous sentences, Harvard/IEEE sentences and environmental sounds also produced significant gains over baseline, performing similarly on the MRT generalization test (all $p>0.319$). When the scores were grouped by material type, however, subjects who received training on words (MRT and PB) performed significantly better than subjects trained on sentences ($p<0.001$) or environmental sounds ($p=0.027$).

Training also produced a significant main effect on performance on the PB materials [$F(5,144)=24.86$, $p<0.001$; Fig. 2(b)]. Overall, it did not matter what type of training subjects received, as performance was significantly higher than pretest for all training conditions (MRT training 43.44% correct $p<0.001$, AS training 44.08% correct $p<0.001$, HS training 44.08% correct $p<0.001$, ENV training 43.68% correct $p<0.001$). The main effect for training condition is carried entirely by the gains in performance relative to the pretest, because no significant differences were observed between performance across the five training conditions (all $p>0.867$). This pattern indicates that when identifying words that are highly discriminable, training with any type of material will provide an equivalent benefit.

A one-way ANOVA on the anomalous sentences [Fig. 2(c)] revealed a significant main effect of training [$F(5,144)=22.986$, $p<0.001$]. Comparison with the pretest revealed that all types of training produced significant increases in performance relative to the baseline (33.6% correct, all $p<0.001$). Subjects who were trained the anomalous sentences (61.7% correct) performed as well as those who were trained on the meaningful Harvard/IEEE sentences (58.62% correct, $p=0.902$) and significantly better than those trained on PB, MRT, and environmental sounds (all $p<0.004$) sentences. Training on MRT (47.74% correct), PB (46.53% correct) and environmental sounds (47.74% correct) provided equivalent benefit when recognizing the anomalous sentences, however (all $p>0.0998$).

A significant main effect of training condition was also observed for the Harvard/IEEE sentences [$F(5,144)=22.444$, $p<0.001$; Fig. 2(d)]. The comparison of each of the training conditions to the Harvard/IEEE sentence pretest revealed that all subjects performed significantly better than the baseline (40% correct) regardless of the type of training they received (all $p<0.001$). As was the case for the PB materials, the training effect was carried entirely by the gains in performance relative to the pretest, as there were no significant differences between performance across the five training conditions (MRT 67.01% correct, PB 66.21% correct, HS 63.90% correct, AS 67.04% correct, ENV 68.51% correct, all $p>0.719$).

The effect of training on the recognition of the environmental sounds showed a different pattern of results [Fig. 2(e)]. A one-way ANOVA revealed a significant main effect of training group on performance [$F(7,192)=4.452$, $p<0.001$], however unlike the training effects observed for the other stimulus materials, subjects only showed gains relative to baseline (38.2% correct) when they were explicitly trained on the environmental sounds (46.40% correct, $p=0.013$). Subjects trained on all other materials failed to show any differences as compared to baseline (MRT 37.6% correct $p=1.00$; PB 35.2% correct $p=0.822$; HS 34.93% correct $p=0.999$; AS 34.93% correct $p=0.764$). Pre- (39%) and posttest (37%) performance of subjects in the control group did not differ from subjects in the AS, HS, PB, or MRT training groups (all $p>0.7$) indicating that subjects who were exposed to but not explicitly trained on the environ-

mental sounds perform as well as subjects who are trained on speech. In addition, explicit training on environmental sounds produced significantly better posttest performance compared to all other groups (all $p<0.03$), who did not differ from one another (all $p>0.557$). That these scores did not differ from the baseline (for either the experimental or control groups), however, suggests that training on the speech materials was equally ineffective when transferring to environmental sounds. In effect, when asked to identify environmental sounds, training with speech materials is as effective as not receiving any training at all.

One potential concern with the current study is the use of a blocked design, which could lead to order effects at testing. Although all groups performed the posttest in block 4, the generalization materials were presented in different blocks to each group due to the block design. To assess the potential order effect a univariate ANOVA was conducted on each type of material using presentation order as the between subjects variable and training as the covariate. Although an order effect was observed for the MRT words [$F(2,122)=10.185$, $p<0.001$], post-hoc Tukey tests revealed that the only differences between blocks was between posttest (block 4) and when the MRT stimuli were presented in block 3 (Harvard/IEEE and anomalous sentence training, $p=0.002$). This result may be more apparent than real, because performance on MRT words in block 3 did not differ from performance in blocks 5 or 6 ($p=0.122$, $p=1.00$) respectively, suggesting that additional practice effects may not have significantly influenced performance on the MRT words during generalization.

For the rest of the generalization data, order effects were either not observed, or could be accounted for entirely by training. For PB words, tests of generalization always occurred in block 6, and no significant effect of order was observed [$F(2,122)=0.975$, $p=0.325$]. Similar results were observed for Harvard/IEEE sentences [$F(2,122)=1.003$, $p=0.37$] in which generalization was tested in blocks 5 or 6. For anomalous sentences, generalization always occurred in block 5, and a significant main effect of order was observed [$F(2,122)=22.77$, $p<0.001$]. A post-hoc Tukey test revealed that this effect was entirely confounded with training, because subjects performed significantly better in the posttest (block 4) than during generalization (block 5, $p<0.001$). Generalization to environmental sounds also showed a significant main effect of order [$F(2,122)=24.650$, $p<0.001$], but the effect was confounded with training, as post-test performance in block 4 was significantly higher than generalization in block 7 ($p<0.001$). Thus for all data order effects were not apparent, or were completely accounted for by training. For MRT words a potential order effect was indicated, but since performance did not differ between early and late blocks, this effect may be more apparent than real.

The specific gains from training are displayed in Fig. 3. Gain scores were computed by subtracting performance at pretest from performance during generalization and posttest for all materials. Positive gain scores indicate improvement after training. Overall, the largest gains from training were observed for the MRT words [Fig. 3(a)], for which training on either MRT words or PB words produced significantly higher gains from training than either type of sentences or environmental sounds. A similar result was observed for anomalous sentences [Fig. 3(c)], where training on either anomalous or meaningful sentences produced significantly higher gains than for words or environmental sounds. Gains for PB words [Fig. 3(b)] and meaningful Harvard/IEEE sentences [Fig. 3(d)] were uniform across all training conditions. Finally, training on environmental sounds [Fig. 3(e)] was the only form of training that produced a significant benefit in the identification of environmental sounds.

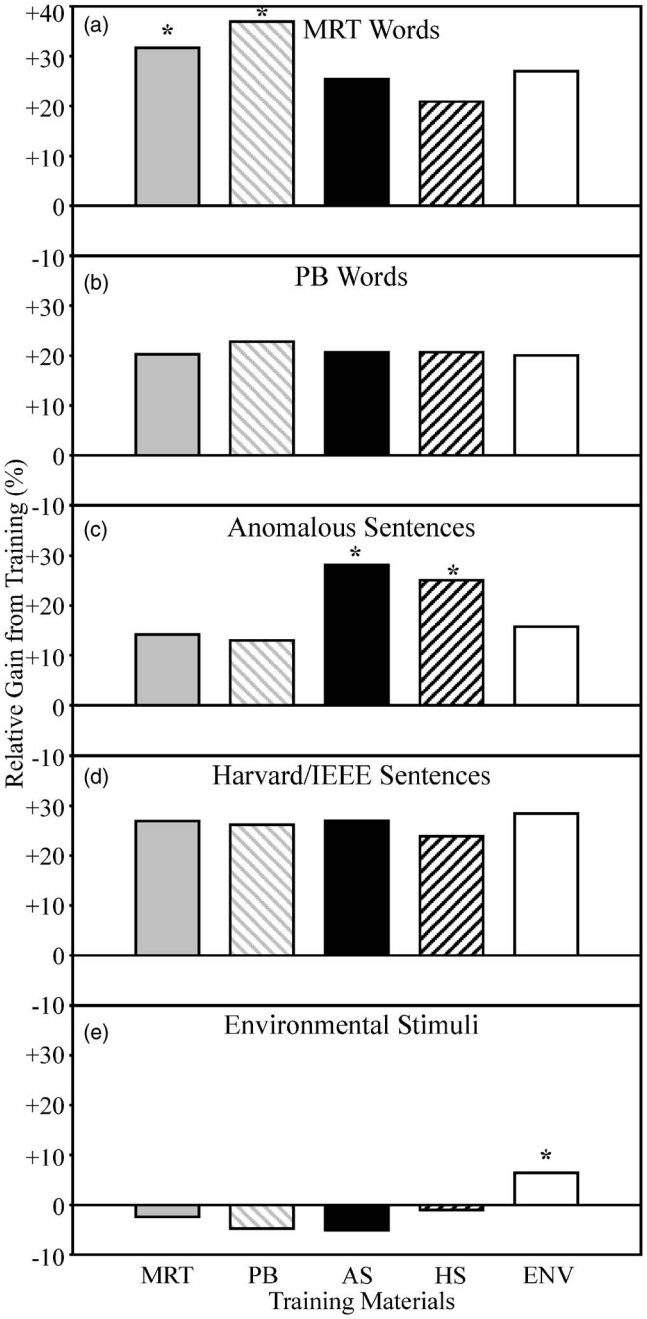


FIG. 3. Bar graph displaying the relative gains from training for the MRT words (A), PB words (B), anomalous sentences (C), Harvard/IEEE sentences (D), and environmental sounds (E). The type of training that a subject received is indicated along the x axis and coded by color (MRT: gray; PB: gray striped; AS: black; HS: black striped; ENV: white). Gain scores were computed by subtracting the posttest or generalization scores from the scores at pretest. Asterisks indicate statistically significant differences from the other groups ($p<0.05$).

IV. DISCUSSION

Taken together, our results showed that the specific type of stimulus materials used during training had a significant effect on perceptual learning. Subjects showed significant pre- to posttest improvement in all training conditions, demonstrating that they were able to make effective use of feedback to improve their performance. Generalization effects, however, were not uniform across materials. For some training materials, subjects showed encoding specificity, performing best on the materials on which they were explicitly trained (e.g., Tulving and Thomson, 1973). Subjects who were trained on isolated words (PB or MRT) performed significantly better when identifying MRT stimuli than the other groups. When tested on PB words, all subjects regardless of training, performed equally well. Subjects who were trained on sentences (anomalous or meaningful) performed significantly better when identifying anomalous sentences than the other groups. When tested on Harvard/IEEE sentences, all subjects performed equally well regardless of the materials on which they were trained. The differences in training specificity suggest that differences exist in the difficulty of the materials, with training generalizing more readily to easier tasks (as evidenced by the nonspecific effect of training on generalization to PB words and Harvard/IEEE sentences), but not to more difficult tasks (as evidenced by training specificity that was observed for MRT words, anomalous sentences and environmental sounds).

Although MRT and PB words have similar word frequencies in American English, they come from lexical neighborhoods with different densities. Neighborhood density determines how many confusable "neighbors" a given word has based on phonological similarity (how many different words can be formed by changing a single phone: beach, peach, teach, breach, bleach, etc.) (Luce and Pisoni, 1998). Words from sparse neighborhoods have fewer confusable lexical alternatives and are recognized faster and with higher accuracy than words from more dense neighborhoods (Luce and Pisoni, 1998). The neighborhood density and lexical frequency for the PB, MRT, and meaningful and anomalous sentence keywords were obtained from the Neighborhood Activation Model database (http://neighborhoodsearch.wustl.edu/neighborhood/Home.asp). PB words had a lower average neighborhood density (12.25) compared to MRT words (21.80) suggesting that PB words should be more discriminable and identifiable than MRT words. As anomalous sentences contain keywords that were initially drawn from the Harvard/IEEE sentences they are equal in lexical frequency and neighborhood density (10.83 and 10.72, respectively). The main difference between these sentences is semantic predictability. The differences in the predictability and confusability of the speech stimuli likely produced differences in task difficulty. When task demands were high (due to lower predictability or higher confusability), subjects performed better when they were trained on stimuli of the same general class (e.g., training on words generalized significantly better to other words, sentences generalized significantly better to other sentences), demonstrating transfer appropriate processing (e.g., Morris, *et al.*, 1977). The opposite effect was observed for the more predictable materials: subject performance did not differ across training groups on the PB words and Harvard/IEEE sentences. This suggests that when the task demands are less difficult, such as when identifying words from lower density neighborhoods or highly predictable sentences, all forms of training are equivalent.

It is interesting to note here that training on PB words had a larger effect on the recognition of MRT words ($M=47\%$) than on the PB words themselves ($M=43\%$) [Fig. 3(a) compared to Fig. 3(b)]. This is likely due to the poorer performance during the pretest for MRT words ($M=6\%$) relative to PB words ($M=23\%$): subjects had more of an opportunity to improve their performance on the MRT words even though they were being trained on PB words. The differences in the performance at pre-test likely stem from the differences in task difficulty: MRT words are composed of syllables that differ only on word initial or word final consonants and are from higher density lexical neighborhoods, whereas PB words are phonetically balanced for phoneme occurrence in American English and come from lower density lexical neighborhoods. For the MRT words, all groups did improve significantly relative to the low baseline, but training on MRT and PB words had an additional effect, improving performance significantly more than either of the sentence training conditions or the environmental sound training condition. Therefore, the low baseline for the MRT words at pretest cannot account for the differences in performance, and the higher performance of subjects trained on MRT or PB words must be due to training.

Although training on words and environmental sounds generalized to anomalous sentences, it is unknown how much exposure to anomalous sentences without explicit training or feedback would improve performance. Based on the previous research of Davis and colleagues (2005), exposure to vocoded meaningful sentences without feedback only produced gains in performance of 11%. It is likely that much of this effect is driven by the semantic predictability of the materials, and that improvement from exposure to anomalous sentences in which sentence context cannot be used to predict upcoming words, would be much lower. Additionally, Burkholder (2005) demonstrated that giving subjects unprocessed auditory feedback during training on vocoded anomalous sentences improved performance across three phases of training by 10%. Given that these subjects were provided with feedback, we would expect the gains from exposure to anomalous sentences to be far less than 10%. In the present study, subjects trained on MRT and PB words and environmental sounds showed improvement on the recognition of anomalous sentences that exceeded 10% over the anomalous sentence pre-test (gains of 14%, 13%, and 16%, respectively), suggesting that such training improved performance on anomalous sentences that was greater than would be expected from exposure alone. Future work, however, will have to determine how much of an effect exposure has on the recognition of vocoded anomalous sentences.

Overall, the specific type of materials upon which subjects were trained was most strongly correlated with materials of a similar class. For both MRT and PB training, posttest

performance was most strongly correlated with performance on the PB/MRT generalization blocks, respectively. For anomalous and Harvard/IEEE training groups, posttest performance was most strongly correlated with performance on the Harvard/IEEE and anomalous sentence generalization blocks, respectively. These findings provide further support for encoding specificity and transfer appropriate processing. Except for three instances, performance on speech stimuli was not correlated with performance on environmental sounds. Performance on MRT and PB materials was significantly correlated with performance on the environmental sounds for the PB training, anomalous and Harvard/IEEE training, suggesting some bidirectionality to training on speech. Interestingly, environmental stimulus training was not correlated with performance on any tests of speech generalization.

One intriguing finding from this study was the clear asymmetry in training that was observed for the environmental sounds (Fig. 3). Subjects explicitly trained on environmental sounds performed significantly better than baseline on all speech materials, suggesting that training on complex nonspeech stimuli produces robust generalization to speech. The inverse, however, was never observed: training on speech consistently failed to produce performance that differed from the environmental baseline. Compared to the control group, explicit training had a significant effect on posttest recognition of environmental sounds, indicating that explicit training generalized to new environmental stimuli that subjects had not heard previously in the experiment. In addition, subjects who were trained on speech did not differ from subjects in the control group who were exposed to the environmental sounds but did not receive explicit feedback or training. Thus, it appears that explicit training on complex non-speech materials leads to improved performance on speech materials, but training on speech materials does not produce gains in the perception of complex non-speech abilities. This suggests that explicit environmental sound training may have increased subjects' attentional sensitivity to the spectrotemporal characteristics of the stimuli, which may have enhanced their utilization of similar information for the identification of speech. However, training on speech may predispose the listener to make lexical judgments about the signal, placing them in a processing mode that focuses on stimulus meaning, a higher order cognitive variable, and reducing their attention to the lower order acoustic features. Further research will be necessary to determine the extent to which such bidirectionality exists.

Recent neuroimaging studies investigating the encoding of environmental sounds have suggested that similar cortical regions may be involved during the processing of environmental sounds and speech sounds (Lewis *et al.*, 2004). These cortical regions include canonical auditory areas required for the recognition of sound (primary auditory cortex), the identification of auditory speech stimuli (superior temporal gyrus, posterior superior temporal sulcus), semantic processing and accessing of lexical information during sound, picture and action naming (posterior medial temporal gyrus, pMTG) (Lewis *et al.*, 2004). These cortical areas (the pMTG and pSTS in particular) showed bilateral activation in response to environmental sounds, but tend to be left lateralized during speech perception tasks (Lewis *et al.*, 2004). This difference may partially explain the asymmetry that we observed for training with environmental sounds and speech. Perhaps training with environmental sounds activated cortical regions implicated in the processing of speech stimuli, leading to efficient generalization to speech. Due to different task demands, training with speech may have utilized additional lateralized cortical regions that would not necessarily facilitate generalization to environmental sounds.

Additionally, other recent neuroimaging studies have demonstrated that the functional connectivity between cortical regions may be differentially altered due to task demands when identifying speech (Obleser *et al.*, 2007). This may facilitate generalization in one case (environmental sounds to speech), but not the other (speech to environmental sounds). Taken together with the recent findings of Kidd and colleagues (2007), a general auditory ability for the recognition of familiar sounds may contribute to the identification of both speech and environmental sounds, providing further support for using general auditory training to enhance the perceptual abilities of CI users.

The present findings are similar to those of Gygi and colleagues, who found that the most important information for recognition of environmental sounds occupies an identical frequency range as that for speech (Gygi *et al.*, 2004). If the important information for environmental sounds overlaps substantially with speech, then training subjects to better utilize the spectro-temporal information in this frequency region more efficiently should improve generalization to speech, as we report here. Training on speech alone may not be sufficient to increase generalization to environmental sounds, because subjects may focus their attention on the higher order lexical and semantic attributes of the signal rather than the spectrotemporal information itself. The finding that training on speech does not generalize to environmental sounds conflicts with the earlier findings of Burkholder (2005), who reported that training on speech did generalize to environmental sounds. However, Burkholder did not use a pre-/posttest design, so the baseline performance levels for environmental sounds were not known. Although training on speech materials in the present study led to performance levels for environmental sounds that were greater than zero, they did not exceed the pretraining baseline. This result suggests that subjects in the Burkholder study (2005) may not have performed any differently after training than subjects who were naïve to the stimulus processing conditions.

Surprenant and Watson (2001), and more recently Kidd and colleagues (2007) reported a significant correlation between subject's ability to discriminate nonspeech stimuli based on spectrotemporal cues and their identification of speech in noise. The authors suggested that common higher order acoustic processes may contribute to both speech and nonspeech processing capabilities. These differences could account for the substantial differences in performance of subjects who receive hearing aids, and cochlear implants alike: auditory sensitivity at a peripheral level may not be the sole cause of variability in outcome and benefit; rather the inabil-

ity to utilize and manipulate auditory information at higher levels may supersede the benefits of an auditory prosthesis (Surprenant and Watson, 2001). This relationship may not be completely bidirectional, however, given our findings that training on environmental sounds generalizes to speech, but training on speech does not generalize to environmental sounds.

One important methodological difference between the results of the present study and earlier investigations using environmental sounds is the use of open-set testing procedures in all conditions. Gygi and colleagues reported identification scores of up to 66% correct using 6-channel noise vocoded stimuli (Gygi *et al.*, 2004), and Shafiro found that performance reaches asymptote with 16 channels (66%), but large stimulus specific effects were observed (Shafiro, 2004). Moreover, Reed and Delhorne (2005) found that CI users show higher levels of performance (79% correct). One commonality between all three earlier studies is that testing used closed-set forced-choice procedures. Under open-set testing, average performance after training (46% correct) was substantially lower than the performance observed in the previous studies. Given that the closed-set procedures necessarily limit subjects to a certain set of responses, open-set testing allows subjects to record their actual impressions of the stimuli in a way that would be more appropriate to real world listening environments (see Clopper *et al.*, 2006).

If subjects are indeed learning to utilize the residual spectrotemporal information more efficiently during training on environmental sounds, we should expect to see benefits for CI users as well. One problem, however, is that not all environmental sounds are perceptually equivalent. As Shafiro (2004) and Burkholder (2005) noted, some environmental sounds may be inherently more identifiable than others. The trading relations between the number of spectral bands needed for environmental sound recognition found by Shafiro (2004) suggests that some stimuli may not be readily identifiable when processed by a vocoder. Given that the amount of acoustic information differs across acoustic environments and task demands, the spectral resolution of the current generation of cochlear implants may be insufficient to provide significant benefit under all listening situations (Shannon *et al.*, 2004; Shannon, 2005). Training CI users to better make perceptual distinctions based on the acoustic information that they do have may provide additional benefit when listening to nonspeech sounds or identifying speech in noise.

Further, an important theoretical question is also raised here. Although several previous studies have failed to show substantive differences for the perception of speech as processed by a noise and sinewave vocoder (Dorman *et al.*, 1997), other studies have found that for non-speech tasks, performance is actually better for sinewave vocoded speech (Gonzales and Oliver, 2005). Gender and talker identification were significantly better for stimuli processed using a sinewave vocoder than when processed using a noise vocoder (Gonzales and Oliver, 2005). The authors suggest that the sinewave carriers may have introduced less distortion, thus preserving more accurate and robust detail in the amplitude envelopes that could be useful to the listener. A comparison of the two methods revealed more residual periodic information in the sinewave vocoder processed signal as compared to the noise vocoder processed signal, forming the basis for their claim (Gonzales and Oliver, 2005). It may be the case that a sinewave vocoder may produce better, more robust results for studies using music and environmental sounds than would a noise vocoder: for stimuli that carry more salient spectral information, less distortion and better preserved periodicities in the envelope may translate to heightened recognition. Whether performance on these types of stimuli differ from performance of cochlear implant users remains an open question.

Our findings also replicate and extend the recent studies conducted by Davis and colleagues (2005) and Burkholder (2005). Training using orthographic feedback paired with a repetition of the processed version of the sentence produced keyword identification scores (71% correct) that were nearly identical to those reported by Davis in the last block of training (75% correct). We also found that training on anomalous sentences produced excellent generalization to meaningful sentences, as was reported previously by both Davis *et al.* (2005) and Burkholder (2005). Moreover, the gains from training exceeded 11%, which was the benefit observed by Davis and colleagues (2005) for exposure to the stimuli without feedback. Thus, access to syntactic structure without relying on sentence context enhances sentence recognition. Our extension to include single PB words and CVCs also provides additional support for this conclusion: training on all materials produced excellent generalization to the meaningful Harvard/IEEE sentences. The results observed for training on environmental sounds, however, suggest that learning to recognize the acoustic form of a stimulus enhanced selective attention to spectro-temporal information, and bottom-up perceptual encoding processes.

We also replicated the findings of Fu and colleagues, who showed that giving CI users explicit training on CV and CVCs does indeed produce gains in sentence intelligibility (Fu *et al.*, 2006). The similar patterns of performance observed with normal hearing subjects listening to acoustic simulations of a cochlear implant provides further support for the utility of the vocoder as an effective model of electric hearing. By studying the perceptual learning of CI simulated speech in normal hearing listeners we can simultaneously learn about the neural and cognitive mechanisms that underlie speech and language processing in general, and expand our knowledge about effective rehabilitation and training programs to assist newly implanted individuals. By formalizing training paradigms that utilize a wide variety of stimulus materials, we may be able to provide cochlear implant users with tools that will bootstrap onto a variety of tasks and difficult listening conditions above and beyond those on which they were trained (i.e., increase "carry-over" effects). Given the substantial variability in performance among cochlear implant users that cannot be attributed to individual differences in etiology and duration of deafness, the question remains as to how differences in postimplantation experience contribute to outcome and benefit. Providing explicit instruction as to the important information in the signal may help to account for a portion of this variability, thereby allowing us

to disentangle the role of experience and provide a more objective assessment of cochlear implant user success.

In summary, we demonstrated that the type of stimulus materials used during training affects generalization to new materials. Although all forms of training provided some benefit, generalization of training was not uniform, and was highly context and task specific. When the task was easy, such as was the case when identifying contextually rich meaningful sentences or highly discriminable isolated words, all five training conditions provided equivalent benefits. When the task was more difficult, such as was the case when identifying highly confusable CVCs or sentences lacking semantic context, subjects who were trained on materials of a similar nature to those on which they were being tested performed significantly better. However, the addition of environmental sounds revealed a unique asymmetry: training on environmental sounds generalized to the recognition of speech, but training on speech did not generalize to environmental sounds. This pattern of performance suggests that a wide variety of stimulus materials should be used during training to maximize perceptual learning and promote robust generalization to novel acoustic sounds.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Qian-Ji Fu for developing and maintaining TigerCIS, a tool that we find invaluable as speech scientists. They would also like to thank Althea Bauernschmidt and Lawrence Phillips for their assistance in data collection and analysis, and Luis Hernandez for providing technical assistance and advice in the design and implementation of the experimental procedures. This work supported by NIH-NIDCD Training Grant No. T32-DC00012 and NIH-NIDCD Research Grant No. R01-DC00111.

Bradlow, A. R., Toretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Burkholder, R. A. (**2005**). "Perceptual learning of speech processed through an acoustic simulation of a cochlear implant," *Research on Spoken Language Processing Technical Report No*. 13, Speech Research Laboratory, Indiana University, Bloomington, IN.

Chiu, C.-Y. P. (**2000**). "Specificity of auditory implicit and explicit memory: Is perceptual priming for environmental sounds exemplar specific?," Mem. Cognit. **28**, 1126–1139.

Chiu, C.-Y. P., and Schacter, D. L. (**1995**). "Auditory priming for nonverbal information: Implicit and explicit memory for environmental sounds," Conscious Cogn **4**, 440–458.

Clark, G. M. (**2002**). "Learning to understand speech with the cochlear implant," In *Perceptual Learning*, edited by M. Fahle and T. Poggio (MIT Press, Cambridge, MA), pp. 147–160.

Clopper, C. G., Pisoni, D. B., and Tierney, A. T. (**2006**). "Effects of open-set and closed-set task demands on spoken word recognition," J. Am. Acad. Audiol **17**, 331–349.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise vocoded sentences," J. Exp. Psychol. **134**, 222–241.

Dorman, M., and Loizou, P. (**1998**). "The identification of consonants and vowels by cochlear implants patients using a 6-channel CIS processor and by normal hearing listeners using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," J. Acoust. Soc. Am. **102**, 2993–2996.

Egan, J. P. (**1948**). "Articulation testing methods," Laryngoscope **58**, 955–991.

Fu, Q.-J., Galvin, J., Wang, X., and Nogaki, G. (**2006**). "Moderate auditory training can improve speech performance of adult cochlear implant patients," ARLO **6**, 106–111.

Gaver, W. W. (**1993**). "What in the world do we hear?: An ecological approach to auditory event perception," Ecological Psychol. **5**, 1–29.

Gonzales, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.

Gygi, B., Kidd, R. R., and Watson, C. S. (**2004**). "Spectral-temporal factors in the identification of environmental sounds," J. Acoust. Soc. Am. **115**, 1252–1265.

Herman, R., and Pisoni, D. B. (**2000**). "Perception of 'elliptical speech' by an adult hearing-impaired listener with a cochlear implant: some preliminary findings on coarse-coding in speech perception," *Research on Spoken Language Processing Progress Report No*. 24 Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 87–112.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (**1965**). "Articulation-testing methods: Consonantal differentiation with a closed-response set," J. Acoust. Soc. Am. **37**, 158–66.

IEEE. (**1969**). "IEEE recommended practice for speech quality measurements," *IEEE Report No*. 297.

KarlJ. R., and Pisoni, D. B. (**1994**). "Effects of stimulus variability on recall of spoken sentences: A first report," *Research on Spoken Language Processing Progress Report No. 19*, Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 145–193.

Kidd, G. R., Watson, C. S., and Gygi, B. (**2007**). "Individual differences in auditory abilities," J. Acoust. Soc. Am. **122**, 418–435.

Lachs, L., McMichael, K., and Pisoni, D. B. (**2003**). "Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events," *Rethinking Implicit Memory*, edited by J. Bowers and C. Marsolek (Oxford University Press, Oxford), pp. 215–235.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., and DeYoe, E. A. (**2004**). "Human brain regions involved in recognizing environmental stimuli," Cereb. Cortex **14**, 1008–1021.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.

Marcell, M. M., Borella, D., Greene, M., Kerr, E., and Rogers, S. (**2000**). "Confrontation naming of environmental sounds," J. Clin. Exp. Neuropsychol **22**, 830–864.

Morris, C. D., Bransford, J. D., and Franks, J. J. (**1977**). "Levels of processing versus transfer appropriate processing," J. Verbal Learn. Verbal Behav. **16**, 519–533.

National Institutes of Health (NIH). (**1995**). "Cochlear implants in adults and children," NIH Consens Statement **13**, 1–29.

Obleser, J., Wise, R. J. S., Dresner, M. A., and Scott, S. K. (**2007**). "Functional integration across brain regions improves speech perception under adverse listening conditions," J. Neurosci. **27**, 2283–2289.

Reed, C. M., and Delhorne, L. A. (**2005**). "Reception of environmental sounds through cochlear implants," Ear Hear. **26**, 48–61.

Shafiro, V. (**2004**). "Perceiving the sources of environmental sounds with a varying number of spectral channels," Unpublished doctoral dissertation, CUNY, New York.

Shannon, R. V. (**2005**). "Speech and music have different requirements for spectral resolution," Int. Rev. Neurobiol. **70**, 121–134.

Shannon, R. V., Fu, Q.-J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Oto-Laryngol., Suppl. **552**, 1–5.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.

Tice R., and Carrell T. (**1998**). Level 16 V2.0.3. University of Nebraska, Lincoln, NE.

Tulving, E., and Thomson, D. M. (**1973**). "Encoding specificity and retrieval processes in episodic memory," Psychol. Rev. **80**, 352–373.

Tye-Murray, N., Tyler, R., Woodward, G., and Gantz, B. (**1992**). "Performance over time with a Nucleus and Ineraid cochlear implant," Ear Hear. **13**, 200–209.

Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech

Kathryn Hopkins,[a] Brian C. J. Moore, and Michael A. Stone
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 15 August 2007; revised 21 November 2007; accepted 21 November 2007)

Speech reception thresholds (SRTs) were measured with a competing talker background for signals processed to contain variable amounts of temporal fine structure (TFS) information, using nine normal-hearing and nine hearing-impaired subjects. Signals (speech and background talker) were bandpass filtered into channels. Channel signals for channel numbers above a "cut-off channel" (CO) were vocoded to remove TFS information, while channel signals for channel numbers of CO and below were left unprocessed. Signals from all channels were combined. As a group, hearing-impaired subjects benefited less than normal-hearing subjects from the additional TFS information that was available as CO increased. The amount of benefit varied between hearing-impaired individuals, with some showing no improvement in SRT and one showing an improvement similar to that for normal-hearing subjects. The reduced ability to take advantage of TFS information in speech may partially explain why subjects with cochlear hearing loss get less benefit from listening in a fluctuating background than normal-hearing subjects. TFS information may be important in identifying the temporal "dips" in such a background.
 [DOI: 10.1121/1.2824018]

PACS number(s): 43.71.Ky, 43.66.Sr, 43.71.Es [MW] Pages: 1140–1153

I. INTRODUCTION

Information in speech is redundant. For normal-hearing subjects, this means that the signal is robust to corruption, and that speech remains intelligible under adverse listening conditions, such as in high levels of background noise. In the normal auditory system, a complex sound like speech is filtered into frequency channels on the basilar membrane. The signal at a given place can be considered as a time-varying envelope superimposed on the more rapid fluctuations of a carrier (temporal fine structure, TFS) whose rate depends partly on the center frequency and bandwidth of the channel. The relative envelope magnitude across channels conveys information about the spectral shape of the signal and changes in the relative envelope magnitude indicate how the short-term spectrum changes over time. The TFS carries information both about the fundamental frequency (F0) of the sound (when it is periodic) and about its short-term spectrum. For example, if at a particular time there is a formant centered at frequency f_x (and hence one or more relatively intense stimulus components near f_x), then channels centered close to f_x will show a TFS synchronized to f_x, and this will be reflected in the patterns of phase locking in those channels (Young and Sachs, 1979). In the mammalian auditory system, phase locking tends to break down for frequencies above 4–5 kHz (Palmer and Russell, 1986), so it is generally assumed that TFS information is not used for frequencies above that limit. The role of TFS in speech perception for frequencies below 5 kHz remains somewhat unclear.

Many studies have assessed the relative importance of TFS and envelope information for speech intelligibility, for normal-hearing subjects. "Vocoder" processing has been used to remove TFS information from speech, so allowing speech intelligibility based on envelope and spectral cues to be measured (Dudley, 1939; Van Tasell *et al.*, 1987; Shannon *et al.*, 1995). A speech signal is filtered into a number of channels (N), and the envelope of each channel signal is used to modulate a carrier signal, typically a noise (for a noise vocoder) or a sine wave with a frequency equal to the channel center frequency (for a tone vocoder). The modulated signal for each channel is filtered to restrict the bandwidth to the original channel bandwidth and the modulated signals from each channel are then combined. For a single talker, provided that N is sufficiently large, the resulting signal is highly intelligible to both normal-hearing and hearing-impaired subjects (Shannon *et al.*, 1995; Turner *et al.*, 1995; Baskent, 2006; Lorenzi *et al.*, 2006b). However, if the original signal includes both a target talker and a background sound, intelligibility is greatly reduced, even for normal-hearing subjects (Dorman *et al.*, 1998; Fu *et al.*, 1998; Qin and Oxenham, 2003; Stone and Moore, 2003), leading to the suggestion that TFS information may be important for separation of a talker and background into separate auditory streams (Friesen *et al.*, 2001).

As well as removing TFS information, vocoder processing also "smears" spectral information, an effect that is greatest when N is small. If the analysis filters are of a similar width to the auditory filters, however, the spectral information that is available to the auditory system is only slightly

[a]Author to whom correspondence should be addressed. Electronic mail: kh311@cam.ac.uk

reduced compared to normal, though TFS information is still absent.

Another method for assessing the roles of different types of temporal information in speech is to attempt to remove envelope information but to leave TFS information (partially) intact. This was first attempted by infinite peak clipping of a wideband speech signal (Licklider and Pollack, 1948), and later by using the Hilbert transform (Bracewell, 1986) to separate envelope and TFS information in each of a number of frequency channels (Smith *et al.*, 2002). The TFS in each channel is preserved, but the envelope cues are removed, and the channel signals are then combined. Effectively, the processing in each channel behaves like a very fast compressor with an infinite compression ratio. For brevity, we will refer to this signal-processing method as "TFS processing." At first sight, this method may appear to remove temporal envelope information and leave TFS intact. However, because envelope and TFS information are correlated, the envelope can be partially re-introduced by filtering in the peripheral auditory system, especially when the channels used in the processing have large bandwidths (Ghitza, 2001). This problem is reduced if the signal is split into many narrow channels before removal of the envelope information, although some envelope cues may remain.

Gilbert and Lorenzi (2006) investigated the extent to which these recovered cues could be used to identify vowel-consonant-vowel (VCV) nonsense syllables. They subjected nonsense syllables to TFS processing, and then passed the resulting signals through an array of filters that simulated filtering in the normal peripheral auditory system. The envelopes of the outputs of these filters were extracted and used to modulate tones at the center frequencies of the filters. This is similar to tone-vocoder processing. The modulated tones were summed and presented to normal-hearing subjects, who were asked to identify the consonant that was presented. When the TFS processing used a small number of broad channels, subjects could identify the consonant accurately from the recovered envelope information. However, when the number of channels used in the TFS processing was eight or more, subjects scored close to chance. The authors concluded that if a signal is filtered into a sufficiently large number of channels before removing envelope cues, any recovered envelope cues are insufficient for intelligibility of VCVs. VCV syllables that are TFS processed with a large number of analysis channels are reasonably intelligible to normally hearing listeners, after some training (Lorenzi *et al.*, 2006b), which suggests that TFS cues alone can convey useful speech information.

Results from several studies have led to the suggestion that the ability to use TFS information is adversely affected by cochlear hearing loss. Much of this work has investigated the discrimination of synthetic complex sounds by hearing-impaired subjects (Lacher-Fougère and Demany, 1998; 2005; Moore and Skrodzka, 2002; Moore and Moore, 2003; Moore *et al.*, 2006; Hopkins and Moore, 2007). For example, Hopkins and Moore (2007) tested the ability of normal-hearing and hearing-impaired subjects to discriminate a harmonic complex tone from a frequency-shifted tone, in which all components were shifted up by the same amount in Hz (de Boer, 1956). The frequency-shifted tone had very similar temporal envelope and spectral envelope characteristics to the harmonic tone, but a different TFS. All tones were passed through a fixed bandpass filter, to reduce excitation-pattern cues. When the filter was centered on the 11th component, so that the components within the passband were unresolved, subjects with moderate cochlear hearing loss performed poorly, while normal-hearing subjects could do the task well. Hopkins and Moore concluded that moderate cochlear hearing loss usually led to a reduced ability to use TFS information.

The reason for this is not clear. One possibility is that the precision of phase locking is reduced by cochlear hearing loss. One study found that phase locking was reduced in animals with induced hearing loss (Woolf *et al.*, 1981), but another study found normal phase locking in such animals (Harrison and Evans, 1979). It is unclear whether the types of pathologies that cause cochlear hearing loss in humans lead to reduced phase locking. Another possible reason for a reduced ability to use TFS information is that TFS information could be decoded by cross correlation of the outputs of two points on the basilar membrane (Loeb *et al.*, 1983; Shamma, 1985). A deficit in this process, produced by a change in the traveling wave on the basilar membrane, would impair the ability to use TFS information even if phase locking were normal. The broader auditory filters typically associated with cochlear hearing loss (Liberman and Kiang, 1978; Glasberg and Moore, 1986) could also lead to a reduced ability to use TFS information. The TFS at the output of these broader filters in response to a complex sound will have more rapid fluctuations and be more complex than normal. Such outputs may be uninterpretable by the central auditory system (Sek and Moore, 1995; Moore and Sek, 1996).

A reduced ability to use TFS information could explain some of the perceptual problems of hearing-impaired subjects (Lorenzi *et al.*, 2006a). TFS information may be important when listening in background noise, especially when the background is temporally modulated, as is often the case when listening in "real life," for example, when more than one person is speaking. Normal-hearing subjects show better speech intelligibility (or lower speech reception thresholds, SRTs) when listening in a fluctuating background than when listening in a steady background (Festen and Plomp, 1990; Baer and Moore, 1994; Peters *et al.*, 1998; Füllgrabe *et al.*, 2006), an effect which is sometimes called "masking release." Hearing-impaired subjects show a much smaller masking release, and it has been suggested that this may be because they are poorer at "listening in the dips" of a fluctuating masker than normal-hearing subjects (Duquesnoy and Plomp, 1983; Peters *et al.*, 1998; Lorenzi *et al.*, 2006b). Reduced audibility may account for some of the reduction in masking release measured for hearing-impaired subjects (Bacon *et al.*, 1998), although the effect persists even when audibility is restored (Peters *et al.*, 1998; Lorenzi *et al.*, 2006a). TFS information may be important in "dip listening" tasks, as it could be used to identify points in the stimulus when the level of the target is high relative to the level of the masker;

if the target and masker do not differ in their TFS, or no TFS information is available, dip listening may be ineffective.

Some studies have investigated the ability of hearing-impaired subjects to use TFS information in speech. Buss *et al.* (2004) showed that there was a correlation between temporal processing as assessed with psychoacoustic tasks and the ability of hearing-impaired subjects to recognize words in quiet. Lorenzi *et al.* (2006b) attempted to measure the ability of young and elderly hearing-impaired subjects to use TFS information in speech more directly. They applied 16-channel TFS processing to VCV nonsense syllables and asked subjects to identify the consonant in each syllable. According to Gilbert and Lorenzi (2006), this number of channels should be sufficient to prevent the use of recovered envelope cues. Hearing-impaired subjects performed poorly at this task, while normal-hearing subjects scored around 90% correct after some training. Lorenzi *et al.* interpreted this result as indicating that the hearing-impaired subjects had a very limited ability to use the TFS information in the speech, whereas it was usable by normal-hearing subjects. Lorenzi *et al.* (2006b) also measured masking release for the young hearing-impaired subjects when listening to unprocessed speech in steady and modulated noise. The amount of masking release was highly correlated with the score obtained for speech in quiet that had been subjected to TFS processing. This result is consistent with the argument made earlier, that the ability to use TFS is important for listening in the dips of a background sound.

A potential problem with the use of TFS-processed signals is that during gaps in the speech in a particular processing channel, low-level recording noise is amplified to the same level as the speech information. This is because the process is equivalent to multi-channel compression with an infinite compression ratio; whatever the original envelope amplitude in a given channel, the output envelope amplitude is constant. Channels with no speech information at a particular time are filled with distracting background sound. As a result, TFS-processed speech sounds harsh and very noisy. This may pose a particular problem to hearing-impaired subjects who, because of their broadened auditory filters, would suffer more from masking between channels. The problem becomes worse as the signal is split into more channels, as this results in more across-channel masking. Also, hearing-impaired listeners would be poorer at recovering any envelope cues that may still be available, again as a result of their broadened auditory filters. This could account for some of the difference in performance between normal-hearing and hearing-impaired subjects when listening to TFS speech. Here, a different approach was used to assess the use of TFS information by normal-hearing and hearing-impaired subjects. Rather than creating a signal that contains speech information only in its TFS, performance was measured as a function of the number of channels containing TFS information; the other channels were noise or tone vocoded, so that they conveyed only envelope information.

II. RATIONALE

Hopkins and Moore (2007) found that subjects with moderate cochlear hearing loss could make little use of TFS information to discriminate complex tones. If similar subjects were completely unable to use TFS information in speech, they would be expected to perform as well when listening to speech that had been vocoded to remove TFS information as when listening to unprocessed speech, provided that N was sufficiently large that the frequency selectivity of the processing was similar to or better than that of the peripheral auditory system of the subject, thus avoiding significant loss of spectral information. However, Baskent (2006) found that hearing-impaired subjects performed better in a phoneme identification task when the syllables were unprocessed than when they were processed with a 32-channel noise-band vocoder. The disparity might arise because hearing-impaired subjects may be able to use TFS information at low carrier frequencies, but may be unable to use it at high frequencies. Hopkins and Moore (2007) showed that hearing-impaired subjects had a greatly reduced ability to discriminate the TFS of complex tones with unresolved components when all components were above 900 Hz, but they did not investigate sensitivity to TFS for lower frequencies. It is possible that subjects with moderate cochlear hearing loss are able to use TFS information below 900 Hz, which could explain why they performed better in the unprocessed condition than in the 32-channel vocoded condition in the study of Baskent (2006). If subjects with moderate cochlear hearing loss can use TFS information only at low carrier frequencies, progressively replacing vocoded information with unprocessed information, starting at low frequencies, should improve performance only up to a cut-off frequency above which TFS information cannot be used. This hypothesis was tested here.

SRTs corresponding to 50% correct keyword identification were measured for signals that were unprocessed for channels up to and including cut-off channel number (CO) and were vocoded for higher-frequency channels. The value of CO, which determined the amount of TFS information available in the signal, was varied from 0 to 32. A competing-talker background was used, because, as described earlier, TFS information may be particularly important for listening in backgrounds that have temporal "dips."

III. METHOD

A. Subjects

Nine normal-hearing subjects and nine hearing-impaired subjects took part in the experiment. The normal-hearing subjects were aged between 18 and 27 years and had audiometric thresholds of 15 dB hearing level (HL) or less at octave frequencies between 250 and 8000 Hz. The audiograms of the test ears of the hearing-impaired subjects are shown in Fig. 1 (subjects HI 1 to HI 9) and the age of each subject is shown in parentheses. All hearing-impaired subjects had air-bone gaps of 15 dB or less, and normal tympanograms, suggesting that their hearing loss was cochlear in origin. Hearing-impaired subjects were tested with the "TEN HL" test, which indicated no cochlear dead region for any subject (Moore *et al.*, 2004).

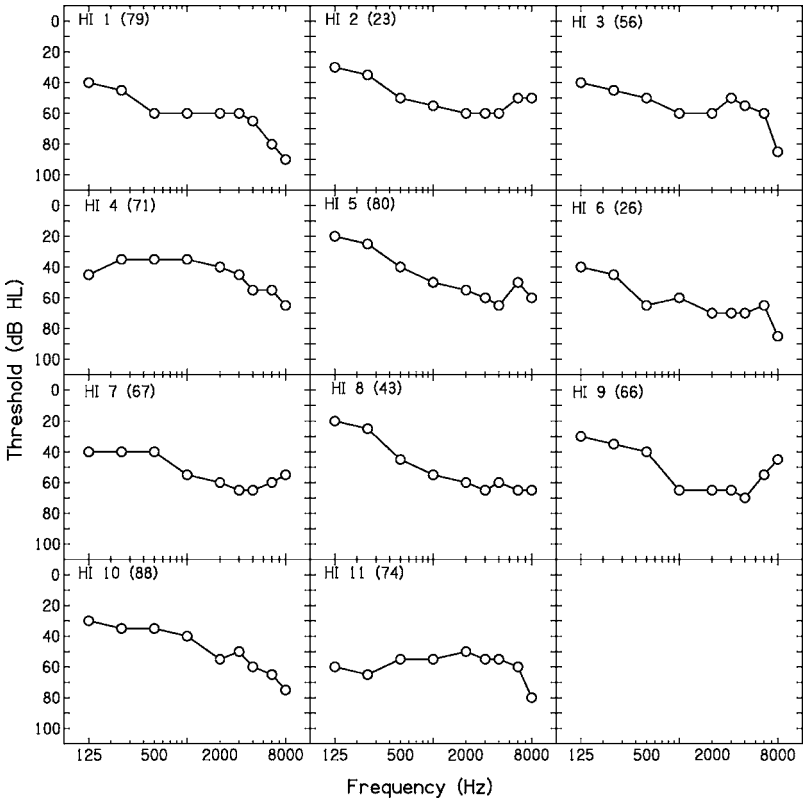


FIG. 1. Air conduction audiometric thresholds of the test ears of the hearing-impaired subjects for experiments one and two. The ages of the subjects (in years) are shown in parentheses.

B. Speech material

Subjects were asked to repeat sentences presented in a competing talker background. The background began 500 ms before the target sentence, and continued after the target sentence had finished for about 700 ms (the exact value depended on the length of the target sentence). Each sentence list was added to a randomly chosen portion of a passage of continuous prose spoken by a competing talker. Long gaps between sentences and pauses for breath were removed from the background passage by hand editing. The same passage was used in both training and testing sessions. Both the target sentences and the competing talker passage had the same long-term spectral shape; for frequencies up to 500 Hz, the spectrum level was roughly constant, and for frequencies above 500 Hz the spectrum level fell by 9 dB per octave. For the training session, IEEE sentences were used (Rothauser *et al.*, 1969). For the testing session, sentences were taken from the adaptive sentence list (ASL) corpus (MacLeod and Summerfield, 1990). Both target and competing talkers were male speakers of British English. The target talker had a fundamental frequency (F0) range of about 130–200 Hz, and the competing talker had a larger F0 range of about 130–280 Hz. The target and background speech were added together at the appropriate signal-to-background ratio (SBR) before processing.

C. Processing and equipment

Speech signals were split into 32 channels with center frequencies spanning the range 100–10,000 Hz, with an array of linear-phase, finite-impulse-response (FIR) filters. The filters had a variable order so that the transition bands of each filter had similar slopes when plotted on a logarithmic frequency scale. Each filter was designed to have a response of −6 dB at the frequencies at which its response intersected with the responses of the two adjacent filters. Channel edges were regularly spaced on an equivalent-rectangular-bandwidth (ERB_N) number scale and each channel was 1 ERB_N wide (Glasberg and Moore, 1990). This filtering was designed to simulate the frequency selectivity of the normal auditory system, so that the processing preserved nearly all of the spectral information available in the original signal. The signals from each channel were time aligned to compensate for the time delays introduced by the bandpass filtering. Stimuli were processed with nine values of CO (0, 4, 8, 12, 16, 20, 24, 28, 32). Channels with channel numbers up to and including CO were not processed further. Channels with channel numbers above CO were vocoded. The signals from these channels were half-wave rectified and these rectified signals were used to modulate white noises.[1] Each modulated noise was subsequently filtered with the initial analysis filters and shaped to have the same spectral shape as the long-term spectrum of the original target speech from that channel. Consequently, envelope fluctuations with frequencies greater than half of the channel bandwidth were attenuated. After processing, the signals from the vocoded and unprocessed channels were added together. All signals were generated with a high-quality 16 bit PC soundcard (Lynx One) at a sampling rate of 22,050 Hz, passed through a Mackie 1202-VLZ mixing desk and presented to the subject monaurally via Sennheiser HD580 headphones. Subjects were seated in a double-walled sound-attenuating chamber.

D. Procedure

Microphones were placed in both the chamber and control room to allow communication between the experimenter

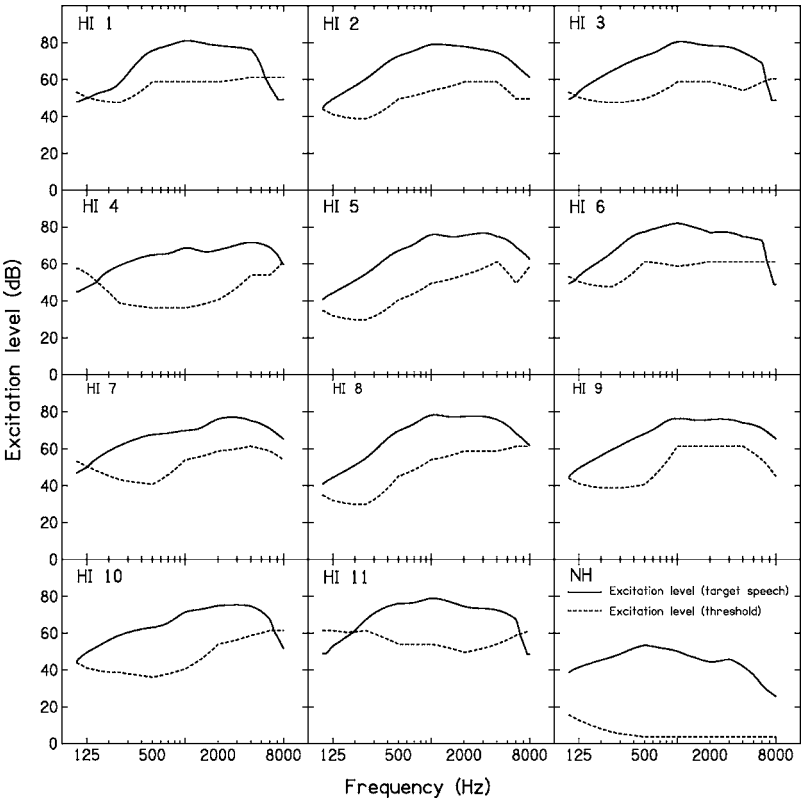


FIG. 2. Excitation levels of the target speech (solid lines) and excitation levels at threshold (dashed lines) for individual hearing-impaired subjects. Excitation levels for normal-hearing subjects are shown for comparison in the bottom-right panel.

and subject, although the control room microphone was only routed to the chamber headphones in the gaps between stimulus presentations. Target speech was presented to the normal-hearing subjects at a constant rms level of 65 dB sound pressure level (SPL), which was equivalent to a spectrum level of 36.6 dB (re 20 μPa) between 100 and 500 Hz (see Sec. III B for a description of the spectral shape). The level of the competing talker was varied to give the appropriate SBR, except when the SBR was less than −16 dB. Below this SBR, the level of the competing talker was not increased further, but instead the level of the target speech was reduced, to prevent the combined signal becoming uncomfortably loud. In practice, this was not necessary for any of the hearing-impaired subjects.

Previous studies have shown that audibility can account for some of the difference in performance between normal-hearing and hearing-impaired subjects listening in a temporally modulated background if stimuli are presented at the same level to both groups of subjects (Bacon *et al.*, 1998; George *et al.*, 2006). To reduce such effects, gains were applied to the combined target and background signal as prescribed by the CAMEQ hearing aid fitting method, according to the audiometric thresholds of each subject (Moore *et al.*, 1998). Gains were specified at audiometric frequencies between 250 and 6000 Hz. The CAMEQ gains are designed to ensure speech audibility between these frequencies. Relatively more gain is prescribed for higher frequencies and this compensates for the increased upward spread of masking that is expected at higher overall levels, so helping to avoid the "rollover" effect on speech intelligibility as overall level increases (Fletcher, 1953; Studebaker *et al.*, 1999). The CAMEQ gains were applied to the processed signals using a linear-phase FIR filter with 443 taps.

E. Audibility calculations

To check that the target speech would be audible for the hearing-impaired subjects after the CAMEQ gains were applied, excitation patterns were calculated for a signal that had the same long-term average spectrum as the target speech signal used for each subject. The spectrum for each subject was obtained by determining the long-term average spectrum of the speech with an overall level of 65 dB SPL and adding the CAMEQ gains at each frequency (with interpolation of gains for frequencies between the values specified by CAMEQ). Mean excitation levels between 100 and 8000 Hz were calculated for each subject using a model similar to that proposed by Moore and Glasberg (2004), but updated to incorporate the middle ear transfer function proposed by Glasberg and Moore (2006). Excitation levels are calculated relative to the excitation evoked by a 1000 Hz tone presented in free field with frontal incidence at a level of 0 dB SPL. The model allows the audiometric thresholds (in dB HL) of the individual subject to be entered. Default values were assumed for the proportion of the hearing loss attributed to outer hair cell and inner hair cell dysfunction. The model also gave estimates of the excitation level at threshold as a function of frequency for each subject. Figure 2 shows the excitation level at threshold and the mean excitation evoked by the (amplified) speech signal for each subject. The excitation level for the target speech was well above threshold excitation level except at very high or very low frequencies for some subjects. For most subjects, and for frequencies between 500 and 5000 Hz (the frequency range that is most important for speech intelligibility), the excitation level of the target speech was more than 15 dB above the excitation

level at threshold, meaning that the entire dynamic range of the speech would have been audible (ANSI, 1997).

F. Training

Previous studies using vocoded speech material have shown large learning effects (Stone and Moore, 2003; 2004; Davis *et al.*, 2005), so a training period was included before testing. Training lasted approximately 1 h, and took place separately to the testing session. First, subjects were played two passages of connected discourse to familiarize them with the task of listening in a competing talker background and to introduce the vocoder-processed speech. The first passage was unprocessed, and the second was vocoded across all 32 channels. The level of the competing talker was initially low, but was increased gradually throughout the passages. Subjects were instructed to listen to the target talker for as long as possible. The hearing-impaired subjects found this difficult, and so were given transcripts of the target passages to follow, which made the task easier.

For the next phase of training, IEEE sentences were presented at a fixed SBR. Six lists were presented, each made up of ten sentences. The sentences were processed with different values of CO and an SBR was selected by the experimenter to yield scores of approximately 70% correct. Subjects were required to repeat each sentence and the number of correctly identified key words was recorded. When subjects did not repeat the sentence perfectly, they were told the correct answer, and the sentence was repeated.

Finally, subjects were given an opportunity to practice the task used in the testing session. Four word lists similar to those in the ASL corpus were used. The same procedure was used as for the testing session, as described below.

G. Testing

Two consecutively presented ASL sentence lists were used for each condition and the order of presentation of conditions was counterbalanced across subjects. The SBR of the target and competing talker was varied adaptively. If a subject identified two or more keywords correctly in a sentence, the next sentence was presented with a SBR that was k dB lower, and if the subject identified fewer than two keywords correctly, the next sentence was presented with a SBR that was k dB higher. Before the third turnpoint was reached, k was equal to 4 dB; subsequently it was equal to 2 dB. The first sentence in each list was initially presented at an adverse SBR, at which the subject was expected to identify no keywords correctly. If the subject scored fewer than two keywords correctly, this sentence was repeated at an SBR that was 4 dB higher until at least two keywords were correctly identified. Subsequent sentences in each list were presented once only. For each sentence list, the total number of keywords presented at each SBR was recorded, as well as the number of keywords that were identified correctly for each SBR. The first sentence in each list was not included in these totals, as subjects could have heard this sentence more than once.

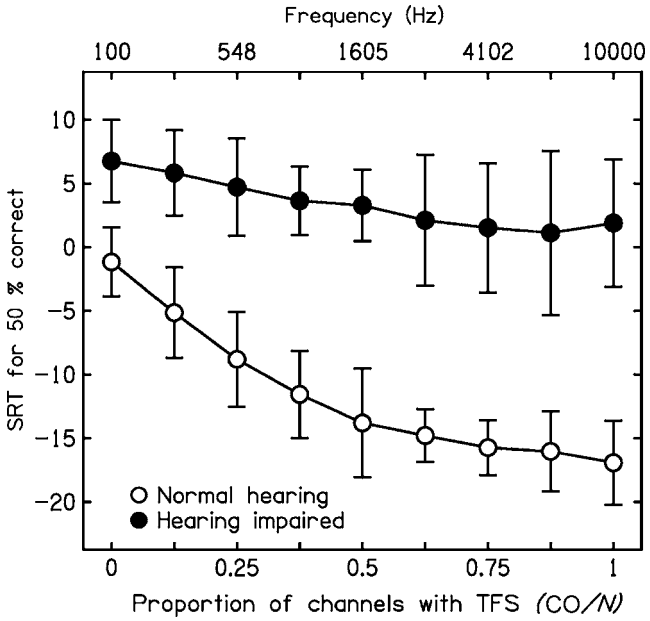


FIG. 3. Mean SRTs for normal-hearing and hearing-impaired subjects, plotted as a function of CO/N. The frequency corresponding to CO/N is shown along the top axis. Error bars show ± one standard deviation across subjects.

H. Analysis

For each SBR, the total keywords presented and keywords correct were summed for the two sentences lists that were presented for each condition. These values were used to perform a probit analysis (Finney, 1971), from which the SRT corresponding to the SBR required for 50% correct identification was estimated for each subject and each condition. In some cases, because of the scatter in the data, the probit analysis failed to fit the data and gave a slope of the psychometric function that was not significantly different from zero. This happened for at least one of the conditions for five of the normal-hearing subjects, but for only one of the hearing-impaired subjects (HI 8). For these cases, the SRT was estimated by plotting the proportion of correctly identified words against the SBR at which the words were presented. A line was drawn by eye to best fit the data points, and this line was used to exclude points from the probit analysis that did not fit the general trend. The probit analysis was then redone. In one case (NH 5, CO=20), after this procedure the probit analysis still did not give a psychometric function with a slope significantly different from zero, so the SRT for this case was treated as a missing data point for the remaining analysis.

An analysis of variance (ANOVA) was performed on all of the data from the normal-hearing and hearing-impaired subjects, with a within-subjects factor of CO and a between-subjects factor of subject type (normal hearing or hearing impaired).

IV. RESULTS

Figure 3 shows the mean data for both normal-hearing and hearing-impaired subjects. Mean SRTs are plotted for each value of CO/N. The hearing-impaired subjects per-

TABLE I. Differences between mean SRTs measured with different values of CO for normal-hearing subjects in experiment one. Equivalent cut-off frequencies (in Hz) are also shown. The LSD calculated using Fisher's LSD procedure was 2.0. Differences equal to or above this value are shown in bold.

CO		0	4	8	12	16	20	24	28	32
	Freq. (Hz)	100	277	548	965	1605	2590	4102	6427	10,000
0	100	0								
4	277	**3.8**	0							
8	548	**7.4**	**3.7**	0						
12	965	**10.2**	**6.4**	**2.8**	0					
16	1605	**12.4**	**8.7**	**5.0**	**2.2**	0				
20	2590	**13.4**	**9.7**	**6.0**	**3.2**	1.0	0			
24	4102	**14.3**	**10.6**	**6.9**	**4.2**	1.9	0.9	0		
28	6427	**14.6**	**10.9**	**7.2**	**4.5**	**2.2**	1.2	0.3	0	
32	10,000	**15.5**	**11.8**	**8.1**	**5.4**	**3.1**	**2.1**	1.2	0.9	0

formed more poorly than the normal-hearing subjects in all conditions, but the difference in performance varied with CO; for larger values of CO, the difference in performance between normal-hearing and hearing-impaired subjects was greater. The main effects of subject type and CO were significant [$F(1,8)=95.2$, $p<0.001$ and $F(8,128)=53.9$, $p<0.001$, respectively], and there was also a significant interaction between subject type and CO [$F(8,128)=12.2$, $p<0.001$].

CO had a greater effect on performance for the normal-hearing subjects than for the hearing-impaired subjects. For example, both subject groups performed better when speech was completely unprocessed (CO=32) than when it was completely vocoded (CO=0), but the difference in performance between these conditions was much greater for the normal-hearing than for the hearing-impaired subjects (mean differences were 15.8 and 4.9 dB, respectively). *Post hoc* Fisher's least-significant-difference (LSD) tests were used to determine whether the SRTs measured with different values of CO were significantly different from each other within each subject group. Tables I and II show the differences between mean scores for each value of CO for the normal-hearing subjects and hearing-impaired subjects, respectively. Values greater than the least significant difference are shown in bold.

Figure 4 shows results for individual hearing-impaired subjects. The mean results for the normal-hearing subjects are shown in the bottom-right panel for comparison. Between-subject variability in overall performance was larger for the hearing-impaired than for the normal-hearing subjects. The pattern of results across conditions also varied more between hearing-impaired subjects. The benefit gained from the additional TFS information that was present when CO was large varied, with some hearing-impaired subjects benefiting little, if at all (for example, HI 4 and HI 5) and others benefiting almost as much as the normal-hearing subjects (HI 8).

V. DISCUSSION

Normal-hearing subjects appear to benefit more than hearing-impaired subjects from the replacement of vocoded speech information with unprocessed speech information. This is consistent with the idea that the hearing-impaired subjects had a reduced ability to use TFS information, which is consistent with previously published results (Lorenzi *et al.*, 2006b; Moore *et al.*, 2006; Hopkins and Moore, 2007). Both groups did, however, improve as CO increased, though the amount of benefit from the additional TFS information varied across hearing-impaired subjects. This may reflect different abilities to use TFS information among hearing-impaired subjects with broadly similar audiometric thresholds, which could account for the weak correlation between audiometric thresholds and the ability to understand speech in noise previously reported for hearing-impaired subjects (Festen and Plomp, 1983; Glasberg and Moore, 1989). Other studies have also reported large individual differences in performance between hearing-impaired subjects when tasks require the use of TFS information (Buss *et al.*, 2004; Moore *et al.*, 2006).

TABLE II. As Table I, but for hearing-impaired subjects.

CO		0	4	8	12	16	20	24	28	32
	Freq. (Hz)	100	277	548	965	1605	1605	4102	6427	10,000
0	100	0								
4	277	0.9	0							
8	548	**2.1**	1.1	0						
12	965	**3.1**	**2.2**	1.1	0					
16	1605	**3.5**	**2.6**	1.4	0.4	0				
20	2590	**4.7**	**3.7**	**2.6**	1.5	1.2	0			
24	4102	**5.2**	**4.3**	**3.2**	**2.1**	1.8	0.6	0		
28	6427	**5.6**	**4.7**	**3.6**	**2.5**	**2.2**	1.0	0.4	0	
32	10,000	**4.9**	**3.9**	**2.8**	1.8	1.4	0.2	0.4	0.8	0

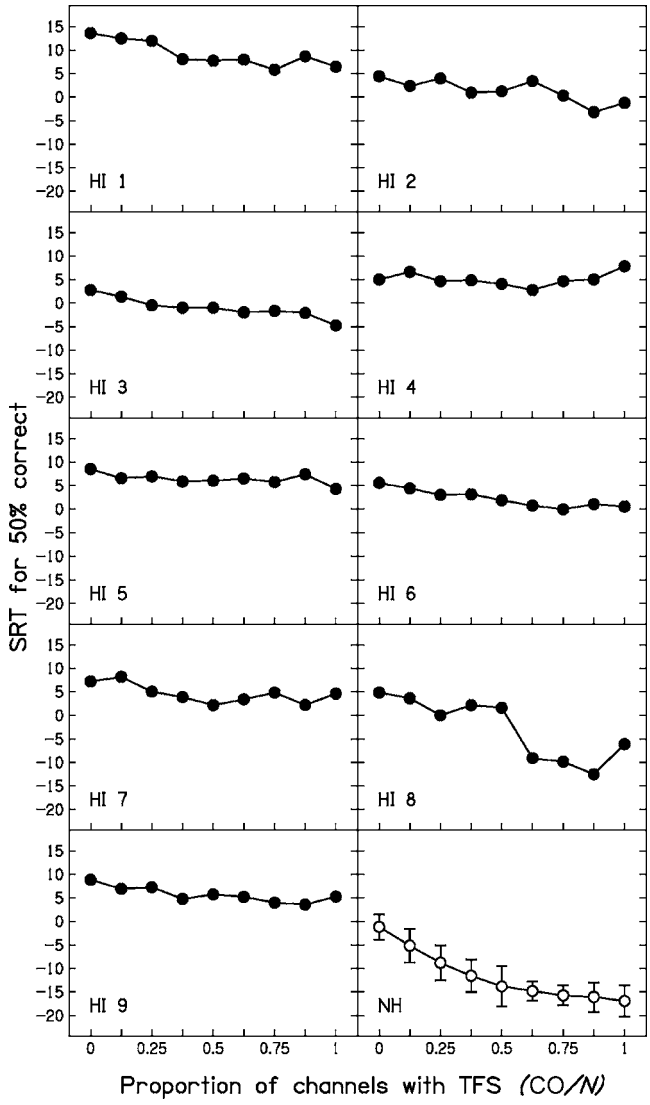


FIG. 4. Individual SRTs for the hearing-impaired subjects, plotted as a function of CO/*N*. Mean SRTs measured for the normal-hearing subjects are shown in the bottom-right panel for comparison.

One possible concern is that the mean age of the normal-hearing subjects was much less than the mean age of the hearing-impaired subjects (21.9 and 56.8 years, respectively), so the reduced benefit from the additional TFS might have been due to age rather than to hearing loss *per se*. Some previous studies have been interpreted as indicating that older subjects with near-normal audiometric thresholds have temporal processing deficits (Pichora-Fuller, 2003; Pichora-Fuller *et al.*, 2006). However, other studies have tested young and elderly normal-hearing subjects listening to target speech in a temporally modulated background similar to that used here, and found relatively small differences in performance between the two groups (Takahashi and Bacon, 1992; Peters *et al.*, 1998; Dubno *et al.*, 2002; Lorenzi *et al.*, 2006a). The differences were much smaller than the difference in performance seen here between the hearing-impaired and normal-hearing subjects in the unprocessed condition (CO=32). It is possible that some of the reduced ability to use TFS information seen for the hearing-impaired subjects in the current study could be attributed to their age, rather than their hearing loss. However, the pattern of results for the two young hearing-impaired subjects tested here (HI 2, 23 years and HI 6, 26 years) did not differ markedly from the pattern of the mean data for the hearing-impaired subjects, suggesting that hearing loss, rather than age was the important factor contributing to the reduced ability to use TFS information. The benefit from the addition of TFS was quantified as the difference between the SRT for CO=0 and CO =32. The correlation between benefit and age for the hearing-impaired subjects was not significant ($r=-0.26$, $p=0.50$). Again, this suggests that age was not the important factor determining the benefit from added TFS information.

Phase locking in the normal auditory system is widely believed to break down for frequencies above 4000–5000 Hz (Palmer and Russell, 1986; Moore, 2003). If this is true, TFS information above 4000–5000 Hz should be unusable, and so no improvement in performance would be expected when TFS information was added in the higher-frequency channels. Consistent with this, the Fisher LSD tests showed no significant difference in performance for the normal-hearing subjects for CO values from 24 to 32; CO =24 corresponds to a cut-off frequency of 4102 Hz. For the hearing-impaired subjects, performance appears to plateau at a lower value of CO. With one exception, LSD tests revealed no significant difference in performance for values of CO from 16 to 32; CO=16 corresponds to a frequency of 1605 Hz. This is consistent with the idea that hearing-impaired subjects may be able to use TFS information at low frequencies, but are unable to use higher-frequency TFS information, even for frequencies where phase locking is believed to be robust in the normal auditory system.

A possible concern when comparing hearing-impaired and normal-hearing subjects is that differences in performance may be explained by differences in audibility. For our experiment, this explanation seems unlikely. Figure 2 shows that the entire dynamic range of speech would have been audible for most of the hearing-impaired subjects for frequencies between 500 and 5000 Hz. Audibility was compromised at very low and high frequencies for some subjects, but this reduced audibility is unlikely to have affected speech intelligibility and cannot account for the large differences between hearing-impaired and normal-hearing subjects.

The improvement in SRT as CO increased has been interpreted so far as reflecting an ability use TFS information, but this is not the only interpretation of these results. Another possibility is connected with the idea that a noise-band vocoder may introduce distracting or masking low-frequency modulations into the signal. Whitmal *et al.* (2007) found that normal-hearing subjects scored better when tested with a tone vocoder than with a noise vocoder (see also Dorman *et al.*, 1997). They suggested that modulations introduced by the noise carrier may have caused a reduction in speech intelligibility. The modulation spectrum of a bandpass filtered noise is triangular (Schwartz, 1970), with more modulation energy at low frequencies. This means that the modulations introduced by the noise carrier are dominated by modulation frequencies that are similar to those thought to be important in understanding speech (Drullman *et al.*, 1994a; 1994b; Shannon *et al.*, 1995). When the number of analysis channels is

large (so the channel widths are small), this is an even greater problem, as higher-frequency modulations are removed when the channel signals are filtered after vocoder processing, leaving the signal even more dominated by low-frequency noise modulations. It is possible that the normal-hearing subjects and the hearing-impaired subjects who showed greater improvement as CO increased did not benefit from the additional TFS information, but performed better because the spurious modulations introduced by the noise carrier were reduced, as the proportion of the signal that was vocoded was reduced.

Another factor that might have influenced the change in performance with increasing CO is connected with the effect of the processing on the representation of high-rate envelope fluctuations. Rosen (1992) suggested that modulations between 50 and 500 Hz are important in providing information on voice periodicity, and this voice periodicity information is important for listening in a competing talker background (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990). In experiment one, the speakers were male, with F0s varying between 130 and 280 Hz. The processing used 32 channels, which were equally spaced on an ERB_N-number scale; each channel was 1 ERB_N wide. This was intended to simulate the frequency selectivity of the normal auditory system. The highest modulation rate that can be carried by a channel is determined by the bandwidth of the channel. The filtering that was used subsequent to modulation of the channel carriers would have attenuated the sidebands produced by the modulation, hence reducing the modulation depth for high rates. As a result, voice periodicity information would have been partially removed from channels tuned to lower center frequencies.

This restriction of periodicity information would not have any important effects for normally hearing subjects, because the filters used in the processing had comparable widths to the "normal" auditory filter. However, hearing-impaired subjects generally have broader auditory filters than normal-hearing subjects, so, for unprocessed speech, higher-frequency modulation sidebands would be attenuated less by the peripheral auditory system. Consequently, post-processing filtering of the vocoded signal into 1 ERB_N-wide channels could reduce the periodicity information available to the hearing-impaired subjects, and this could be a reason for their worse performance with CO=0 than with CO=32.

These possible explanations for the improvement in performance with increasing CO found for the normal-hearing subjects and some of the hearing-impaired subjects were investigated in experiment two.

VI. EXPERIMENT TWO

A. Rationale

Experiment two was broadly similar to experiment one, but a tone vocoder was used rather than a noise vocoder. The carrier signals were sine waves with frequencies equal to the channel center frequencies. No random modulations were introduced by the carrier signals, unlike for the noise vocoder used in experiment one.

Previous work has suggested that subjects with moderate cochlear hearing loss have auditory filters that are between two and four times as broad as those for normal-hearing subjects (Glasberg and Moore, 1986; Moore, 2007). In experiment two, the signal was divided into either 8 or 16 channels before processing, rather than 32, so that each channel was wider, and more comparable to the auditory filters of the hearing-impaired subjects (channels were 4 or 2 ERB_N wide rather than 1, as previously). This avoided the possible loss of modulation at F0 rates. A consequence of splitting the signal into fewer channels before vocoder processing is that more spectral detail from the original signal is lost. If the filters used in the processing are broader than those in the peripheral auditory system, as they would be for normal-hearing subjects, this in itself may lead to poorer performance. To check the effect of decreasing N, and to allow comparison with the results of experiment one, a condition was run with $N=32$ and CO=0, so that stimuli were fully tone vocoded, but with the same value of N as for experiment one.

B. Method

1. Subjects

Five of the normal-hearing and seven of the hearing-impaired subjects who took part in experiment one also took part in experiment two. Four normal-hearing and two hearing-impaired subjects were newly recruited. Recruitment criteria were the same as for experiment one. The audiograms and ages of all of the hearing-impaired subjects used in both experiments are shown in Fig. 1. HI 10 and HI 11 took part in experiment two only, HI 4 and HI 6 took part in experiment one only, and the remaining subjects took part in both experiments.

2. Speech material

ASL lists were used for training, as most of the subjects had already heard these in the testing session for experiment one. For the testing session, Bench-Kowal-Bench (BKB) sentence material was used, which is similar in style to the ASL sentence material (Bench and Bamford, 1979).

3. Processing

Sentences were processed in a similar way as for experiment one, but a tone vocoder was used rather than a noise vocoder, and the signal was split into 8 or 16 channels before processing rather than 32 (so channels were 4 or 2 ERB_N wide rather than 1 ERB_N wide). Signals were split into channels, and the envelope of each channel was extracted as before. Sine waves with frequencies equal to the center frequency of each channel were used as carrier signals rather than noise bands, and these sine waves were modulated with the envelope of the original channel signals. As before, processed channel signals were filtered to remove sidebands that were introduced as a result of the processing, so limiting the frequency of modulation that could be carried in each channel.

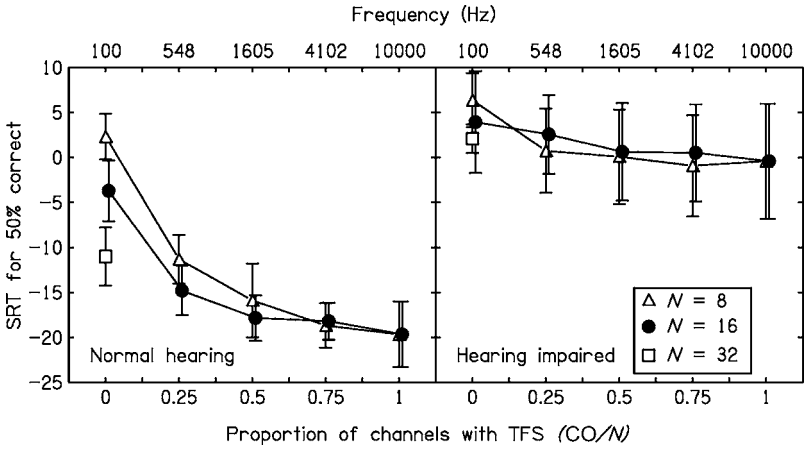


FIG. 5. Mean SRTs for normal-hearing subjects (left) and hearing-impaired subjects (right), plotted as a function of CO/*N*. Error bars show ± one standard deviation across subjects.

4. Conditions and procedure

For $N=8$, values of CO were 2, 4, 6 and 8. For $N=16$, values of CO were 0, 4, 8 and 12 (note that the condition where $N=16$ and CO=16 is the same as $N=8$ and CO=8, so this condition was not retested). For five of the normal-hearing subjects, and five of the hearing-impaired subjects, an additional condition was tested, with $N=32$ and CO=0, but still using a tone vocoder. The procedure for training and testing sessions was the same as for experiment one, except for the differences in sentence material, as noted previously. Data were analyzed in the same way as for experiment one.

C. Results

The results are summarized in Fig. 5. Mean SRTs are plotted against CO/*N*. A given value of CO/*N* corresponds to a fixed frequency, as indicated at the top of the panels in Fig. 5. As in experiment one, the hearing-impaired subjects performed more poorly than the normal-hearing subjects for all conditions, and the difference in performance between the two groups was greatest when CO/$N=1$. An ANOVA was performed with *N* and CO/*N* as within-subject factors and subject type as a between-subject factor. The main effects and two-way interactions were all highly significant ($p<0.001$) and the three-way interaction was also significant [$F(4,60)=3.33$, $p=0.02$].

Differences between mean results for different values of CO/*N* are shown in Tables III and IV for the normal-hearing and hearing-impaired subjects, respectively. For the normal-hearing subjects, SRTs did not differ significantly for CO/$N=0.75$ and 1 (for $N=8$) or for CO/$N=0.5$, 0.75 and 1 (for $N=16$). Thus, performance reached a plateau for higher values of CO/*N*, as found in experiment one. For the hearing-impaired subjects, SRTs reached a plateau at a lower value of CO/*N*. SRTs did not differ significantly for CO/$N=0.25$, 0.5, 0.75 and 1 (for $N=8$) or for CO/$N=0.5$, 0.75 and 1 (for $N=16$). For $N=16$, the SRT for the normal-hearing subjects decreased by 16.0 dB as CO/*N* was increased from 0 to 1. This is similar to, but slightly larger than the decrease found in experiment one for 32-channel processing. For $N=8$, the decrease was larger, at 22.0 dB, because the SRT was higher for $N=8$ than for $N=16$ when the signal was fully vocoded (CO/$N=0$). For the hearing-impaired subjects, the decrease in SRT with increasing CO/*N* was much smaller, 4.1 dB for $N=16$ and 6.4 dB for $N=8$. Thus, as found in experiment one, the benefit of progressively adding TFS information (by increasing CO/*N*) was much smaller for the hearing-impaired than for the normal-hearing subjects, despite the use of a tone vocoder and a smaller number of channels in experiment two.

Fisher LSD tests revealed that normal-hearing subjects performed better for $N=16$ than for $N=8$, except when CO/$N\geq 0.75$ (see Table V). Hearing-impaired subjects performed better for $N=16$ when CO=0, but not for higher values of CO. For CO/$N=0.25$, performance was significantly better for $N=8$ than $N=16$ for the hearing-impaired subjects, and when CO/$N\geq 0.5$, there was no significant difference in performance for $N=8$ and 16.

Three student's *t* tests were performed to assess whether there was a significant effect of number of channels ($N=8$, 16 or 32) when CO=0 (i.e., when the signal was completely vocoded), for those hearing-impaired and normal-hearing subjects who were tested in all conditions. A Bonferroni cor-

TABLE III. As Table I, but for experiment two, which used two values of *N*. The LSD calculated using Fisher's LSD procedure was 2.3.

		$N=8$					$N=16$				
CO/*N*		0	0.25	0.5	0.75	1	0	0.25	0.5	0.75	1
	Freq. (Hz)	100	548	1605	4102	10,000	100	548	1605	4102	10,000
0	100	0					0				
0.25	548	**13.7**	0				**11.1**	0			
0.5	1605	**18.2**	**4.6**	0			**14.1**	**3.0**	0		
0.75	4102	**21.0**	**7.4**	**2.8**	0		**14.5**	**3.4**	0.4	0	
1	10,000	**22.0**	**8.3**	**3.8**	1.0	0	**15.2**	**4.1**	1.1	0.7	0

TABLE IV. As Table III, but for hearing-impaired subjects.

		$N=8$					$N=16$				
CO/N		0	0.25	0.5	0.75	1	0	0.25	0.5	0.75	1
	Freq. (Hz)	100	548	1605	4102	10,000	100	548	1605	4102	10,000
0	100	0					0				
0.25	548	**5.3**	0				1.4	0			
0.5	1605	**5.9**	0.6	0			**3.2**	1.8	0		
0.75	4102	**7.1**	1.8	1.2	0		**3.6**	2.0	0.1	0	
1	10,000	**6.4**	1.1	0.5	0.7	0	**4.2**	**2.8**	1.0	0.8	0

rection for multiple comparisons was applied. The normal-hearing subjects performed significantly better when $N=32$ than when $N=16$ ($p=0.01$), whereas the hearing-impaired subjects did not perform significantly differently for the two conditions ($p=0.12$). However, the SRT of the hearing-impaired subjects did increase when the value of N was decreased further to 8, and the difference in SRT between 32 and 8 channels was significant ($p=0.05$).

Performance in the condition when $N=32$ and CO=0 was much better than for the same condition in experiment one, for both normal-hearing and hearing-impaired subjects (the mean SRTs were 7.8 and 4.7 dB lower, respectively, in experiment two).

D. Discussion

The pattern of results was similar for experiments one and two, for both the normal-hearing and hearing-impaired subjects. This suggests that neither the random amplitude fluctuations introduced by the noise vocoder nor the partial removal of high-rate envelope modulations by the relatively narrow filters used in experiment one entirely explain the (small) benefit of adding TFS information found for the hearing-impaired subjects in experiment one. Rather, the results are consistent with the idea that the improvement in SRT as CO increased resulted mainly from the use of TFS information, and the improvement was smaller for the hearing-impaired than for the normal-hearing subjects because the latter have a greatly reduced ability to use TFS information.

Performance was better for the tone vocoder (experiment two) than for the noise vocoder (experiment one) with matched N, but different sentence material was used for the two experiments, which makes the comparison difficult. Better performance has been reported for the ASL sentence lists used in experiment one than for the BKB lists used in experiment two (MacLeod and Summerfield, 1990). If the same sentence material had been used for the two experiments, an even larger difference might have been observed. Overall, the comparison of results for experiments one and two with $N=32$ and CO=0 is consistent with previous results (Dorman *et al.*, 1997; Whitmal *et al.*, 2007) and with the hypothesis that the random amplitude fluctuations introduced by the noise vocoder have a deleterious effect on performance.

Normal-hearing subjects benefited from the greater spectral information in the vocoded signal when $N=32$ than when $N=16$, whereas the hearing-impaired subjects did not. This is consistent with what would be expected from the greater auditory-filter bandwidths that are typically found for hearing-impaired subjects (Glasberg and Moore, 1986; Moore, 2007). The normal-hearing subjects, who were expected to have relatively sharp filters, benefited significantly from the greater spectral information provided by more channels, while the hearing-impaired subjects, who were expected to have relatively broad filters, benefited little, if at all. These findings are consistent with those of Baskent (2006), who found a similar plateau in performance as N increased above 16 for hearing-impaired subjects.

Previous work has concentrated on reduced frequency selectivity as an explanation for the supra-threshold deficits associated with moderate cochlear hearing loss. Reduced frequency selectivity means that hearing-impaired listeners are more susceptible to masking across frequencies, and this partially explains why they perform poorly when listening in background sounds. The different patterns of performance for the normal-hearing and hearing-impaired subjects in the results presented here cannot be accounted for by differences in across-frequency masking. Similar amounts of masking would be expected in all of the conditions that were tested, so if deficits caused by cochlear hearing loss were only a result of across-frequency masking, a similar pattern of performance would have been expected for the normal-hearing and hearing-impaired subjects. Reduced spectral resolution may account for the differences in performance between the

TABLE V. Differences between mean SRTs measured for $N=8$ and $N=16$ for each value of CO/N, for normal-hearing and hearing-impaired subjects in experiment two. The LSD calculated using Fisher's LSD procedure was 1.7. Differences equal to or above this value are shown in bold.

CO/N	0	0.25	0.5	0.75	1
Freq. (Hz)	100	548	1605	4102	10,000
Normal hearing	**6.0**	**3.5**	**2.0**	0.5	0.7
Hearing impaired	**2.3**	**1.7**	0.4	1.5	0.1

subject groups when CO=0, when no TFS information was available. Indeed, the fact that the hearing-impaired subjects were tested using higher overall sound levels than the normal-hearing subjects might have exacerbated this effect, since auditory filters tend to broaden at high levels (Glasberg and Moore, 1990). However, changes in auditory-filter bandwidth with level tend to be smaller for hearing-impaired than for normally hearing subjects (Moore, 2007), so the effect of level is unlikely to be large. For whatever reasons, speech intelligibility worsens at very high sound levels for both normal-hearing and hearing-impaired subjects (Summers and Cord, 2007), so the higher level used here for the hearing-impaired subjects may have contributed to their poorer overall performance. However, the increasing deficit as TFS information was added is unlikely to reflect this "rollover effect," since the speech level was the same for all values of CO.

Another possible factor that may have influenced our results is that some of the hearing-impaired subjects may not have been able to make effective use of information conveyed by the higher-frequency components in the speech, even though those components would have been audible. In other words, the lack of benefit from adding TFS information may reflect a general lack of ability to use information from the higher-frequency components in speech. However, a reduced ability to use information from the higher-frequency (>2000 Hz) components in speech has mainly been found for subjects with hearing losses greater than about 60 dB (Ching *et al.*, 1998; Hogan and Turner, 1998; Vickers *et al.*, 2001). Hearing-impaired subjects with hearing losses less than 60 dB do seem to be able to make effective use of information from such high-frequency components (Skinner and Miller, 1983; Vickers *et al.*, 2001; Baer *et al.*, 2002). Several of our subjects had hearing losses of 60 dB or less for frequencies up to about 4 kHz, but they still failed to show a clear benefit as CO/*N* was increased above 0.5 (corresponding to a frequency of 1600 Hz). For example, HI 4 had audiometric thresholds of 55 dB or better for all frequencies up to 6000 Hz, but did not show any benefit of increasing CO/*N*.

Overall, it seems likely that the increasing deficit of the hearing-impaired subjects as CO/*N* was increased reflects a different ability to use TFS information between the two groups. It is possible that reduced frequency selectivity may contribute to a reduced ability to use TFS information. The outputs of broader auditory filters would have a more complex TFS than the outputs of narrower filters, as found in normal-hearing subjects. It is possible that such complex outputs may not be interpretable by the central auditory system. Deficits in phase locking would also be expected to reduce the ability to use TFS, as inaccuracies in phase locking would degrade information about TFS available to the central auditory system. Understanding the mechanism responsible for the observed deficit in the ability to use TFS would be an interesting topic for future research.

The individual differences in benefit from the addition of TFS information found here between hearing-impaired subjects may explain the relatively poor correlation between audiometric thresholds and speech intelligibility in noise (Festen and Plomp, 1983; Glasberg and Moore, 1989). For the subjects tested here, the amount of benefit gained from addition of TFS information [(SRT for CO=0)–(SRT for CO=32)] was not significantly correlated with the mean of audiometric thresholds at 250, 500, 1000, 2000 and 4000 Hz ($r=-0.04$, $p=0.92$). The ability to use TFS information may be a factor affecting speech intelligibility that is not well predicted by traditional audiometry.

VII. CONCLUSIONS

Hearing-impaired subjects benefited less than normal-hearing subjects from TFS information added to a vocoded speech signal when listening in a competing talker background. The amount of benefit varied between subjects, with some not benefiting at all. The same general pattern of results was found regardless of whether a noise vocoder or a tone vocoder was used. It is argued that subjects with moderate cochlear hearing loss have a limited ability to use TFS information, especially for medium and high frequencies. This may explain some of the speech perception deficits found for such subjects, especially the reduced ability to take advantage of temporal dips in a competing background.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Christian Lorenzi and one anonymous reviewer for helpful comments on an earlier version of this paper.

[1]Normally, the rectification would be followed by lowpass filtering, or the Hilbert transform would be used to extract the envelope. The omission of this stage in our processing meant that the modulator contained high-frequency components related to the TFS of the signal. However, these high-frequency components resulted in sidebands that were removed by the subsequent bandpass filtering. Listening tests and physical measurements confirmed that the processing used here gave results that were almost identical to those obtained when the Hilbert transform was used to extract the envelope.

ANSI (**1997**). *ANSI S3.5-1997, Methods for the Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).
Assmann, P. F., and Summerfield, A. Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.
Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.
Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.
Baer, T., Moore, B. C. J., and Kluk, K. (**2002**). "Effects of lowpass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies," J. Acoust. Soc. Am. **112**, 1133–1144.
Baskent, D. (**2006**). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels," J. Acoust. Soc. Am. **120**, 2908–2925.
Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).
Bracewell, R. N. (**1986**). *The Fourier Transform and its Applications* (McGraw–Hill, New York), pp. 267–272.
Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.
Buss, E., Hall, J. W., III, and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Ching, T., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.
Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.
de Boer, E. (**1956**). "Pitch of inharmonic signals," Nature (London) **178**, 535–536.
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.
Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897–2907.
Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.
Duquesnoy, A. J., and Plomp, R. (**1983**). "The effect of a hearing-aid on the speech-reception threshold of hearing-impaired listeners in quiet and in noise," J. Acoust. Soc. Am. **73**, 2166–2173.
Festen, J. M., and Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am. **73**, 652–662.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Finney, D. J. (**1971**). *Probit Analysis* (Cambridge University Press, Cambridge).
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., 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.
Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). "Masking release for consonant features in temporally fluctuating background noise," Hear. Res. **211**, 74–84.
George, E. L., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.
Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.
Gilbert, G., and Lorenzi, C. (**2006**). "The ability of listeners to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am. **119**, 2438–2444.
Glasberg, B. R., and Moore, B. C. (**2006**). "Prediction of absolute thresholds and equal-loudness contours using a modified loudness model," J. Acoust. Soc. Am. **120**, 585–588.
Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.
Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Harrison, R. V., and Evans, E. F. (**1979**). "Cochlear fiber responses in guinea pigs with well defined cochlear lesions," Scand. Audiol. Suppl. **9**, 83–92.
Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.
Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.
Lacher-Fougère, S., and Demany, L. (**1998**). "Modulation detection by normal and hearing-impaired listeners," Audiology **37**, 109–121.
Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.
Liberman, M. C., and Kiang, N. Y. S. (**1978**). "Acoustic trauma in cats: Cochlear pathology and auditory-nerve activity," Acta Oto-Laryngol., Suppl. **358**, 1–63.
Licklider, J. C. R., and Pollack, I. (**1948**). "Effects of differentiation, integration and infinite peak clipping upon the intelligibility of speech," J. Acoust. Soc. Am. **20**, 42–52.
Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). "Spatial cross correlation: A proposed mechanism for acoustic pitch perception," Biol. Cybern. **47**, 149–163.
Lorenzi, C., Husson, M., Ardoint, M., and Debruille, X. (**2006a**). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception," Int. J. Audiol. **45**, 487–495.
Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006b**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.
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.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).
Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, 2nd ed. (Wiley, Chichester).
Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.
Moore, B. C. J., and Moore, G. A. (**2003**). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," Hear. Res. **182**, 153–163.
Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.
Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**, 327–335.
Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (**1998**). "Development and evaluation of a procedure for fitting multi-channel compression hearing aids," Br. J. Audiol. **32**, 177–195.
Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.
Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**2004**). "New version of the TEN test with calibrations in dB HL," Ear Hear. **25**, 478–487.
Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1–15.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Pichora-Fuller, M. K. (**2003**). "Cognitive aging and auditory information processing," Int. J. Audiol. **42** Suppl. 2, 2S26–32.
Pichora-Fuller, M. K., Schneider, B. A., Benson, N. J., Hamstra, S. J., and Storzer, E. (**2006**). "Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry," J. Acoust. Soc. Am. **119**, 1143–1155.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "I.E.E.E. recommended

practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.
Schwartz, M. (**1970**). *Information Transmission, Modulation, and Noise* (McGraw–Hill, Kogakusha, Tokyo).
Sek, A., and Moore, B. C. J. (**1995**). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**, 2479–2486.
Shamma, S. A. (**1985**). "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Skinner, M. W., and Miller, J. D. (**1983**). "Amplification bandwidth and intelligibility of speech in quiet and noise for listeners with sensorineural hearing loss," Audiology **22**, 253–279.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.
Stone, M. A., and Moore, B. C. J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am. **114**, 1023–1034.
Stone, M. A., and Moore, B. C. J. (**2004**). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," J. Acoust. Soc. Am. **116**, 2311–2323.
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.
Summers, V., and Cord, M. T. (**2007**). "Intelligibility of speech in noise at high presentation levels: Effects of hearing loss and frequency region," J. Acoust. Soc. Am. **112**, 1130–1137.
Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1410–1421.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.
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.
Vickers, D. A., Moore, B. C. J., and Baer, T. (**2001**). "Effects of lowpass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies," J. Acoust. Soc. Am. **110**, 1164–1175.
Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Heifer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.
Woolf, N. K., Ryan, A. F., and Bone, R. C. (**1981**). "Neural phase-locking properties in the absence of outer hair cells," Hear. Res. **4**, 335–346.
Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403.

A probabilistic framework for landmark detection based on phonetic features for automatic speech recognition[a)]

Amit Juneja[b)] and Carol Espy-Wilson
Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742

(Received 15 November 2007; accepted 20 November 2007)

A probabilistic framework for a landmark-based approach to speech recognition is presented for obtaining multiple landmark sequences in continuous speech. The landmark detection module uses as input acoustic parameters (APs) that capture the acoustic correlates of some of the manner-based phonetic features. The landmarks include stop bursts, vowel onsets, syllabic peaks and dips, fricative onsets and offsets, and sonorant consonant onsets and offsets. Binary classifiers of the manner phonetic features—syllabic, sonorant and continuant—are used for probabilistic detection of these landmarks. The probabilistic framework exploits two properties of the acoustic cues of phonetic features—(1) sufficiency of acoustic cues of a phonetic feature for a probabilistic decision on that feature and (2) invariance of the acoustic cues of a phonetic feature with respect to other phonetic features. Probabilistic landmark sequences are constrained using manner class pronunciation models for isolated word recognition with known vocabulary. The performance of the system is compared with (1) the same probabilistic system but with mel-frequency cepstral coefficients (MFCCs), (2) a hidden Markov model (HMM) based system using APs and (3) a HMM based system using MFCCs. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2823754]

PACS number(s): 43.72.Ne, 43.72.Bs, 43.72.Ar [ADP] Pages: 1154–1168

I. INTRODUCTION

In a landmark-based automatic speech recognition (ASR) system, the front-end processing (or low-level signal analysis) involves the explicit extraction of speech-specific information. This speech-specific information consists of the acoustic correlates of the linguistic features (Chomsky and Halle, 1968) which comprise a phonological description of the speech sounds. The processing occurs in two steps. The first step consists of the automatic detection of acoustic events (also called landmarks) that signal significant articulatory changes such as the transition from a more open to a more closed vocal tract configuration and vice versa (i.e., changes in the manner of articulation), a sudden release of air pressure and changes in the state of the larynx. There is evidence that the auditory system responds in a distinctive way to such acoustic events (e.g., Delgutte and Kiang, 1984). The second step involves the use of these landmarks to extract other relevant acoustic information regarding place of articulation that helps in the classification of the sounds spoken. Given the extensive variability in the speech signal, a complete ASR system would integrate this front-end processing with a lexical access system that handles pronunciation variability and takes into account prosody, grammar, syntax and other higher-level information.

State-of-the-art ASR systems are based on hidden Markov modeling (HMM) and the standard parametrization of the speech signal consists of Mel-frequency cepstral coefficients (MFCCs) and their first and second derivatives (Rabiner and Juang, 1993; Young *et al.*, 2006). The HMM framework assumes independence of the speech frames so that each one is analyzed and all of the MFCCs are looked at in every frame. In contrast, a landmark-based approach to speech recognition can target level of effort where it is needed. This efficiency can be seen in several ways. First, while each speech frame may be analyzed for manner-of-articulation cues resulting in the landmarks, analysis thereafter is carried out only at significant locations designated by the landmarks. This process in effect takes into account the strong correlation among the speech frames. Second, analysis at different landmarks can be done with different resolutions. For example, the transient burst of a stop consonant may be only 5 ms long. Thus, a short temporal window is needed for analysis. On the other hand, vowels which are considerably longer (50 ms for a /schwa/ to 300 ms for an /ae/) need a longer analysis window. Third, the acoustic parameters (APs) used to extract relevant information will depend upon the type of landmark. For example, at a burst landmark, appropriate APs will be those that characterize the spectral shape of the burst (maybe relative to the vowel to take into account contextual influences) to distinguish between labial, alveolar and velar stops. However, at a vowel landmark, appropriate APs will be those that look at the relative spacing of the first three formants to determine where

[a)] Portions of this work have appeared in "Segmentation of Continuous Speech Using Acoustic-Phonetic Parameters and Statistical Learning," International Conference on Neural Information Processing, Singapore, 2002, and "Speech segmentation using probabilistic phonetic feature hierarchy and support vector machines," International Joint Conference on Neural Networks, Portland, Oregon, 2003, and "Significance of invariant acoustic cues in a probabilistic framework for landmark-based speech recognition," in the proceedings of From Sound to Sense: 50+ Years of Discoveries in Speech Communication, June 11–13, 2004, MIT, Cambridge, MA.

[b)] Author to whom correspondence should be addressed. Electronic mail: amjuneja@gmail.com

the vowel fits in terms of the phonetic features *front*, *back*, *high* and *low*.

Another prominent feature of a landmark ASR system is that it is a tool for uncovering and subsequently understanding variability. Given the physical significance of the APs and a recognition framework that uses only the relevant APs, error analysis often points to variability that has not been accounted for. For example, an early implementation of the landmark-based ASR system (Bitar, 1997) used zero crossing rate as an important measure to capture the turbulent noise of strident fricatives. The zero crossing rate will not be large, however, during a voiced strident fricative /z/ when it contains strong periodicity. In this case, the high-frequency random fluctuations are modulated by the low-frequency periodicity. This situation occurs when the /z/ is produced with a weakened constriction so that the glottal source is comparatively stronger (and, therefore, the supraglottal source is weaker) than it is during a more canonically produced /z/. Spectrographically, a weakened /z/ shows periodic formant structure at low frequencies like a sonorant consonant and some degree of turbulence at high frequencies like an obstruent. This understanding led to the development of an aperiodicity/periodicity/pitch (APP) detector which, along with fundamental frequency information, provides a spectro-temporal profile of aperiodicity and periodicity (Deshmukh *et al.*, 2005). The APP detector, however, was not used in this work due its computational requirements.

A significant amount of work has gone into understanding the acoustic correlates of the linguistic features (Stevens, 2002). Studies have shown that the acoustic correlates of the phonetic features can be reliably and automatically extracted from the speech signal (Espy-Wilson, 1987; Bitar, 1997; Ali, 1999; Carbonell *et al.*, 1987; Glass, 1984; Hasegawa-Johnson, 1996) and that landmarks can be automatically detected (Bitar, 1997; Salomon, 2000; Ali, 1999; Liu, 1996). Stevens (2002) has laid out a model for lexical access based on acoustic landmarks and phonetic features. However, to date, no one has implemented a complete ASR system based on a landmark approach. The previous landmark-detection systems (Bitar, 1997; Salomon, 2000; Ali, 1999; Liu, 1996) performed well, but they lacked a probabilistic framework for handling pronunciation variability that would make the systems scalable to large-vocabulary recognition tasks where higher-level information has to be integrated. For example, it could not be demonstrated in these systems that a voiced obstruent realized as a sonorant consonant will ultimately be recognized correctly due to higher-level constraints. These systems were primarily rule based and it has been pointed out (Rabiner and Juang, 1993) that in rule-based systems, the difficulty in the proper decoding of phonetic units into words and sentences increases sharply with an increase in the rate of phoneme insertion, deletion and substitution. In this work, a probabilistic framework is developed that selectively uses knowledge-based APs for each decision and it can be constrained by a high-level pronunciation model of words and probability densities of durations of phonetic units. Since recognition can be constrained by higher-level knowledge, the system does not have to decode phonetic units into words in a separate step.

Probabilistic frameworks exist for segment-based (Glass *et al.*, 1996; Zue *et al.*, 1989; Halberstadt, 1998) and syllable-based (Chang, 2002) ASR. But these systems are not targeted at selectively using knowledge-based acoustic correlates of phonetic features for detection of landmarks or for place of articulation detection. Many HMM-based approaches to speech recognition have used knowledge-based APs (Bitar and Espy-Wilson, 1996; Deshmukh *et al.*, 2002; Hosom, 2000), or the concept of phonetic features (Deng and Sun, 1994; Kirchhoff, 1999; Eide *et al.*, 1993). However, these were not landmark-based methods in that they did not involve an initial step of segmenting or detecting events in speech.

In this paper, a probabilistic framework for a landmark-based ASR system called event-based system (EBS) (Bitar, 1997) is presented. The focus of this paper is on the implementation and performance of the probabilistic landmark detection module of the framework. The framework was discussed in brief earlier (Juneja and Espy-Wilson, 2004) but it was not sufficiently detailed because of the limited space available. Initial results during the development of the probabilistic landmark detection system were reported earlier (Juneja and Espy-Wilson, 2002, 2003) but these only involved statistical classifiers for phonetic features and did not involve the probabilistic scoring with duration constraints. Because of the lack of probabilistic duration constraints these initial systems resulted in numerous insertions of segments of very small durations (5–10 ms) and such small segments of continuant sounds had to be removed to get a good recognition accuracy. Also because of the lack of probabilistic scoring, these systems could not be constrained by pronunciation models. In this work, the complete probabilistic framework for landmark detection is described in detail along with the details of the experiments.

The performance of the framework is demonstrated with the broad-class recognition and landmark detection task using the TIMIT database (NIST, 1990) and a vocabulary-constrained broad class recognition task on isolated digits using the TIDIGITS database (LDC, 1982). These recognition tasks require only the first step of the front-end processing where information is sought regarding only the manner features. For comparison, a traditional HMM-based recognition system is implemented. In one case, the traditional MFCCs and their first and second derivatives serve as input to the HMM system. In another implementation, the same APs used in EBS serve as input to the HMM system. Finally, the APs in the EBS system are replaced with the MFCCs. This four-way comparison allows the evaluation of the effectiveness of not only the probabilistic landmark framework as compared to the statistical HMM system, but also the knowledge-based APs with the MFCCs.

II. METHOD

A. Overview

Figure 1 shows a block diagram of EBS and it highlights the part of EBS that is the focus of this paper. To start the landmark detection process, the knowledge-based APs (Bitar, 1997) shown in Table I for each of the phonetic features—

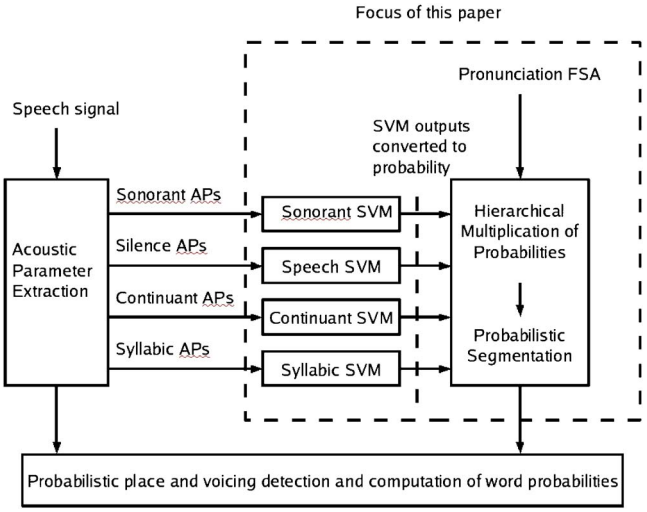


FIG. 1. (Color online) Overview of the landmark-based speech recognition system.

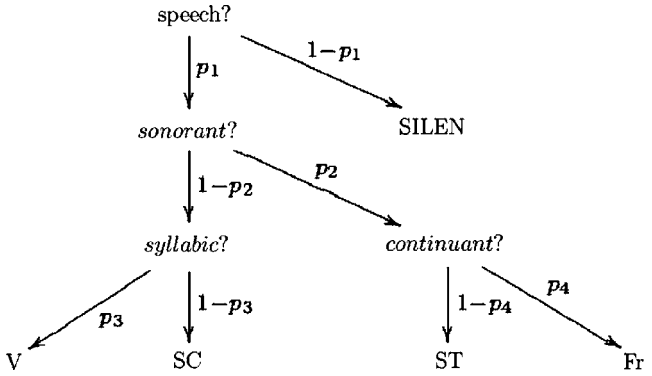


FIG. 2. Probabilistic Phonetic Feature Hierarchy.

sonorant, *syllabic*, *continuant*—and silence are automatically extracted from each frame of the speech signal. Then, a support vector machine (SVM) (Vapnik, 1995) based binary classifier is applied at each node of the hierarchy shown in Fig. 2 such that only the relevant APs for the feature at that node serve as input to the classifier. Probabilistic decisions obtained from the outputs of SVMs are combined with class dependent duration probability densities to obtain one or more segmentations of the speech signal into the broad classes—vowel (V), fricative (Fr), sonorant consonant (SC—including nasals semivowels), stop burst (ST) and silence (SILEN—including stop closures). A segmentation is then used along with the knowledge-based measurements to deterministically find landmarks related to each of the broad class segments. For a fixed vocabulary, segmentation paths can be constrained using broad class pronunciation models.

The phonetic feature hierarchy shown in Fig. 2 is the upper part of a complete hierarchy that has manner features at the top, place features at the lower nodes and phonemes at the lowest level. Several studies have provided evidence for a hierarchical organization of the phonetic features (e.g., Clements, 1985). Probabilistic hierarchies with phonemes at the terminal nodes have been used before in speech recognition (Halberstadt, 1998; Chun, 1996) where the use of such hierarchies occurs after an initial segmentation step. EBS uses the hierarchy as a uniform framework for obtaining manner-based landmarks and place and voicing feature detection. The complete hierarchy used by EBS is shown in Juneja, 2004.

Figure 3 shows the two kinds of landmarks that EBS is expected to extract (the landmark locations in this figure are hand marked)—abrupt landmarks and nonabrupt landmarks. In the previous implementation of EBS (Bitar, 1997), the maxima and minima of the APs $E[640\ \text{Hz}, 2800\ \text{Hz}]$ and $E[2000\ \text{Hz}, 3000\ \text{Hz}]$ were used in a rule-based system to obtain the nonabrupt landmarks that occur at syllabic peaks and syllabic dips. Thresholds on energy onsets and energy offsets were used to obtain the abrupt stop burst landmarks. Figure 3 shows that $E[640\ \text{Hz}, 2800\ \text{Hz}]$ has maxima in the vowel nuclei at the syllabic peak landmarks and minima in the sonorant consonant regions at the syllabic dip landmarks. Spurious peaks or dips caused insertions and peaks or dips that are not fully realized caused deletions in this system. There was no way to recover from such errors.

The presented framework can be viewed as a probabilistic version of the system in (Bitar, 1997) as it finds the landmarks using the following two steps:

1. The system derives multiple probabilistic segmentations from statistical classifiers (that use relevant APs as input) taking into account the probability distributions of the durations of the broad class segments. The probabilistic duration models penalize the insertions of very small broad class segments (for example, 5–10-ms-long vowels) by assigning low duration probabilities to such sounds.

TABLE I. APs used in broad class segmentation. f_s: sampling rate F3: third formant average [a,b]: frequency band [aHz,bHz], E[a,b]: energy in the frequency band [aHz,bHz].

Phonetic Feature	APs
Silence	(1) $E[0, F3\text{-}1000]$, (2) $E[F3, f_s/2]$, (3) ratio of spectral peak in [0,400 Hz] to the spectral peak in $[400, f_s/2]$, (4) Energy onset (Bitar, 1997) (5) Energy offset (Bitar, 1997)
sonorant	(1) $E[0, F3\text{-}1000]$, (2) $E[F3, f_s/2]$, (3) Ratio of $E[0, F3\text{-}1000]$ to $E[F3\text{-}1000, f_s/2]$, (4) $E[100,400]$
syllabic	(1) $E[640, 2800]$ (2) $E[2000, 3000]$ (3) Energy peak in [0,900 Hz] (4) Location in Hz of peak in [0,900 Hz]
continuant	(1) Energy onset (Bitar, 1997), (2) Energy offset (Bitar, 1997), (3) $E[0, F3\text{-}1000]$, (4) $E[F3\text{-}1000, f_s/2]$

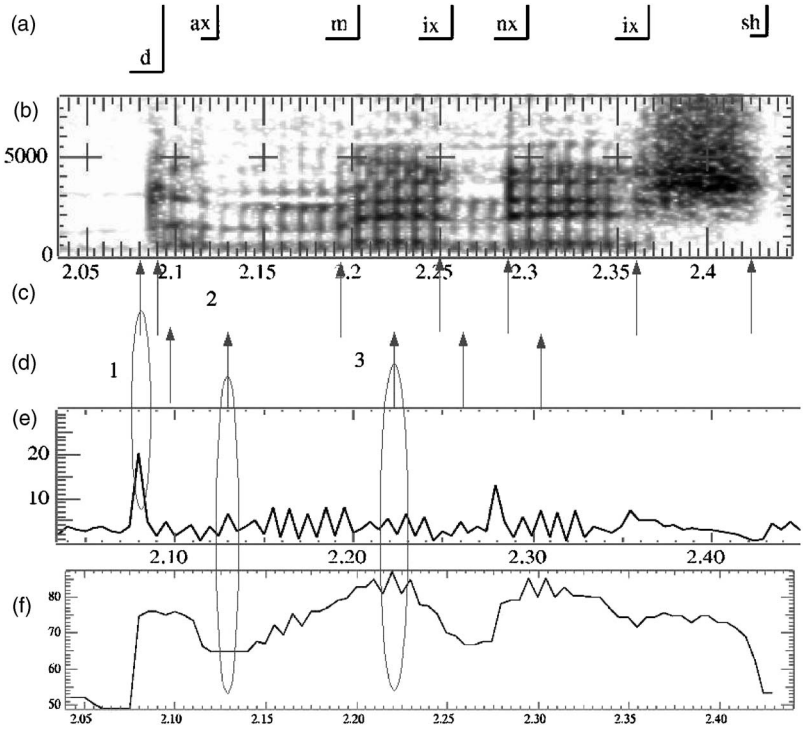


FIG. 3. Illustration of manner landmarks for the utterance "diminish" from the TIMIT database. (a) Phoneme Labels, (b) Spectrogram, (c) Landmarks characterized by sudden change, (d) Landmarks in stable regions of speech, (e) Onset waveform (an acoustic correlate of phonetic feature *−continuant*), (f) $E[640,2800]$ (an acoustic correlate of *syllabic* feature). The ellipses show how landmarks were obtained in (Bitar, 1997) using certain APs. Ellipse 1 shows the location of stop burst landmark for the consonant /d/ using the onset. Ellipse 2 shows the location of syllabic dip for the nasal /m/ using the minimum of $E[640,2800]$. Ellipse 3 shows that the maximum of the $E[640,2800]$ can be used to locate a syllabic peak landmark of the vowel /ix/.

2. The landmarks are then derived using the boundaries of the broad classes as abrupt landmarks and the maxima and minima of the AP $E[640\ \text{Hz},2800\ \text{Hz}]$ inside individual sonorant segments to get the nonabrupt landmarks. The global maximum inside the vowel segment is used to get the syllabic peak landmark and the global minima inside the sonorant consonant is used to get the syllabic dip landmark. Therefore, presence of multiple peaks or dips does not cause insertions at this step.

SVMs are used for the purpose of binary classification (although other classifiers, for example, neural networks or Gaussian mixture models could be used as well) of phonetic features because of their ability to generalize well to new test data after learning from a relatively small amount of training data. Additionally, SVMs have been shown to perform better than HMMs for phonetic feature detection in speech (Niyogi, 1998; Keshet *et al.*, 2001) and for phonetic classification from hand-transcribed segments (Clarkson and Moreno, 1999; Shimodaira *et al.*, 2001). The success of SVMs can be attributed to their property of large margin classification. Figure 4 shows two types of classifiers for linearly separable data: (1) a linear classifier without maximum margin and (2) a linear classifier with maximum margin. It can be seen from Fig. 4 that the maximum margin classifier is more robust to noise because a larger amount of noise (at least half of the margin for the samples shown) is required to let a sample point cross the decision boundary. It has been argued (Vapnik, 1995) that a maximization of the margin leads to the minimization of a bound on the test error. Mathematically, SVMs select a set of N_{SV} support vectors $\{\mathbf{x}_i^{\text{SV}}\}_{i=1}^{N_{\text{SV}}}$ that is a subset of l vectors in the training set $\{\mathbf{x}_i\}_{i=1}^{l}$ with class labels $\{y_i\}_{i=1}^{l}$, and find an optimal separating hyperplane $f(\mathbf{x})$ (in the sense of maximization of margin) in a high dimensional space $\mathcal{H}$,

$$f(\mathbf{x}) = \sum_{i=1}^{N_{\text{SV}}} y_i \alpha_i K(\mathbf{x}_i^{\text{SV}}, \mathbf{x}) - b. \tag{1}$$

The space $\mathcal{H}$ is defined by a linear or nonlinear kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ that satisfies the Mercer conditions (Burges, 1998). The weights α_i, the set of support vectors $\{\mathbf{x}_i^{\text{SV}}\}_{i=1}^{N_{\text{SV}}}$ the

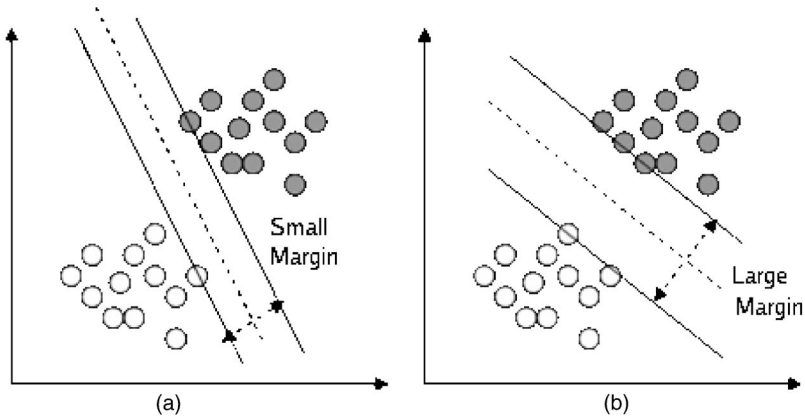


FIG. 4. (a) small margin classifiers, (b) maximum margin classifiers.

TABLE II. An illustrative example of the symbols B, L and U.

	/z/	/ɪ/	/r/	/o/	/w/
$U\Rightarrow$	u_1	u_2	u_3	u_4	u_5
	−sonorant	*+sonorant*	*+sonorant*	*+sonorant*	*+sonorant*
	+continuant	*+syllabic*	*−syllabic*	*+syllabic*	*−syllabic*
	+strident	*−back*	*−nasal*	*+back*	*−nasal*
	+voiced	*+high*	*+rhotic*	*−high*	*+labial*
	+anterior	*+lax*		*+low*	
$B\Rightarrow$	Fr	V	SC	V	SC
$L\Rightarrow$	l_1	l_2	l_3	l_4	l_5
	Fon	VOP	Son	VOP	Son
	Foff	P	D	P	D
			Soff		Soff

bias term b are found from the training data using quadratic optimization methods. Two commonly used kernels are the radial basis function (RBF) kernel and the linear kernel. For the RBF kernel, $K(\mathbf{x}_i,\mathbf{x})=\exp(-\gamma|\mathbf{x}_i-\mathbf{x}|^2)$ where the parameter γ is usually chosen empirically by cross validation from the training data. For the linear kernel, $K(\mathbf{x}_i,\mathbf{x})=\mathbf{x}_i.\mathbf{x}+1$.

B. Probabilistic framework

The problem of speech recognition can be expressed as the maximization of the posterior probability of sets of phonetic features where each set represents a sound or a phoneme. A set of phonetic features include (1) manner phonetic features represented by landmarks and (2) place or voicing phonetic features found using the landmarks. Mathematically, given an acoustic observation sequence O, the problem can be expressed as

$$\hat{U}\hat{L} = \arg\max_{U,L} P(U,L|O) = \arg\max_{U,L} P(L|O)P(U|O,L), \tag{2}$$

where $L=\{l_i\}_{i=1}^{M}$ is a sequence of landmarks and $U=\{u_i\}_{i=1}^{N}$ is the sequence bundles of features corresponding to a phoneme sequence. The meaning of these symbols is illustrated in Table II for the digit "zero." Computation of $P(L|O)$ is the process of probabilistic detection of acoustic landmarks given the acoustic observations and the computation of $P(U|L,O)$ is the process of using the landmarks and acoustic observations to make probabilistic decisions on place and voicing phonetic features. The goal of the rest of the paper is to show how $P(L|O)$ is computed for a given landmark sequence and how different landmark sequences and their probabilities can be found given an observation sequence.

There are several points to note with regards to the notation in Table II.

1. l_i denotes a set of related landmarks that occur during the same broad class. For example, the syllabic peak (P) and the vowel onset point (VOP) occur during a vowel. The VOP should occur at the start of the vowel and P should occur during the vowel when the vocal tract is most open. Also, certain landmarks may be redundant in the sequence. For example, when a vowel follows a sonorant consonant, the sonorant consonant offset and the vowel onset are identical.
2. Each set of landmarks l_i, as shown in Table III, is related to a broad class B_i of speech selected from the set: vowel (V), fricative (Fr), sonorant consonant (SC), stop burst (ST), silence (SILEN). For example, P and VOP are related to the broad class V. Let $B=\{B_i\}_{i=1}^{M}$ denote the sequence of broad classes corresponding to the sequence of sets of landmarks L. Note that, in this paper, that ST denotes the burst region of a stop consonant, and the closure region is assigned the broad class SILEN.
3. The number of broad classes M and the number of bundles of phonetic features N may not be the same in general. This difference may occur because a sequence of sets of landmarks and the corresponding broad class sequence may correspond to one set of phonetic features or two sets of phonetic features. For example, SILEN-ST could be the closure and release of one stop consonant, or it could be that the closure corresponds to one stop consonant and the release corresponds to another stop consonant (e.g., the cluster /kt/ in the word "vector"). Likewise, one set of landmarks or the corresponding broad class may correspond to two sets of place features. For example, in the word "omni" with the broad class sequence V-SC-V, the SC will have the features of the sound /m/ (calculated using the SC onset) as well as the sound /n/ (calculated using SC offset).

The landmarks and the sequence of broad classes can be obtained deterministically from each other. For example, the sequence $B=\{\text{SILEN},\text{Fr},\text{V},\text{SC},\text{V},\text{SC},\text{SILEN}\}$ for "zero" in Table II will correspond to the sequence of sets of landmarks L shown. Therefore

TABLE III. Landmarks and corresponding broad classes.

Broad class segment	Landmark type
Vowel (V)	Syllabic peak (P)
	Vowel onset point (P)
Stop (ST)	Burst
Sonorant consonant (SC)	Syllabic dip (D)
	SC onset (Son)
	SC offset (Soff)
Fricative (Fr)	Fricative onset (Fon)
	Fricative offset (Foff)

$$P(L|O) = P(B(L)|O), \tag{3}$$

where $B(L)$ is a sequence of broad classes for which the landmark sequence L is obtained. Note that the symbols B, U and L contain information about the order in which the broad classes or landmarks occur, but they do not contain information about the exact start and end times of each of those units. The equivalence of broad classes and landmarks is not intended as a general statement and it is assumed to hold only for the landmarks and broad classes shown in Table III.

C. Segmentation using manner phonetic features

Given a sequence of T frames $O=\{o_1, o_2, \ldots, o_T\}$, where o_t is the vector of APs at time t (t is in the units of frame numbers), the most probable sequence of broad classes $B=\{B_i\}_{i=1}^{M}$ and their durations $D=\{D_i\}_{i=1}^{M}$ have to be found. The frame o_t is considered as the set of all APs computed at frame t to develop the probabilistic framework, although EBS does not use all of the APs in each frame. The probability $P(B|O)$ can be expressed as

$$P(B|O) = \sum_{D} P(B,D|O). \tag{4}$$

The computation of $P(B,D|O)$ for a particular B and all D is a very computationally intensive task in terms of storage and computation time. Therefore, an approximation is made that is similar to the approximation made by Viterbi decoding in HMM-based recognition systems and the SUMMIT system (Glass *et al.*, 1996),

$$P(B|O) \approx \max_{D} P(B,D|O). \tag{5}$$

Because the probabilities $P(B|O)$ calculated this way for different B will not add up to one, the more correct approximation is

$$P(B|O) \approx \frac{\max_D P(B,D|O)}{\sum_B \max_D P(B,D|O)}. \tag{6}$$

Provided that a frame at time t lies in the region of one of the manner classes, the posterior probability of the frame being part of a vowel at time t can be written as (see Fig. 2)

$$P_t(V|O) = P_t(+\,speech, +\,sonorant, +\,syllabic|O)$$
$$= P_t(+\,speech|O)P_t(+\,sonorant| + speech, O), \tag{7}$$

$$P_t(+\,syllabic| + sonorant, +\,speech, O), \tag{8}$$

where P_t is used to denote the posterior probability of a feature or a set of features at time t. A similar expression can be written for each of the other manner classes.

Calculation of the posterior probability for each feature requires only the acoustic correlates of that feature. Furthermore, to calculate the posterior probability of a manner phonetic feature at time t, only the acoustic correlates of the feature in a set of frames $\{t-s, t-s+1, \ldots, t+e\}$, using s previous frames and e following frames along with the current frame t, are required to be used. Let this set of acoustic correlates extracted from the analysis frame and the adjoining frames for a feature f be denoted by x_t^f. Then Eq. (8) can be rewritten as

$$P_t(V|O) = P_t(+\,speech|x_t^{speech})P_t(+\,sonorant| + speech, x_t^{sonorant})$$
$$P_t(+\,syllabic| + sonorant, +\,speech, x_t^{syllabic}). \tag{9}$$

The probability $P(B,D|O)$ can now be expanded in terms of the underlying manner phonetic features of each broad class. Denote the features for class B_i as the set $\{f_1^i, f_2^i, \ldots, f_{N_{B_i}}^i\}$, the broad class at time t as b_t, and the sequence $\{b_1, b_2, \ldots, b_t\}$ as b^t. Note that B is the broad class sequence with no information about the duration of each broad class in the sequence. On the other hand, b_t denotes a broad class at frame t. Therefore, the sequence b^t includes the information of durations of each of the broad classes until time t. Using this notation, the posterior probability of a broad class sequence B and durations D can be expanded as

$$P(B,D|O) = \prod_{i=1}^{M} \prod_{t=1+\Sigma_{j=1}^{i-1}D_j}^{D_i+\Sigma_{j=1}^{i-1}D_j} P_t(B_i|O, b^{t-1}). \tag{10}$$

The variable t in the above equation is the frame number, the limits of which can be explained as follows. $\Sigma_{j=1}^{i-1}D_j$ is the sum of the durations of the $i-1$ broad classes before the broad class i, and $\Sigma_{j=1}^{i}D_j$ is the sum of durations of the first i broad classes. Then, $\Sigma_{j=1}^{i}D_j - \Sigma_{j=1}^{i-1}D_j$ is the duration of the ith broad class. Therefore, numbers $\{1+\Sigma_{j=1}^{i-1}D_j, 2+\Sigma_{j=1}^{i-1}D_j, \ldots, D_i+\Sigma_{j=1}^{i-1}D_j\}$ are the frame numbers of the frames that occupy the ith broad class. When the lower and upper limits of t are specified as $1+\Sigma_{j=1}^{i-1}D_j$ and $D_i+\Sigma_{j=1}^{i-1}D_j$, respectively, it means that the product is taken over all the frames of the ith broad class.

Making a stronger use of the definition of acoustic correlates by assuming that the acoustic correlates of a manner feature at time t are sufficient even if b^{t-1} is given,

$$P(B,D|O) = \prod_{i=1}^{M} \prod_{t=1+\Sigma_{j=1}^{i-1}D_j}^{D_i+\Sigma_{j=1}^{i-1}D_j} \prod_{k=1}^{N_{B_i}} P_t(f_k^i|x_t^{f_{k^i}}, f_1^i, \ldots, f_{k-1}^i, b^{t-1}). \tag{11}$$

Now expanding the conditional probability,

$$P(B,D|O) = \prod_{i=1}^{M} \prod_{t=1+\Sigma_{j=1}^{i-1}D_j}^{D_i+\Sigma_{j=1}^{i-1}D_j} \prod_{k=1}^{N_{B_i}} \frac{P_t(f_k^i, x_t^{f_k^i}, f_1^i, \ldots, f_{k-1}^i, b^{t-1})}{P_t(x_t^{f_k^i}, f_1^i, \ldots, f_{k-1}^i, b^{t-1})}. \tag{12}$$

Splitting the priors gives

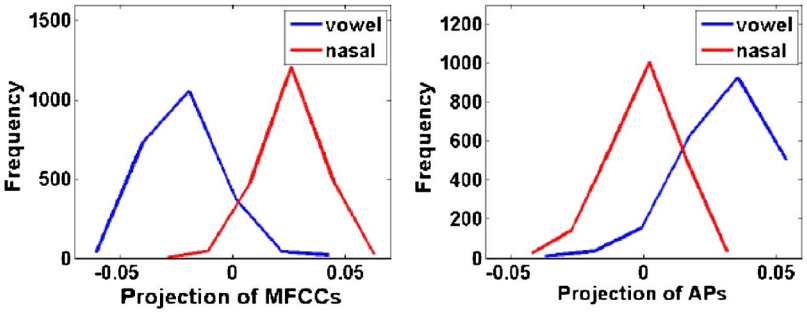


FIG. 5. (Color online) (a) Projection of 39 MFCCs into a one-dimensional space with vowels and nasals as discriminating classes, (b) similar projection for four APs used to distinguish +sonorant sounds from −sonorant sounds. Because APs for the sonorant feature discriminate vowels and nasals worse than MFCCs, they are more invariant.

$$P(B,D|O) = \prod_{i=1}^{M} \prod_{t=1+\Sigma_{j=1}^{i-1}D_j}^{D_i+\Sigma_{j=1}^{i-1}D_j} \prod_{k=1}^{N_{B_i}} P_t(f_k^i|f_1^i,\ldots,f_{k-1}^i,b^{t-1}) \times \frac{P_t(x_t^{f_k^i}|f_1^i,\ldots,f_k^i,b^{t-1})}{P_t(x_t^{f_k^i}|f_1^i,\ldots,f_{k-1}^i,b^{t-1})}. \quad (13)$$

The probability terms not involving the feature vector $x_t^{f_k^i}$ can now be combined to get the prior probabilities of the broad class sequence and the sequence dependent durations, that is,

$$\prod_{i=1}^{M} \prod_{t=1+\Sigma_{j=1}^{i-1}D_j}^{D_i+\Sigma_{j=1}^{i-1}D_j} \prod_{k=1}^{N_{B_i}} P_t(f_k^i|f_1^i,\ldots,f_{k-1}^i,b^{t-1}) = P(B,D) = P(B)P(D|B). \quad (14)$$

Now given the set $\{f_1^i,\ldots,f_{k-1}^i\}$ or the set $\{f_1^i,\ldots,f_k^i\}$, $x_t^{f_k^i}$ is assumed to be independent of b^{t-1}. This independence of the APs from the previous broad class frames is hard to establish, but it can be shown to hold better for the knowledge-based APs than for the mel-frequency cepstral coefficients (MFCCs) (see Fig. 5) under certain conditions as discussed in Sec. III B. In words, this independence means that given a phonetic feature or the phonetic features above that feature in the hierarchy, the APs for that phonetic feature are assumed to be invariant with the variation of the broad class labels of the preceding frames. For example, the APs for the feature *sonorant* in a *+sonorant* frame are assumed to be invariant of whether the frame lies after vowel, nasal or fricative frames. This is further discussed in Sec. III B. Making this independence or invariance assumption and applying Eq. (14) in Eq. (13),

$$P(B,D|O) = P(B)P(D|B)\prod_{i=1}^{M} \prod_{t=1+\Sigma_{j=1}^{i-1}D_j}^{D_i+\Sigma_{j=1}^{i-1}D_j} \times \prod_{k=1}^{N_{B_i}} \frac{P_t(x_t^{f_k^i}|f_1^i,\ldots,f_k^i)}{P_t(x_t^{f_k^i}|f_1^i,\ldots,f_{k-1}^i)}, \quad (15)$$

which can be rewritten as

$$P(B,D|O) = P(B)P(D|B)\prod_{i=1}^{M} \prod_{t=1+\Sigma_{j=1}^{i-1}D_j}^{D_i+\Sigma_{j=1}^{i-1}D_j} \times \prod_{k=1}^{N_{B_i}} \frac{P_t(f_k^i|x_t^{f_k^i},f_1^i,\ldots,f_{k-1}^i)}{P_t(f_k^i|f_1^i,\ldots,f_{k-1}^i)}. \quad (16)$$

The posterior $P_t(f_k^i|x_t^{f_k^i},f_1^i,\ldots,f_{k-1}^i)$ is the probability of the binary feature f_k^i obtained using the APs $x_t^{f_k^i}$ and it is obtained in this work from an SVM-based binary classifiers as described in Sec. I below. The term $P_t(f_k^i|x_1^i,\ldots,f_{k-1}^i)$ normalizes the imbalance of the number of positive and negative samples in the training data. The division on the right side of Eq. (16) can be considered as the conversion of a posterior probability to a likelihood by division by a prior. The prior is computed as the division of the number of training samples for the positive value of the feature to the number of training samples for the negative value of the feature.

1. Training and application of binary classifiers

One SVM classifier was trained for each of the phonetic features shown in Fig. 2. The input to the classifier is the set of APs shown in Table I for that feature. The sounds used to get the training samples of class +1 and class −1 for each SVM are shown in Table IV. For the feature *continuant*, the

TABLE IV. Sounds used in training of each classifier.

Phonetic feature	Sounds with +1 label	Sounds with −1 label
speech	All speech sounds excluding stop closures	Silence, pauses and stop closures
sonorant	Vowels, nasals and semivowels	Fricatives, affricates and stop bursts
syllabic	Vowels	Nasals and semivowels
continuant	Fricatives	Stop bursts

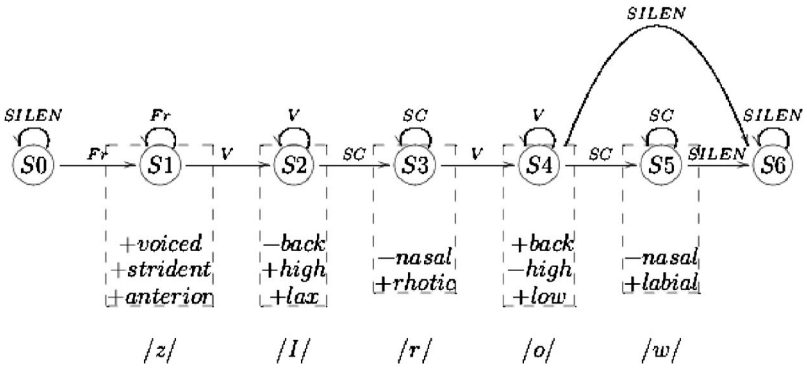


FIG. 6. A phonetic feature-based pronunciation model for the word "zero."

stop burst frame identified as the first frame of a stop consonant using TIMIT labeling was used to extract APs representative of the value −1 of that feature. For the +1 class of the feature *continuant*, the APs were extracted from all of the fricative frames. For the other features, all frames for each of the classes were extracted as training samples. Flap (/dx/), syllabic sonorant consonants (/em/, /el/, /en/, /er/ and /eng/) and diphthongs (/iy/, /ey/, /ow/, /ay/, /aw/, and /uw/) were not used in the training of the feature *syllabic*, and affricates (/jh/ and /ch/) and glottal stops were not used in training of the feature *continuant*, but these sounds were used for frame-based testing. The reason for not using these sounds for training is that they have different manifestations. For example, the affricates /ch/ and /jh/ may appear with or without a clear stop burst. However, such information is not marked in the TIMIT hand-transcribed labels.

SVM Light (Joachims, 1998), an open-source toolkit for SVM training and testing, was used for building and testing the classifiers. Two types of SVM kernels were used—linear and radial basis function (RBF)—to the build corresponding two types of classifiers. The optimal number of adjoining frames s and e used in each classifier as well as the optimal SVM related parameters (e.g., the bound on slack variables α_i and γ for RBF kernels) were found using cross validation over a separate randomly chosen data set from the TIMIT training set.

A SVM outputs a real number for a teat sample. To convert this real number into a probability, the real space of the SVM outputs is divided into 30 bins of equal sizes between −3 and +3. This range was chosen empirically from observations of many SVM outputs. After the SVMs are trained, the proportion of samples of class +1 to the total numbers of training samples in each of the bins is noted and the proportions for all of the bins are stored in a table. While testing, the bin corresponding to the real number obtained for a particular test sample is noted and its probability is looked up from the stored table.

2. Probabilistic segmentation

A Viterbi-like probabilistic segmentation algorithm (Juneja, 2004) takes as input the probabilities of the manner phonetic features—*sonorant*, *syllabic*, *continuant*—and silence from the SVM classifiers and outputs the probabilities $P(B|O)$ under the assumption of Eq. (5). The algorithm is similar to the one used by Lee (1998). The algorithm operates on the ratio of posterior probabilities on the right side of Eq. (16), unlike the algorithm developed by Lee (1998) where segment scores of observations in speech segments are used. Another difference is that the transition points in the segmentation algorithm in the current work are obtained as those frames at which the ranking of the posterior probabilities of the broad classes changes. In the work by Lee (1998), the transitions points were calculated from the points of significant change in the acoustic representation.

D. Deterministic location of landmarks

Once a broad class sequence with the start and end times of each of the broad classes is found, the landmarks are located deterministically. Fon and Foff are allotted the start and end times of the broad class Fr. Son and Soff are assigned the start and end times of the broad class SC. The stop burst B is found as the location of the maximum value of the temporal onset measure within a 60 ms window centered at the first frame of the segment ST. VOP is assigned the first frame of the segment V, and P is assigned the location of highest value of $E[640,2800]$ in the segment V. The syllabic dip D for an intervocalic SC is assigned the location of the minimum in $E[640,2800]$ in the segment SC. For prevocalic and postvocalic SC, D is assigned the middle frame of the SC segment.

E. Constrained landmark detection for word recognition

For isolated word or connected word recognition, manner class segmentation paths are constrained by a pronunciation model in the form of a finite state automata (FSA) (Jurafsky and Martin, 2000). Figure 6 shows an FSA-based pronunciation model for the digit "zero" and the canonical pronunciation /z I r ow/. The broad manner class representation corresponding to the canonical representation is Fr-V-SC-V-SC (the last SC is for the off glide of /ow/). The possibility that the off-glide of the final vowel /ow/ may or may not be recognized as a sonorant consonant is represented by a possible direct transition from the V state to the SILEN state. Starting with the start state S0, the posterior probability of a particular path through the FSA for zero can be calculated using the likelihood of a transition along a particular broad class B_i as

$$\prod_{k=1}^{N_{B_i}} \frac{P_t(f_k^i|x_t^{f_k^i}, f_1^i, \ldots, f_{k-1}^i)}{P_t(f_k^i|f_1^i, \ldots, f_{k-1}^i)}.$$

The likelihoods of all of the state transitions along a path are multiplied with the prior $P(B)$ and the duration densities $P(D|B)$ using the durations along that path. The probabilistic segmentation algorithm gives for an FSA and an observation sequence the best path and the posterior probability computed for that path. Note that word posterior probabilities can be found by multiplying the posterior probability $P(L|O)$ of the landmark sequence with the probability $P(U|OL)$ of the place and voicing features computed at the landmarks (Juneja, 2004), about computing complete word probabilities is out of the scope of this paper.

III. EXPERIMENTS AND RESULTS

A. Database

The "si" and "sx" sentences from the training section of the TIMIT database were used for training and development. For training the SVM classifiers, randomly selected speech frames were used because SVM training with all of the available frames was impractical. For training the HMMs, all of the si and sx sentences from the training set were used. All of the si and sx sentences from the testing section of the TIMIT database were used for testing how well the systems perform broad class recognition. The 2240 isolated digit utterances from the TIDIGITS training corpus were used to obtain word-level recognition results. If spoken canonically, the digits are uniquely specified by their broad class sequence. Thus, word-level results are possible for this constrained database. Note that the TIMIT database is still used for training since the TIDIGITS database is not transcribed. Thus, this experiment not only shows how well the systems perform word-level recognition, but it also allows for cross-database testing.

B. Sufficiency and invariance

In this section, an illustration of how the APs satisfy the assumptions of the probabilistic framework better than the MFCCs is presented. Although it is not clear how sufficiency and invariance can be rigorously established for certain parameters, some idea can be obtained from classification and scatter plot experiments. For example, sufficiency of the four APs used for the sonorant feature detection—$E[0,F3]$, $E[100\ \text{Hz},400\ \text{Hz}]$, $E[F3,f_s/2]$, ratio of the $E(0,F3)$ to the energy in ($F3$, half of sampling rate)[1]—can be viewed in relation to the 39 mel-frequency cepstral coefficients (MFCCs) in terms of classification accuracy of the sonorant feature. Two SVMs with linear kernels were trained, one for the APs and one for the MFCCs, using a set of 20,000 randomly selected sample frames of each of the *+sonorant* and *−sonorant* frames from dialect region 1 of the TIMIT training set. The same number of samples were extracted from dialect region 8 of the TIMIT training set for testing. A frame classification accuracy of 93.0% was obtained on data using the APs and SVMs, which compares well to 94.2% accuracy obtained using the MFCCs and SVMs. Note that for the two SVMs the same speech frames were used for training as well as testing and only the types of acoustic features were different.

In Eq. (16), the APs $x_t^{f_k^i}$ for a manner feature were assumed to be independent of the manner class labels of the preceding frames b^{t-1} when either $\{f_1^i,\ldots,f_k^i\}$ or $\{f_1^i,\ldots,f_{k-1}^i\}$ was given. For example, for the feature $f_k^i=+sonorant$, the set $\{f_1^i,\ldots,f_k^i\}$ is $\{+speech,+sonorant\}$ and $\{f_1^i,\ldots,f_{k-1}^i\}$ is $\{+speech\}$. Consider the case where $\{f_1^i,\ldots,f_k^i\}$ is given, that is, the value of the feature whose APs are being investigated is known. A typical case where the assumption may be hard to satisfy is when the APs for the *sonorant* feature are assumed to be invariant of whether the analysis frame lies in the middle of a vowel region or the middle of a nasal region (both vowels and nasals are *+sonorant*).That is, b^{t-1} will be composed of nasal frames in one case and vowel frames in the other case.

Such independence can roughly be measured by the similarity in the distribution of the vowels and nasals based on the APs for the feature *sonorant*. To test this independence, *sonorant* APs were extracted from dialect region 8 of the TIMIT training set from each of the nasal and vowel segments. Each set of APs was extracted from a single frame located at the center of the vowel or the nasal. The APs were then used to discriminate vowels and nasals using Fischer linear discriminant analysis (LDA). Figure 5(a) shows the distribution of the projection of the 39 MFCCs extracted from the same 200 frames into a one-dimensional space using LDA. A similar projection is shown for the four *sonorant* APs in Fig. 5(b). It can be seen from these figures that there is considerably more overlap in the distribution of the vowels and the nasals for the APs of the *sonorant* feature than for the MFCCs. Thus, the APs for the sonorant feature are more independent of the manner context than are the MFCCs. The overlap does not show worse performance of the APs compared to MFCCs because the *sonorant* APs are not meant to separate vowels and nasals. They separate vowels, nasals and semivowels (i.e., sonorants) from fricatives, stop consonants and affricates (i.e., obstruents). Thus, the APs for the feature *sonorant* are invariant across different *+sonorant* sounds but successfully discriminate *+sonorant* sounds from *−sonorant* sounds. Further discussion of the sufficiency and the invariance properties of the APs can be found in Juneja (2004).

C. Frame-based results

The SVMs for each feature utilized APs extracted from the analysis frame as well as s starting frames and e ending frames. The values of the two variables e and s were obtained for each classifier by performing validation over a subset of the TIMIT training data (Juneja, 2004). Training was performed on randomly picked samples (20,000 samples for each class) from the si and sx sentences of the TIMIT training set. The binary classification results on the whole of the TIMIT test set at the optimal values of s and e are shown in Table V in two cases—(1) when all the frames were used for testing and (2) when only the middle one-third portion of each broad class was used for testing. The difference in the results indicates the percentage of errors that are made due to boundary or coarticulation effects. Note that in the presented landmark-based system, it is not important to classify each frame correctly. The results on the middle one-third segment

TABLE V. Binary classification results for manner features in %. Accuracy on middle frames is not shown for the feature *continuant* because the feature distinguishes the stop releases from the beginning of fricatives.

Feature	*s*	*e*	Accuracy on middle frames	Accuracy on all frames
sonorant	4	1	96.59	94.39
syllabic	16	24	85.00	80.06
Speech/silence	3	2	94.60	93.50
continuant	4	4	…	95.58

are more representative of the performance of the system because if the frames in a stable region are correctly recognized for a particular manner feature, this would mean that the corresponding landmarks may still be correctly obtained. For example, if the middle frames of an intervocalic sonorant consonant are correctly recognized as *syllabic*, then the correct recognition of frames near the boundary is not significant because landmarks for the sonorant consonant will be obtained accurately. For the feature *continuant*, the classification error on middle frames is not relevant because the SVM is trained to extract the stop burst as opposed to a certain stable region of speech. Also the transient effects at broad class boundaries are minimized by low probability density values of very small broad class durations.

Figures 7–10 show the most significant sources of error for each of the phonetic features. The errors include misclassifications of the +*feature* sounds as −*feature*, and vice versa. For the feature *sonorant*, it can be seen that the sounds /v/ and the glottal stop /q/ are often detected as +*sonorant*. The sound /v/ is often manifested as a sonorant consonant so that the assignment of +*sonorant* for /v/ is expected. In the case of the glottal stop, a separate detector is required either at the broad class recognition level or further down the hierarchy to recognize glottalization because it can be significant for lexical access, especially in the detection of the consonant /t/ (Stevens, 2002). For the feature *syllabic*, classification accuracy for nasals as -syllabic is above 90%. But the semivowels—/ɹ/, /r/, /l/ and /w/ have lower accuracies which is expected because of the vowel-like behavior of these sounds. About 30% of the frames of reduced vowels are also misrecognized as sonorant consonants. This typically happened when a sonorant consonant followed a stressed vowel and preceded a reduced vowel such that the reduced vowel is confused as a continuation of the sonorant consonant. A similar result was shown by Howitt (2000) where the vowel landmarks were missed for reduced vowels more than other vowels. The performance of the feature *continuant* is 95.6% which indicates the accuracy on classification of onset frames of all nonsonorant sounds. That is, an error was counted if a stop burst was wrongly classified as −*continuant* or a fricative onset was wrongly classified as a stop burst. The major source of error is the misclassification of 13.7% of fricative onsets as stop bursts.

D. Sequence-based results

The SVM models obtained in the frame-based analysis procedure were used to obtain broad class segmentation as well as the corresponding landmark sequences for all of the si and sx sentences of the TIMIT test set using the probabilistic segmentation algorithm. Not all broad class sequences were allowed as the segmentation paths were constrained using a pronunciation graph such that (1) SCs only occur adjacent to vowels, (2) ST is always preceded by SILEN (for stop closure) and (3) each segmentation path starts and ends with silence. The same pronunciation graph was used for both the EBS system and the HMM system. The duration probability for each broad class was modeled by a mixture of Rayleighs using a single Rayleigh density for the classes SC,

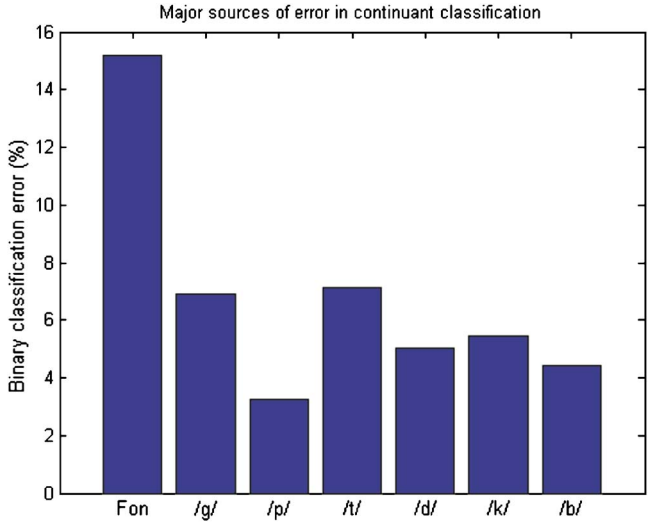


FIG. 7. (Color online) Sounds with high error percentages for the feature *sonorant*; "voic-stop" represents voiced stop consonants.

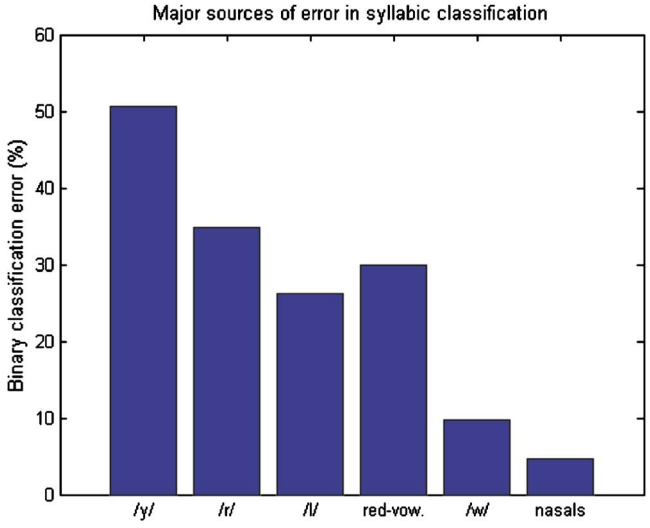


FIG. 8. (Color online) Sounds with high error percentages for the feature *continuant*. Fon represents fricative onsets.

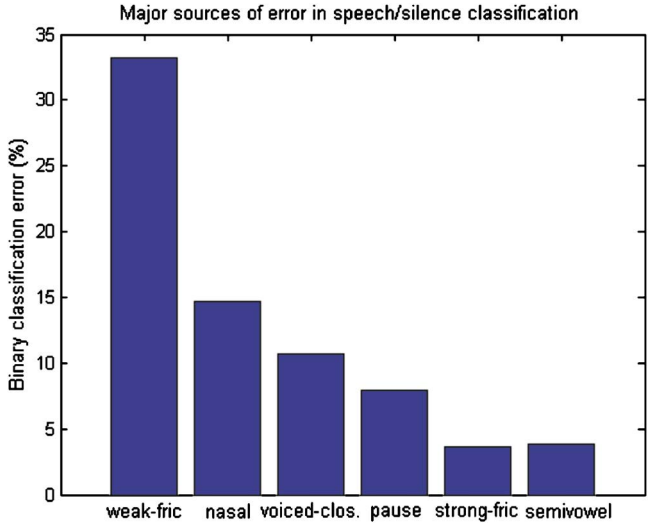


FIG. 9. (Color online) Sounds with high error percentages for the feature *syllabic*. "Red-vow" represents reduced vowels.

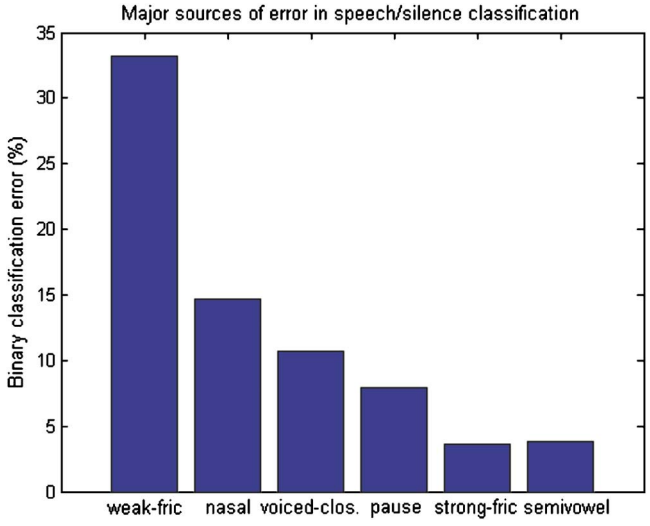


FIG. 10. (Color online) Sounds with high error percentages in speech silence distinction. "Weak-fric" represents weak fricatives and "strong-fric" represents strong fricatives. "Voiced-clos" represents closures of voiced stop consonants.

V, Fr and ST, and a mixture of two Rayleigh densities for SILEN (one density targets short silence regions like pauses and closures and the other density targets beginning and ending silence). The parameter for each Rayleigh density was found using the empirical means of the durations of each of the classes from the TIMIT training data.

For the purpose of scoring, the reference phoneme labels from the TIMIT database were mapped to manner class labels. Some substitutions, splits and merges as shown in Table VI were allowed in the scoring process. Specifically, note that two identical consecutive broad classes were allowed to be merged into one since the distinction between such sounds is left to the place classifiers. Also note that affricates were allowed to be recognized as ST+Fr as well as Fr, and similarly diphthongs—-/iʏ/, /eʏ/, /ow/, /aʏ/, /aw/, and /uw/—were allowed to be recognized as V+SC as well as V because the off glides may or may not be present. Scoring was done on the sequences of hypothesized symbols without using information of the start and end of the broad class segments which is similar to word level scoring in continuous speech recognition.

The same knowledge-based APs were used to construct an 11-parameter front end for a HMM-based broad class segmentation system. The comparison with the HMM-based system does not show that the presented system performs superior or inferior to the HMM-based systems, but it shows an acceptable level of performance. The HMM-based systems have been developed and refined over decades and the work presented in this paper is only the beginning of the development of a full speech recognition system based on phonetic features and acoustic landmarks.

All the HMMs were context-independent three-state (excluding entry and exit states) left-to-right HMMs with diagonal covariance matrices and eight-component mixture observation densities for each state. All the si and sx utterances from the TIMIT training set were used for training the HMM broad classifier. A HMM was built for each of the five broad classes, and a separate HMM was built for each of the special sounds—affricate, diphthong, glottal stop, syllabic sonorant consonant, flap /dx/ and voiced aspiration /hv/—making a total of 11 HMMs. Only the five broad class models were used in testing and the HMMs for the special sounds were ignored, so that the training and testing sounds of HMM and EBS were identical. The HMM models were first initialized and trained using all of the training segments for each model separately (for example, using semivowel and nasal segments for the sonorant consonant model), and then improved using embedded training on the concatenated HMMs for

TABLE VI. Allowed splits, merges and substitutions.

Reference	Allowed hypothesis	Reference	Allowed hypothesis
V+V	V	SC+SC	SC
Fr+Fr	Fr	SILEN+SILEN	SILEN
/q/+V, V+/q/	V	/q/	ST, SC
/t/, /p/, /k/, /g/, /d/	ST+Fr	/v/	SC, Fr
/em/, /en/, /er/, /el/	V+SC	/ch/, /jh/	ST+Fr
/hv/	SC, Fr	/dx/	SC
/dx/	SILEN+ST	/iʏ/, /ow/, /eʏ/, /oʏ/, /aw/, /uw/, /ow/	V+SC

TABLE VII. Broad class segmentation results in percent. Correctness (Corr)/Accuracy (Ace) are shown when the system is scored on the basis of numbers of deletions, insertions and substitutions of broad classes. A "-" in a cell means that the particular system was computationally too intensive to get a result from.

	EBS (RBF)	EBS (linear)	HMM
	Corr/Acc	Corr/Acc	Corr/Acc
11 APs	86.2/79.5	84.0/77.1	80.9/73.7
39 MFCCs	-	86.1/78.2	86.8/80.0

each sentence. Triphone models and other similarly improved HMM models may give better results than the ones presented in this paper, but the focus here is to build a base line HMM system to which EBS's performance can be compared.

The results are shown in Table VII. The results are also shown for EBS for two different front ends—AP and MFCC (including MFCCs, their delta and acceleration coefficients which gives a 39-parameter front end). The performance of all of the systems, except when EBS is used with MFCCs, is comparable although the HMM-MFCC system gives the maximum accuracy. The inferior performance of the MFCCs with EBS is perhaps because of the better agreement of APs with the invariance assumptions of the probabilistic framework. Similarly, better performance of MFCCs in the HMM framework may be because of better agreement with the diagonal covariance assumption of the HMM system applied here. That is, APs are not processed by a diagonalization step prior to application to the HMM systems while MFCCs go through such a process. These are possible explanations of these results and they are open to further investigation.

An example of landmarks generated by EBS on a test sentence of TIMIT is shown in Fig. 11 which also shows how errors in the system can be analyzed. The pattern recognizer calls the /dh/ in "the" (as marked by the ellipse) a sonorant constant (SC) instead of the correct broad class Fr. The reason is that the parameter $E[0,F3]/E[F3,f_s/2]$ does not dip adequately as it usually does in most *–sonorant sounds*. This indicates that improved APs, for example, from the APP detector (Deshmukh *et al.*, 2005) that directly captures the aperiodicity, are needed to correct errors like this one.

The confusion matrix of various landmarks for EBS using the AP front end is shown in Table VIII without including the sounds—diphthongs, syllabic sonorant consonants, flaps, /v/, affricates and the glottal stop /q/. For this latter set of sounds the confusion matrix is shown in Table IX. There is a considerable number of insertion errors. Insertions are common in any speech recognition system because typical speaking rates vary from training segments to test segments. There are sudden onsets of vowels and fricatives that give rise to stop burst insertions; 68% of stop burst insertions were at the begining of fricative segments and 46% were at the beginning of the sentences possibly because speakers are highly likely to start speaking with a sudden onset. High-frequency noise in the silence regions and aspiration at the end or beginning of vowels cause fricative insertions; 44% of all fricative insertions occur with an adjoining silence region, 43% of the rest of the fricative insertions have an adjoining vowel.

E. Word-level results and constrained segmentation results

The SVM and HMM models obtained by training on the TIMIT database were then applied to the isolated digits of the TIDIGITS database in both the vocabulary constrained

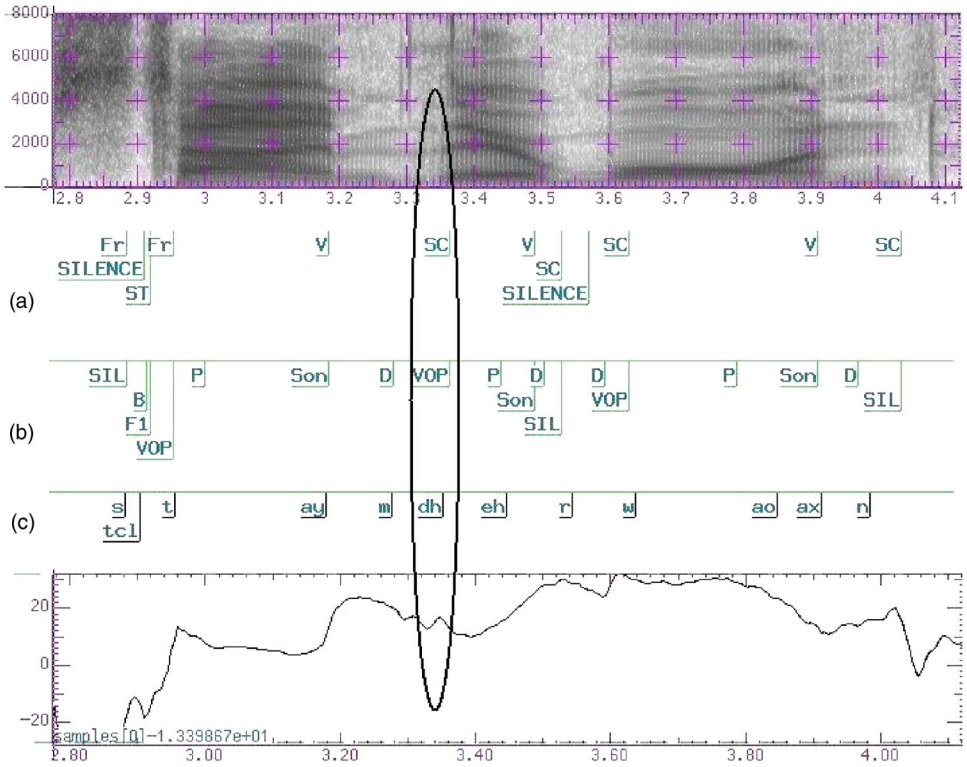


FIG. 11. (Color online) Top: spectrogram of the utterance, "time they're worn." A: Broad class labels, B: Landmark labels, C: phoneme labels, bottom: ratio of $E[0,F3]$ to $E[F3,f_s/2]$. Broad class and phoneme labels are marked at the end of each sound, and the landmark labels show the time instant of each landmark. The ellipse shows an error made by the system on this utterance. $E[0,F3]/E[F3,f_s/2]$ does not dip in the /dh/ region which makes the pattern recognizer call the fricative a *+sonorant* sound.

TABLE VIII. Confusion matrix for landmarks with exclusion of affricates, syllabic sonorant consonants, /v/, glottal stop /q/, diphthongs and flap /dx/. Only nonredundant landmarks are shown. For example, VOP implies presence of a syllabic peak P and vice versa, therefore, only VOP is used in the confusion matrix.

	Total	Fon	SIL	VOP	Son	B	Deletions	Correct (%)
Fon	6369	5607	10	1	136	185	430	88.03
SIL	10,232	15	9281	12	104	0	820	90.71
VOP	12,467	50	56	11,146	18	24	1173	89.40
Son	5504	155	65	1	4565	95	1290	70.82
B	9152	448	2	24	104	2755	797	84.98
Insertions	3439	682	692	206	1038	821		

and the unconstrained modes. In the unconstrained mode, the models were tested in exactly the same way as on the TIMIT database. To get the results on constrained segmentation, the segmentation paths were constrained using the broad class pronunciation models for the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The segmentation was identically constrained for both the HMM system and EBS. The results are shown in Table X for EBS (with linear as well as RBF kernels) and for the HMM systems trained on TIMIT and tested on TIDIGITS. On moving from unconstrained to constrained segmentation, a similar improvement in performance of the EBS (RBF) and HMM-AP systems can be seen in this table. This result shows that EBS can be constrained in a successful manner as for the HMM system. The overall performance of EBS using RBFs is also very close to the HMM-AP system. HMM-AP system shows better generalization than the HMM-MFCC system over cross-database testing which may be attributed to better speaker independence of the APs compared to the MFCCs (Deshmukh *et al.*, 2002).

Figure 12 shows an example of the output of the unconstrained probabilistic segmentation algorithm for the utterance "two" with canonical pronunciation /t uw/. The two most probable landmark sequences obtained from the algorithm are shown in this figure. The landmark sequence obtained with the second highest probability for this case is the correct sequence. It is hoped that once probabilistic place and voicing decisions are made, the second most probable sequence of landmarks will yield an overall higher posterior word probability for the word two.

Finally, word level accuracies were obtained for all of the systems. The state-of-the-art word recognition accuracy using word HMM models on TIDIGITS is above 98% (Hirrsh and Pearce, 2000). Recognition rates of 99.88% have also been obtained when using word HMM models for recognition of the TI-46 isolated digit database (Deshmukh *et al.*, 2002). These results were obtained using the same topology as in the present experiments (i.e., three-state HMMs with eight-mixture components). The difference is that instead of three-state HMM word models, we are now using three-state HMM broad class models to make the comparison with EBS. Note that a full word recognition system including place features is not presented here and only broad class models are presented. Therefore, a complete segmentation for a digit was scored as correct if it was an acceptable broad class segmentation for that digit.

The results are shown in Table XI. A fully correct segmentation of 68.7% was obtained using the EBS-AP system. About 84.0% of the digits had a correct segmentation among the top two choices. Note that the top two or three choices can be combined with place information to get final probabilities of words. A significant increase in correct recognition in the top two choices over the top one choice shows that there is a good scope of recovery of errors when place information is added. An accuracy of 67.6% was obtained by the HMM-AP system and an accuracy of 63.8% was obtained by the HMM-MFCC system. These results further confirm the comparable performance of the EBS and the HMM-AP systems. Specifically this result shows that a system that selectively uses knowledge based APs for phonetic feature detection can be constrained as well as the HMM systems for limited vocabulary tasks and can also give a similar performance in terms of recognition accuracy.

IV. DISCUSSION

A landmark-based ASR system has been described for generating multiple landmark sequences of a speech utterance along with a probability of each sequence. The land-

TABLE IX. Confusion matrix for affricates, syllabic sonorant consonants (SSCs), /v/, glottal stop /q/, diphthongs and flap /dx/. Empty cells indicate that those confusions were scored as correct but the exact number of those confusions were not available from the scoring program.

	Total	Fon	SIL	VOP	Son	B	Deletions	Correct (%)
/q/	927	2		0	5			99.25
Diph	4390	23	18	3991	25	5	328	90.91
SSCs	1239	11	14	1071	27	1	115	86.44
/v/	710		40	2		4	198	65.63
/dx/	632	40	6	0			75	80.85
/ch/, /jh/	570	562	0	1	0		7	98.60
/hv/	233		0	0		3	43	80.26

TABLE X. Broad class results on TIDIGITS (Correct/Accurate in percent).

	EBS (linear)	EBS(RBF)	HMM-MFCC	HMM-AP
Constrained	91.7/82.8	92.6/85.2	92.4/84.3	92.3/85.8
Unconstrained	89.5/64.0	93.0/74.3	88.6/74.1	84.2/72.9

mark sequences can be constrained using broad class pronunciation models. For unconstrained segmentation on TIMIT, an accuracy of 79.5% is obtained assuming certain allowable splits, merges and substitutions that may not affect the final lexical access. On cross database constrained detection of landmarks, a correct segmentation was obtained for about 68.7% of the words. This compares well with a correct segmentation for about 67.6% of the words for the HMM system using APs and 63.8% for the HMM system using MFCCs. The percentage accuracy of broad class recognition improved from 74.3% for unconstrained segmentation to 84.2% for constrained segmentation which is very similar to the improvement from 72.9 to 85.5% for the HMM system using APs. These results show that EBS can be constrained by a higher level pronunciation model similar to the HMM systems.

The comparison with previous work on phonetic feature detection is very difficult because of the different test conditions, definitions of features and levels of implementation used by different researchers. At the frame level, the 94.4% binary classification accuracy on the *sonorant* feature compares well with previous work by Bitar (1997) where an accuracy of 94.6% for sonorancy detection on the same database was obtained. The *continuant* result of 95.6% is not directly comparable with previously obtained stop detection results (Bitar, 1997; Liu, 1996; Niyogi, 1998). In the work by Niyogi (1998) results were presented at a frame rate of 1 ms, and in the work by Liu (1996) and Bitar (1997), results were not presented at the frame level. A full probabilistic landmark detection system was not developed in the research cited above. An 81.7% accuracy on the *syllabic* feature may seem low, but note that there is usually no sharp boundary between vowels and semivowels. Therefore, a very high accuracy at the frame level for this feature is not only very difficult to achieve, but also it is not very important as long as sonorant consonants are correctly detected. The authors have not been able to find a previous result to which this number can be suitably compared. At the sequence level, the overall accuracy of 79.5% is comparable to 77.8% accuracy obtained in a nonprobabilistic version of EBS (Bitar, 1997). Note that the most significant improvement over the system by Bitar (1997) is that the current system can be constrained for limited vocabulary and it can be used for obtaining multiple landmark sequences instead of one. There are various other systems to which segmentation results can be compared (Salomon *et al.*, 2004; Ali, 1999), but the comparison is omitted in this work because the purpose of this paper is to present how the ideas from such systems can be applied to a practical speech recognizer.

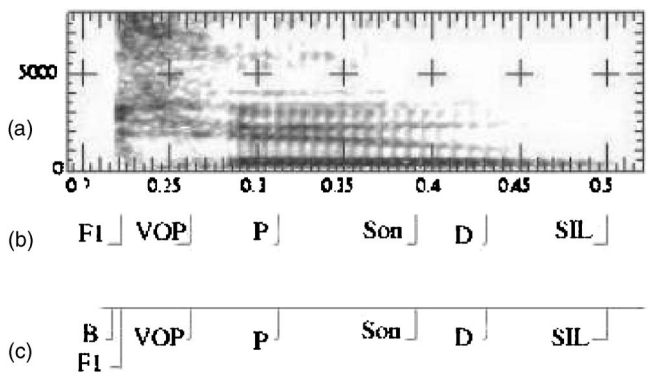


FIG. 12. A sample output of the probabilistic landmark detection for the digit "two." The spectrogram is shown in (a).Two most probable landmark sequences (b) and (c) are obtained by the probabilistic segmentation algorithm. The first most probable sequence (b) has a missed stop consonant but the second most probable sequence gets it.

The complete system for word recognition is currently being developed. There has been some success in small vocabulary isolated word recognition (Juneja, 2004) and in landmark detection for large vocabulary continuous speech recognition (Hasegawa-Johnson *et al.*, 2005). EBS benefits directly from research in discriminative APs for phonetic features, therefore, the system will improve as more powerful APs are designed for various phonetic features. By the use of APs specific to each phonetic feature EBS provides a platform for the evaluation of new knowledge gained on discrimination of different speech sounds. EBS provides easier evaluation of newly designed APs than HMM based systems. If certain APs give better performance in binary classification of phonetic features and are more context independent than currently used APs, then they will give overall better recognition rates. Therefore, complete speech recognition experiments are not required in the process of designing the APs. In the future, apart from the research that will be carried out on the automatic extraction of APs for all the phonetic features, further research will be done on better glide detection and incorporation of previous research (Howitt, 2000) on detection of vowel landmarks. APs to separate nasals from semivowels (Pruthi and Epsy-Wilson, 2003) and to detection nasalization in vowels (Pruthi, 2007) will be integrated along with an improved formant tracker Xia and Espy-Wilson (2000). Studies of pronunciation variability derived from previous work (Zhang, 1998) as well as continuing research will be integrated into EBS.

TABLE XI. Percent of TIDIGITS isolated digits with fully accurate broad class sequence.

EBS (RBF)	HMM-AP	HMM-MFCC
68.7	67.6	63.8

ACKNOWLEDGMENTS

This work was supported by Honda Initiation Grant No. 2003 and NSF Grant No. BCS-0236707. The authors would like to thank Om Deshmukh at the University of Maryland for help with the HMM experiments.

[1] $F3$ was computed as an average over the third formant values obtained in voiced regions using on the ESPS formant tracker (Entropic, 1997). The same average value of $F3$ was used in each speech frame for computation of manner APs.

Ali, A. M. A. (**1999**). "Auditory-based acoustic-phonetic signal processing for robust continuous speech recognition," Ph.D. thesis, University of Pennsylvania.

Bitar, N. (**1997**). "Acoustic analysis and modeling of speech based on phonetic features," Ph.D. thesis, Boston University.

Bitar, N., and Espy-Wilson, C. (**1996**). "A knowledge-based signal representation for speech recognition," *International Conference on Acoustics, Speech and Signal Processing*, Atlanta, GA, 29–32.

Burges, C. (**1998**). "A tutorial on support vector machines for pattern recognition," Data Min. Knowl. Discov., **2**, 2, 121–167.

Carbonell, N., Fohr, D., and Haton, J. P. (**1987**). "Aphodex, an acoustic-phonetic decoding expert system," Int. J. Pattern Recognit. Artif. Intell. **1**, 31–46.

Chang, S. (**2002**). "A syllable, articulatory-feature, stress-accent model of speech recognition," Ph.D. thesis, University of California, Berkeley.

Chomsky, N., and Halle, N. (**1968**). *The Sound Pattern of English* (Harper & Row, New York).

Chun, R. (**1996**). "A hierarchical feature representation for phonetic classification," Master's thesis, Massachusetts Institute of Technology.

Clarkson, P., and Moreno, P. J. (**1999**). "On the use of support vector machines for phonetic classification," *International Conference on Acoustics, Speech and Signal Processing*, Phoenix, AZ, 485–488.

Clements, G. N. (**1985**). "The geometry of phonological features," Phonology Yearbook **2**.

Delgutte, B., and Kiang, N. Y. S. (**1984**). "Speech coding in the auditory nerve: Iv. sounds with consonant-like dynamic characteristics," J. Acoust. Soc. Am. **75**, 897–907.

Deng, L., and Sun, D. X. (**1994**). "A statistical framework for automatic speech recognition using the atomic units constructed from overlapping articulatory features," J. Acoust. Soc. Am. **100**, 2500–2513.

Deshmukh, O., Espy-Wilson, C, and Juneja, A. (**2002**). "Acoustic-phonetic speech parameters for speaker independent speech recognition," *International Conference on Acoustics, Speech and Signal Processing*, Orlando, FL, 593–596.

Deshmukh, O., Espy-Wilson, C, and Salomon, A. (**2005**). "Use of temporal information: Detection of the periodicity and aperiodicity profile of speech," IEEE Trans. Speech Audio Process. **13**, 776–786.

Eide, E., Rohlicek, J., Gish, H., and Mitter, S. (**1993**). "A linguistic feature representation of the speech waveform," *International Conference on Acoustics, Speech and Signal Processing* **93**, Minneapolis, MN, 483–486.

Entropic (**1997**). "Entropic signal processing system 5.3.1," Company out of business.

Espy-Wilson, C. (**1987**). "An acoustic phonetic approach to speech recognition: Application to the semivowels," Ph.D. thesis, Massachusetts Institute of Technology.

Glass, J. (**1984**). "Nasal consonants and nasalized vowels: An acoustic study and recognition experiment," Master's thesis, Massachusetts Institute of Technology.

Glass, J., Chang, J., and McCandless, M. (**1996**). "A probabilistic framework for feature-based speech recognition," *International Conference on Spoken Language Processing*, Philadelphia, PA, 2277–2280.

Halberstadt, A. K. (**1998**). "Heterogeneous acoustic measurements and multiple classifiers for speech recognition," Ph.D. thesis, Massachusetts Institute of Technology.

Hasegawa-Johnson, M. (**1996**). "Formant and burst spectral measurements with quantitative error models for speech sound classification," Ph.D. thesis, Massachusetts Institute of Technology.

Hasegawa-Johnson, M., Baker, J., Borys, S., Chen, K., Coogan, E., Greenberg, S., Juneja, A., Kirchhoff, K., Livescu, K., Mohan, S., Muller, J., Sonmez, K., and Wang, T. (**2005**). "Landmark-based speech recognition: Report of the 2004 Johns Hopkins summer workshop," *IEEE International Conference on Acoustic, Speech and Signal Processing*, Philadelphia, PA, 213–216.

Hirrsh, H. G., and Pearce, D. (**2000**). "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," *Proc. ISCA ITRW ASR2000*, Paris, France, 181–188.

Hosom, J. P. (**2000**). "Automatic time alignment of phonemes using acoustic-phonetic information," Ph.D. thesis, Oregon Graduate Institute of Science and Technology.

Howitt, A. W. (**2000**). "Automatic syllable detection for vowel landmarks," Ph.D. thesis, Massachusetts Institute of Technology.

Joachims, T. (**1998**). "Making large-scale support vector machine learning practical," Advances in Kernel Methods: Support Vector Machines.

Juneja, A. (**2004**). "Speech recognition based on phonetic features and acoustic landmarks," Ph.D. thesis, University of Maryland, College Park.

Juneja, A., and Espy-Wilson, C. (**2002**). "Segmentation of continuous speech using acoustic-phonetic parameters and statistical learning," *International Conference on Neural Information Processing*, Singapore.

Juneja, A., and Espy-Wilson, C. (**2003**). "Speech segmentation using probabilistic phonetic feature hierarchy and support vector machines," *International Joint Conference on Neural Networks*, Portland, OR, 675–679.

Juneja, A., and Espy-Wilson, C. (**2004**). "Significance of invariant acoustic cues in a probabilistic framework for landmark-based speech recognition," *From Sound to sense:* 50+ *Years of Discoveries in Speech Communication* (MIT, Cambridge MA), pp. C–151 to C–156.

Jurafsky, D., and Martin, J. H. (**2000**). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (Prentice–Hall, Englewood Cliffs, NJ).

Keshet, J., Chazan, D., and Bobrovsky, B. (**2001**). "Plosive spotting with margin classifiers," Proceeding of Eurospeech, **3**, 1637–1640.

Kirchhoff, K. (**1999**). "Robust speech recognition using articulatory information," Ph.D. thesis, University of Bielefeld, Germany.

LDC (**1982**). "A speaker-independent connected-digit database," Http://www.ldc.upenn.edu/Catalog/docs/LDC93S10/, last viewed on January 30, 2007.

Lee, S. (**1998**). "Probabilistic segmentation for segment-based speech recognition," Master's thesis, Massachusetts Institute of Technology.

Liu, S. A. (**1996**). "Landmark detection for distinctive feature based speech recognition," J. Acoust. Soc. Am. **100**, 3417–3430.

NIST (**1990**). "Timit acoustic -phonetic continuous speech corpus," NTIS Order No. PB91 -5050651996.

Niyogi, P. (**1998**). "Distinctive feature detection using support vector machines," *International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, 425–428.

Pruthi, T. (**2007**). "Analysis, vocal-tract modeling and automatic detection of vowel nasalization," Ph.D. thesis, University of Maryland, College Park.

Pruthi, T., and Espy-Wilson, C. (**2003**). "Automatic classification of nasals and semivowels," *International Conference on Phonetic Sciences*, Barcelona, Spain.

Rabiner, L., and Juang, B. (**1993**). *Fundamentals of Speech Recognition* (Prentice–Hall, Englewood Cliffs, NJ).

Salomon, A. (**2000**). "Speech event detection using strictly temporal information," Master's thesis, Boston University.

Salomon, A., Espy-Wilson, C, and Deshmukh, O. (**2004**). "Detection of speech landmarks from temporal information," J. Acoust. Soc. Am. **115**, 1296–1305.

Shimodaira, H., Noma, K., Nakai, M., and Sagayama, S. (**2001**). "Support vector machine with dynamic time-alignment kernel for speech recognition," Proceeding of Eurospeech, **3**, 1841–1844.

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

Vapnik, V. (**1995**). *The Nature of Statistical Learning Theory* (Springer-Verlag, Berlin).

Xia, K., and Espy-Wilson, C. (**2000**). "A new formant tracking algorithm based on dynamic programming," *International Conference on Spoken Language Processing*, Beijing, China, **3**, 55–58.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2006**). *HTK Documentation* (Microsoft Corporation and Cambridge University Engineering Department), http://htk.eng.cam.ac.uk/, last viewed January 30, 2007.

Zhang, Y. (**1998**). "Towards implementation of a feature-based lexical-access system," Master's thesis, Massachusetts Institute of Technology.

Zue, V., Glass, J., Philips, M., and Seneff, S. (**1989**). "The MIT summit speech recognition system: A progress report," DARPA Speech and Natural Language Workshop, pp. 179–189.

Vibrational frequencies and tuning of the African mbira

L. E. McNeil[a)]
Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3255, USA

S. Mitran[b)]
Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3250, USA

(Received 4 June 2007; revised 22 October 2007; accepted 2 December 2007)

The acoustic spectrum of the mbira, a musical instrument from Africa that produces sound by the vibration of cantilevered metal rods, has been measured. It is found that the most prominent overtones present in the spectrum have frequencies that are approximately 5 and 14 times the lowest frequency. A finite-element model of the vibration of the key that takes into account the acoustic radiation efficiency of the various normal modes reveals that the far-field power spectrum is dominated by modes involving predominately transverse motion of the key. Modes involving longitudinal motion do not radiate efficiently, and therefore contribute little to the sound produced. The high frequencies of the dominant overtones relative to the fundamental make it unlikely that the tunings of the mbira that are used by expert musicians are determined by matching the fundamental frequencies of the upper keys with the overtones of the lower keys. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2828063]

PACS number(s): 43.75.Kk, 43.75.Bc [NHF] Pages: 1169–1178

I. INTRODUCTION

The *mbira* is a musical instrument found in multiple forms in many countries across the continent of Africa, especially in Zimbabwe, the Democratic Republic of the Congo, Mozambique, and Angola.[1] One form, known as the *mbira dzavadzimu* ("mbira of the ancestors") is particularly characteristic of the melodic music of the Shona people of Zimbabwe, Mozambique, and Zambia. The keys are put into vibration by stroking the free end with the thumb or forefinger, and each one produces a distinct musical pitch. The sound is sometimes amplified by placing the mbira inside a hollow shell such as a gourd that serves as a resonating chamber. The music played on the mbira is typically polyphonic, so that multiple keys are sounded simultaneously or alternately in sequence.

Like all melodic instruments with multiple vibrating parts, the pitches produced by the individual keys are related to one another in an organized way that is referred to as the *tuning* of the instrument. Throughout Africa many mbira tunings are used, which are related to melody and chord patterns prevalent within the local culture and also reflect the artistry of individual master players. It is reasonable to ask, however, whether those tunings are related to the musical acoustics of the instrument itself.

Because more than one mode of vibration is excited when the key is stroked, the sound produced by the vibration of the key will contain multiple frequencies. The lowest of these frequencies, which determines the pitch experienced by the listener, is called the *fundamental*. The higher frequencies, which together determine the overall characteristics of the musical sound, are referred to as *overtones*. We set out to examine the frequencies of vibration of the keys of the mbira and to explore their possible relationship to the tuning of the instrument.

II. EXPERIMENT AND CALCULATION

A. The mbira

The instrument used for these measurements is shown in Fig. 1. It was made by Josephat Mandaza in Chitungwiza, Zimbabwe in 2002 and was provided to us by Dr. Louise Meintjes of the Duke University Department of Music. The body of the instrument is made of a single piece of solid wood approximately 18.5×21.7 cm and 2.5 cm thick. A 2.5-cm-diam hole in the body near one end provides a convenient finger-hold. The 22 vibrating rods or keys range in length from 9.5 to 17.5 cm. They are clamped at one end and cantilevered from a bridge made of a metal strip embedded in the wood body, such that the freely vibrating length ranges from 3.5 to 10.5 cm. They are arranged in three groups: nine on the right-hand side of the instrument (designated R1–R9), six upper keys on the left-hand side (designated L1–L6), and seven lower keys on the left-hand side (designated B1–B7). The flattened tongue-like shapes of the freely vibrating portion of the keys are approximately 0.5–0.8 cm wide at the bridge and 0.5–1.5 cm wide at the widest point (near the free end). The cross section of each key is pentagonal, being approximately 0.09 cm thick at the edges and 0.15 cm thick in the center. The keys curve upward from the bridge (presumably to make them easier to play) such that the free end is raised above the bridge by about 10% of the key's length. The material of which the keys are made is unknown, but

a)Electronic mail: mcneil@physics.unc.edu.
b)Electronic mail: mitran@amath.unc.edu.

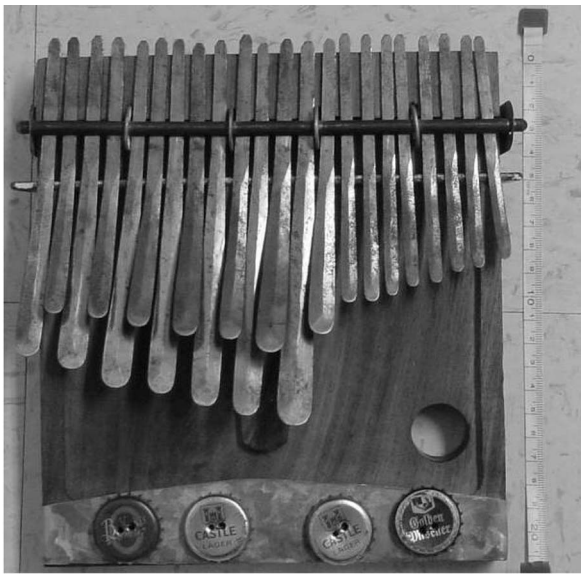


FIG. 1. Photo of the mbira used in the measurements.

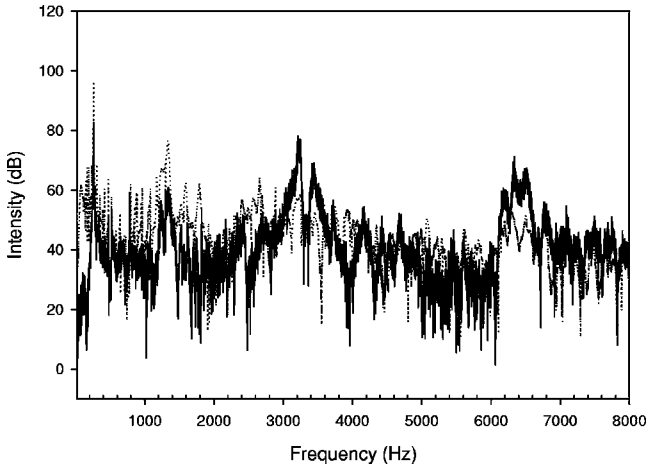


FIG. 2. Power spectrum of audio signal from key B7 (fundamental frequency 260 Hz), background subtracted (solid line); and the same spectrum recorded with all of the keys free to vibrate (dotted line).

from their appearance and stiffness and the presence of a small amount of rust on them they can be presumed to be some sort of mild steel.

At the opposite end of the instrument body from the bridge, a metal strip is nailed across the width of the wood. Four metal beer-bottle caps are attached to the strip with wire such that they are free to rattle against the metal when the instrument is played. Traditionally, shells were used for this function,[2] but in more recent times bottle caps have become a convenient substitute. The resulting buzzing sound is considered to be an important characteristic of the instrument in its indigenous use in music-making, but it does not affect the pitches produced by the keys and for the purpose of the experiments described here it was undesirable. For that reason, before making the measurements we damped the vibration of the bottle caps by stuffing them with wool fleece. The vibrations of keys other than the one being measured were similarly damped.

B. Acoustic measurements

We made the acoustic measurements in a sound isolation chamber in the UNC-CH Department of Linguistics. We made sound recordings with a Shure KSM44 microphone using the omnidirectional pickup pattern and 80-Hz high-pass filter. The microphone preamplifier was a Mackie 1202-VLZ PRO mixer. We used the channel's insert-send as the signal output to minimize contributed noise and distortion. We used an Echo Digital Audio Indigo IO for the analog-to-digital signal conversion. The conversion parameters were 24-bit quantization and 22 050-Hz sample rate, and we stored the samples on a hard drive using Adobe AUDITION 1.5 software. We used the same software to perform the Fourier analysis of the audio samples. An initial baseline measurement to evaluate the environmental noise recorded that all frequencies from 100 Hz to the Nyquist frequency were at least 102 dB below full scale. We held the mbira close to the microphone with one hand and one key was set into vibration by pressing it with a finger on the other hand and allowing the finger to slide off the end of the key. Recording continued until no further vibration of the body of the mbira could be perceived with the hand holding the instrument, approximately 10 s. (This was a longer time than the sound was audible.)

An example of the power spectrum resulting from the Fourier transform of the audio signal is shown in Fig. 2. The spectrum contains a strong, relatively sharp peak 50–80 dB above background at the fundamental frequency of the key's vibration. This mode of vibration is responsible for the pitch the key sounds, and ranges from 117 to 948 Hz (approximately $B_2^\flat$ to $B_5^\flat$ in the American system of pitch notation). A second peak, typically 30–50 dB above background, appears at the frequency of the next higher mode of vibration. In most of the spectra a third (and sometimes a fourth) peak is also distinguishable from the background level. The frequencies recorded for each key are given in Table I. Also shown in Fig. 2 (dotted line) is the spectrum recorded from that key with all of the keys undamped, as they would be when the instrument is played. As can be seen, the damping has little if any effect on the spectrum except at the higher frequencies, where the sound is less intense.

C. Frequency calculations

We have computed the vibration frequencies of some of the mbira keys using two finite element methods. In the first method, a mbira key is treated as a one-dimensional body and discretized using curved beam (more concisely, arch) elements as described in the following. Arch elements were introduced after an initial study showed poor behavior of a model constructed using standard, straight beam elements. In the second method, a key is modeled as a three-dimensional body using standard brick elements (the C3D6 element of the CALCULIX finite element program was used). Two methods were used to provide a check against one another and to resolve a number of interesting questions that arise in comparisons of experimental and computational results.

The behavior of straight elastic beams is well understood and accurately described by the classical Euler–Bernoulli and Timoshenko beam theories. Curved beams ex-

TABLE I. Frequencies recorded for the keys of the mbira.

Key	Fundamental (Hz)	Overtone 1 (Hz)	Overtone 2 (Hz)
B1	117	646	1705
B2	141	722	1883
B3	152	784	2033
B4	171	890	2290
B5	191	965	2530
B6	212	1080	2882
B7	260	1341	3220
L1	234	1176	3005
L2	353	1613	3990
L3	310	1502	2874
L4	392	2310	
L5	420	2000	5060
L6	475	2167	5597
R1	288	1515	3940
R2	474	2365	6116
R3	525	2480	6137
R4	587	2764	
R5	631	3022	
R6	704	2944	5200
R7	788	1576	3941
R8	852		
R9	948		

hibit more complicated behavior because of the nonlinear coupling among bending, twist, shear, and extension. Due to their importance in structural engineering, there have been many studies on the behavior of curved beams (e.g., Ref. 3). An important question is whether the scale disparity between beam length and thickness can be used to separate the three-dimensional elasticity problem into a one-dimensional problem along the beam length and a two-dimensional problem in the beam cross section. The two-dimensional problem can typically be easily approximated, and the overall complexity is reduced to solving a one-dimensional problem instead of the full three-dimensional elasticity problem. That such a splitting is possible has been established by Berdichevsky.[4] The problem of determining a specific splitting procedure is still a research subject. Specific characteristics of the problem, including amount of deformation, curvature, nature of applied forces, come into play, and some splittings are able to capture the above-mentioned effects while others fail (see Ref. 5). Even though one-dimensional curved beam theories exhibit these deficiencies, the overall economy in computational effort makes them attractive.

There are a number of formulations for curved finite elements. The basic steps in the construction of the finite element model are: (1) define an approximation function for the displacements as a function of longitudinal position; (2) construct the elemental stiffness and mass matrices using a variational formulation; (3) assemble the elemental matrices to form a global eigenproblem; and (4) solve the eigenproblem for a specific set of boundary conditions. Steps (2)–(4) are standard procedures which can be found in a number of finite element texts, e.g., Bathe.[6] The main feature which distinguishes one arch finite element formulation from another is how displacements are approximated along the beam length. A specific formulation has been introduced to study this problem.

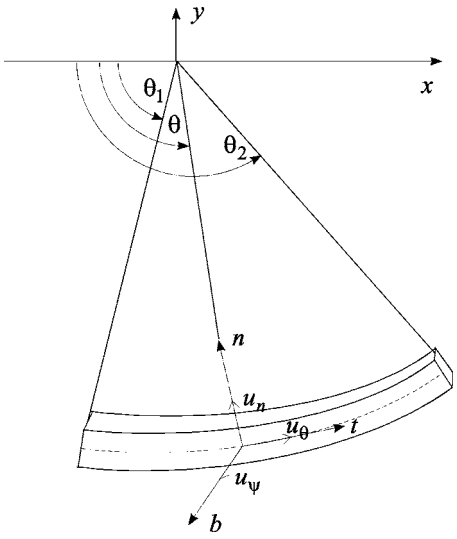


FIG. 3. Diagram showing finite element coordinate system and deformations.

We use a local Frenet triad (t,n,b) to describe the deformations in the finite element cross section, with the coordinates denoting the tangential, normal, and binormal direction, respectively. (See Fig. 3 for the definition of the coordinates.) We assume that the finite element is small enough to consider the radius of curvature R constant over the element's length; hence the longitudinal position can also be specified by the angle $\theta=t/R$. We assume that plane cross sections remain plane during deformation. We consider the beam element to undergo three types of deformation: (a) a tangential deformation u_θ; (b) a normal deformation u_n; (c) a cross-section rotation u_ψ. All these deformations are defined at the element mean fiber, i.e., the line along the thickness of the element that does not change in length during deformation. In general, there exists some series expansion of the displacement u in terms of the θ angle, $u=u(\theta)$. We desire for this expansion to have as few terms as possible while still providing sufficient accuracy. The distinguishing feature of the finite element approximation used here is the use of a mixed, Taylor-trigonometric expansion to approximate the displacements. The tangential deformation is approximated as

$$u_\theta = a_1 + a_2 \cos\theta + a_3 \sin\theta + a_4\theta\cos\theta + a_5\theta\sin\theta. \quad (1)$$

If we neglect shear deformations (relative slipping of adjacent finite element cross sections) the expansions for (u_n,u_ψ) can be written using the coefficients from the expansion for u_θ,

$$u_n = Ca_1\theta + a_2\sin\theta - a_3\cos\theta + a_4(\sin\theta - \theta\cos\theta) + a_5(\cos\theta + \theta\sin\theta) + a_6,$$

$$u_\psi = \frac{C}{R}a_1\theta + \frac{2}{R}a_4 \sin\theta + \frac{2}{R}a_5 \cos\theta + \frac{1}{R}a_6, \tag{2}$$

where $C=1+I_b/AR^2$, A is the cross-sectional area, and I_b is the moment of inertia around the binormal axis. Note that twisting of the cross section is not taken into account in this model. Normal stresses are assumed to vary linearly across the cross section. The resultant forces from a cross section can be related to the displacements through

$$F_t = \frac{dF_\theta}{d\theta},$$

$$F_\theta = \frac{EA}{R}\left(\frac{du_\theta}{d\theta} - u_t\right) - \frac{M_b}{R},$$

$$M_b = \frac{EI_b}{R^2}\left(u_t + \frac{d^2u_t}{d\theta^2}\right). \tag{3}$$

Once the longitudinal displacement functions and the forces in the cross section have been established, standard finite element procedures lead to the formation of an eigenproblem

$$\omega^2\mathbf{Mu} = \mathbf{Ku}. \tag{4}$$

In Eq. (4), ω is the pulsation, $\mathbf{M}$ the mass matrix, $\mathbf{K}$ the stiffness matrix, and $\mathbf{u}$ the vector of displacements. In order to completely define the problem, boundary conditions must be imposed. Given the construction of the mbira, the vertical displacements at the point where the keys are clamped are set as null. The displacements in the other two directions are impeded by frictional contact between the key and the soundboard or upper clamping bar. In the computations carried out here, these displacements were also set as null. The behavior at the fret is more difficult to capture simply. Displacements along the vertical orientation (perpendicular to the soundboard) are impeded only in one direction (downwards). Lateral and longitudinal displacements are impeded due to friction between the key and the fret but not necessarily nullified. We consider the effect of a number of boundary conditions at the fret below. Solving the eigenproblem gives the vibration frequencies $\nu_i=\omega/2\pi$ and associated vibration modes $\mathbf{r}_i$.

D. Acoustic radiation from a vibrating mbira key

The sound produced by a struck mbira key is a superposition of the vibration modes from the above-presented finite element computation. Not all modes are excited by the initial striking of the key and some modes are more efficient sources of acoustic radiation than others. Furthermore, the soundboard is also excited by a vibrating key and radiates acoustic waves. All these effects have to be taken into account in order to compare a measured acoustic power spectrum with the results from a computational simulation.

1. Key playing

We adopt a simple model of how a key is played. We assume that a key is deformed from its equilibrium position by a force acting normal to the key mean fiber and then instantaneously released. The initial shape $\mathbf{u}_0$ is determined by solving the problem

$$\mathbf{Ku}_0 = \mathbf{f}, \tag{5}$$

where $\mathbf{f}$ is a vector of forces applied at each node within the finite element model. For the model of key playing considered here all components are null except those corresponding to the end node of a key.

2. Vibration mode amplitudes

The initial shape $\mathbf{u}_0$ is decomposed on the vibration modes $\mathbf{R}=[\mathbf{r}_1,\mathbf{r}_2\cdots\mathbf{r}_N]$ by solving the linear system

$$\sum_{j=1}^{N}(\mathbf{r}_i\cdot\mathbf{r}_j)\mathbf{a}_j = \mathbf{r}_i\cdot\mathbf{u}_0, \quad i=1,2,\ldots,N. \tag{6}$$

The amplitudes $\mathbf{a}_j$ specify the degree of initial excitation of each vibration mode.

3. Acoustic efficiency

Mbira keys are curved arches with significant coupling among extensional, twisting, and bending vibration modes. Only weak acoustic radiation is expected from an extensional or twisting mode, due to the small key surface area that sets the surrounding air into motion. The keys radiate directly and also set the soundboard into motion. The surface area of the soundboard is much larger than that of the keys and, conceivably, extensional modes could radiate efficiently by exciting the soundboard. However, the coupling of a key to the soundboard is accomplished by ensuring *vertical* contact at three points: the edge of the soundboard, the clamp pressing on the top of a key, and the fret pressing against the bottom of a key. The contact in the other directions is through friction forces that result from the vertical clamping forces. Longitudinal vibrations of a key produce forces which act against the frictional contact and are not expected to produce significant excitation of the soundboard. We therefore concentrate on the relative acoustic efficiency of sound radiation from the key motion itself, understanding that the intensity of acoustic waves emitted by bending modes of a key is amplified by the soundboard while those associated with extensional or twisting modes are not.

The acoustic field produced by a key is the solution of

$$\phi_{tt} - c^2\nabla^2\phi = 0, \tag{7}$$

along with boundary conditions given by the key motion. In Eq. (7) ϕ is the acoustic velocity potential and c is the sound speed. Rather than solve this complicated problem in a domain with a moving boundary, we carry out a simple estimate. We assume that only the normal displacement associated with a vibration mode makes a significant contribution to the overall acoustic field. Assuming the vibration amplitudes to be small, the above-presented problem is replaced by $\phi_{tt}-c^2\nabla^2\phi=Ayy_t$ with the source term yy_t defined only on the mbira key, and A a scaling factor which plays no role here and shall be taken as unity. Here y is the displacement of the key normal to the mean fiber, and the source term arises from integrating the kinematic condition $\nu=\phi_y=y_t$

which states that the acoustic velocity equals the key velocity for positions on the key (for an extended discussion of how this equation is obtained see Refs. 7 and 8). We further assume the key to be infinitely thin. The solution to Eq. (7) can be expressed as the convolution

$$\phi(\vec{x},t) = \int_0^L \int_0^\infty G(\vec{x} - \xi\vec{e}_\xi, t-\tau) y(\xi) y_t(\xi) d\tau d\xi \tag{8}$$

with $\vec{\mathbf{e}}_\xi$ being the unit vector tangent to the key mean fiber and $G(\vec{x},t) = \delta(|\vec{x}| - ct)/(4\pi|\vec{x}|)$ the fundamental solution for the wave equation. The acoustic intensity is given by $I = pu$ with p the acoustic overpressure and u the acoustic velocity. In terms of the velocity potential we have: $I \sim \phi_t \partial\phi/\partial y$. Use of the Dirac delta allows the time integral to be evaluated and we are led to the far-field estimate

$$I(\omega) \sim \omega \int_0^L y^2(\xi,\omega) d\xi \sim \omega a^2(\omega) \mathbf{r}^n(\omega) \cdot \mathbf{r}^n(\omega), \tag{9}$$

stating that the far acoustic field is proportional to the eigenmode frequency and the squared normal-to-mean-fiber amplitude.

The estimate (9) allows us to predict the acoustic power spectrum obtained by playing a key. When a key is played a beam eigenmode is excited with amplitude $a(\omega)$ obtained by solving Eq. (6). For each eigenmode we compute the integral (9) and thereby obtain a prediction of the acoustic wave intensity produced by that eigenmode. Note that modes for which the transverse motion $y^2(\xi,\omega)$ is small will contribute less to the overall acoustic spectrum even if their initial excitation by stroking the key $a(\omega)$ is comparable to that of other modes.

As mentioned, for a curved beam bending and extensional modes are coupled. In order to classify which modes behave more like bending modes (i.e., transverse motion is predominant) and which behave more like extensional modes (i.e., longitudinal motion is predominant) we introduce a criterion. Let z denote the deformation in the local direction of the mean fiber, and introduce the quantity

$$J(\omega) \sim \omega \int_0^L z^2(\xi,\omega) d\xi. \tag{10}$$

We define modes for which $\eta(\omega) = I(\omega)/J(\omega) > 1$ as "transverse" while those for which $\eta(\omega) \leqslant 1$ shall be referred to as "longitudinal."

III. RESULTS

A. Measured frequencies

Figure 4 shows the ratio of the frequencies of the second and third peaks in the spectrum (hereafter referred to as the first and second overtones) recorded for each key to the fundamental frequency produced by the key, plotted versus the fundamental frequency. With the exception of the shortest key, the ratio of the frequency of the first overtone to that of the fundamental is approximately 5, ranging from 4.2 to 5.8. The second overtone is more variable, but for the longest keys (those with the lowest fundamental frequencies), its frequency is approximately 13.6 times the fundamental frequency. This suggests that the modes of vibration of the keys are all very similar.

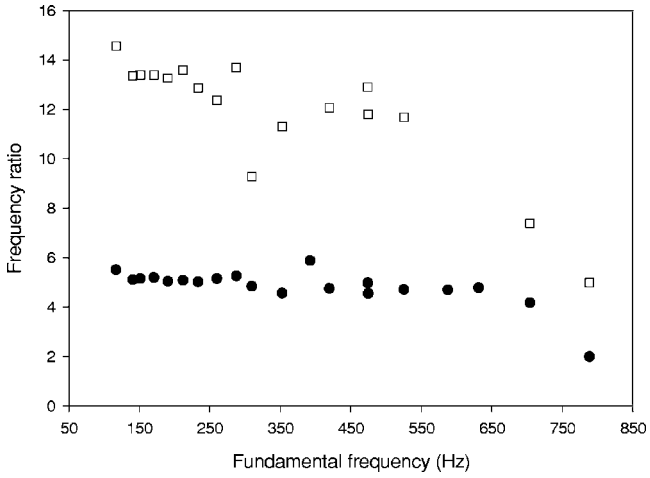


FIG. 4. Ratio of overtone frequencies to fundamental frequency. Closed symbols: First overtone. Open symbols: Second overtone.

B. Calculated frequencies

The exact material properties of the key were not known and could not be measured without damage to the instrument; hence we used the fundamental frequency to calibrate the Young's modulus (E) and density (ρ) of the material of which the key was made. Use of standard values for soft steel ($E = 200$ GPa, $\rho = 7800$ kg/m^3) led to a relative error of $\sim 10\%$ in prediction of the fundamental frequency among the keys for which the computation was carried out (B1, B7, R9). Given this calibration against the fundamental frequency, the main prediction of the model is the frequency of the overtones as well as the overall shape of the far-field acoustic spectrum. We used a smoothing spline procedure to adjust the geometric data for the keys to avoid propagating small measurement errors into the numerical procedure for computing the eigenmodes and frequencies.

The eigenmodes for key B7 computed using the three-dimensional model are shown in Fig. 5. The three-dimensional computations served as a check on the one-dimensional model and also to investigate fret boundary conditions. The computations are however quite expensive. In order to obtain convergence in the first six eigenmodes to within 1 Hz approximately 240,000 brick elements were required. Carrying out the decomposition of the initial deflection onto an eigenmode basis was not feasible. The effect of fret boundary conditions is shown in Table II. When vertical displacements are set to null the key oscillates as a shorter length beam with a consequent increase in frequency (compare mode 1, free displacement and $u_y = 0$ case). Blocking further degrees of freedom changes the overtone frequencies.

Frequencies obtained from the one-dimensional arch element computation of the B7 key are shown in Fig. 6. Given the much simpler computational effort many more eigenmodes were computed, sufficient to accurately represent the power spectrum associated with the initial deformation imposed by stroking a key. Some of the modes are spurious

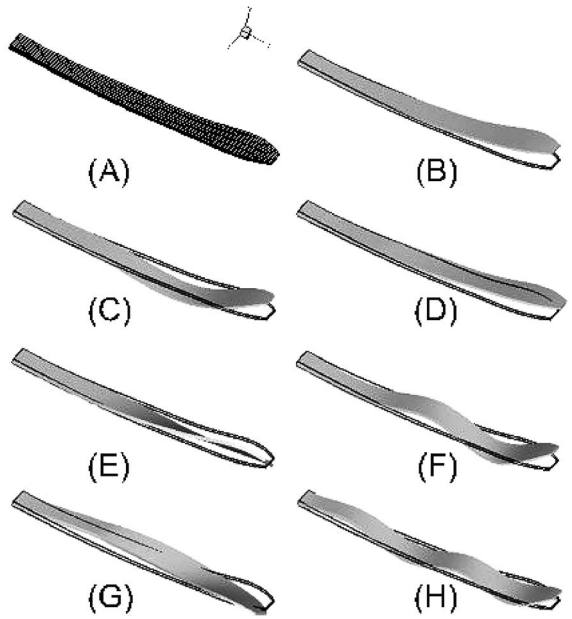


FIG. 5. (A) Three-dimensional finite element discretization. (B)–(E) Eigenmodes corresponding to seven lowest frequencies. At the fret the vertical displacements were set to null, the others were left free. Modes B,C,F,H are predominantly bending modes. Modes E,G are predominantly twisting modes. Mode D is similar to mode B but with pronounced extensional component.

however and result from the simplifying assumptions of the one-dimensional model. In particular modes 1, 2 from the one-dimensional computation arise due to the rapid tapering of a key toward the end. This leads to large cross-section rotations in the finite element model. The deformation for these spurious modes corresponds to bending of the tapered key tip attached to the much stiffer main part of the key. The possibility that spurious modes appear in regions where the simplifying assumptions of one-dimensional beam theory break down is unavoidable. In particular, results should be scrutinized in region of rapid variations in geometry. Spurious modes can also arise from shear locking behavior (details on origin and elimination of spurious modes can be found in Ref. 9). Fortunately, the spurious modes have relatively small transverse displacements and contribute little to the acoustic power spectrum evaluated by Eq. (9) even though their vibration power spectrum is significant (pressing on the end of a key imposes a large initial amplitude on these modes). Hence in this work no additional modification of the finite element method was made to eliminate spurious modes.

TABLE II. Effect of boundary conditions at fret upon frequencies calculated using the three-dimensional finite element model. Displacements u_x, u_y, u_z correspond to directions along key length, perpendicular to key and vertical, perpendicular to key and horizontal.

Fret boundary condition	Mode A	Mode B	Mode C	Mode D	Mode E
Free displacements	171	728	830	2256	2348
$u_y=0$	256	749	1441	2771	3970
$u_x=0$, $u_y=0$	261	1465	1607	2797	4010
$u_x=0$, $u_y=0$, $u_z=0$	261	1465	1848	2848	4010

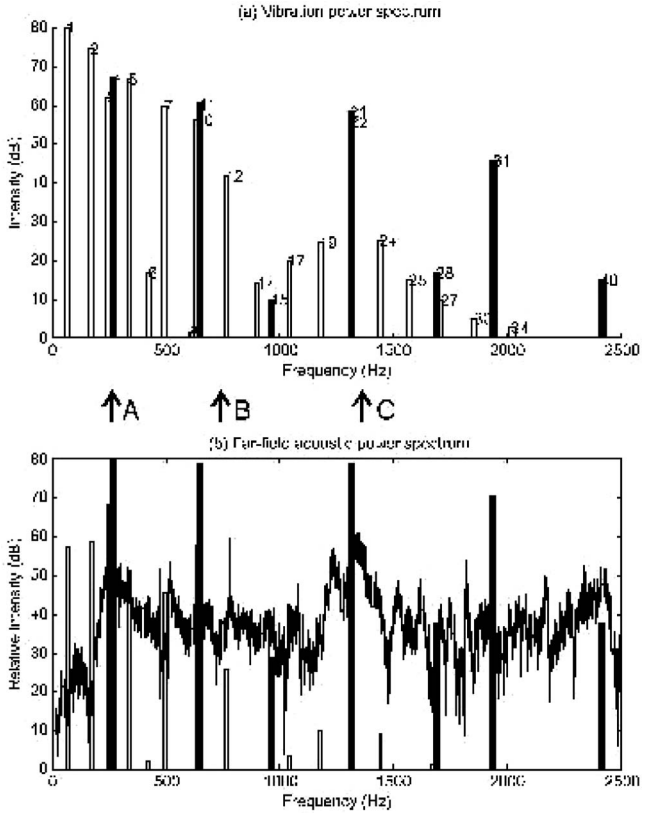


FIG. 6. Results of one-dimensional finite element analysis for the B7 key. Initial deformation produced by a force acting normal to the mean fiber at the end of the key. (a) Vibration power spectrum showing the principal transverse modes (solid bars) 4, 11, 22, 32 as well as a wide variety of longitudinal modes (open bars). (b) Far-field acoustic power spectrum predicted by the finite element analysis (bars) compared with the experimentally measured spectrum (line). Arrows A, B, and C indicate frequencies obtained from the three-dimensional model.

IV. DISCUSSION

A. Overtone frequencies

The vibrational frequencies of a simple cantilever can be calculated by standard techniques.[10] The well-known result for a straight beam of uniform cross section that is clamped at one end is that the frequencies of the first and second overtones are 6.4 and 17.5 times the fundamental frequency. However, the mbira keys are neither uniform in cross section nor straight. There is significant coupling among bending, twisting and compressional deformation.

A typical result from the computation of the B7 key is presented in Fig. 6. The vibration power spectrum shows the initial amplitude of each eigenmode imposed by stroking the key as obtained from solving system (6). The far-field acoustic power spectrum shows the results of evaluating integral $I(\omega)$ from Eq. (9) to obtain the radiated sound intensity. Bars show the predictions of the one-dimensional model, arrows show the frequencies obtained from the three-dimensional model. The strongly excited, one-dimensional, modes 1 and 2 are spurious modes, arising from breakdown of one-dimensional simplifying hypotheses as mentioned earlier. The most efficient acoustic radiator [highest $I(\omega)$] among the low frequency modes is mode 4, which corresponds to bending of the entire key. We calibrated the observed fundamental

frequency to mode 4 to account for the unknown material properties, thus obtaining a value of $E=205$ GPa for the elasticity modulus. We then observed that the next most efficient acoustic radiator is mode 21 at a frequency of 4.9 times that of the fundamental; the associated computed frequency $f=1313$ Hz is within 2% of the measured first overtone of 1341 Hz (Table I). Other vibration modes which are significant sound radiators are modes 11, 31, and 40. All of these show up as peaks within the measured acoustical spectrum [Fig. 6(c)], thereby confirming the overall accuracy of the numerical procedure.

From a modeling point of view it is interesting to compare the three-dimensional and one-dimensional approaches. The three-dimensional approach allows for accurate predictions of the low order modes, but at such a computational cost that obtaining a complete eigenbasis to compute the vibration power spectrum and the far-field acoustic power spectrum becomes prohibitive. Furthermore, contact boundary conditions at the fret introduce an additional unknown which must be explored through (expensive) numerical experimentation. The one-dimensional model sometimes predicts spurious modes, but furnishes a sufficiently accurate description of key vibrations for a correct evaluation of the far-field acoustic spectrum.

A simulation of the effect of the soundboard was also carried out. Forces from transverse modes and from longitudinal modes were applied on a rectangular slab of anisotropic, three-dimensional brick elements in the CALCULIX program. The transverse forces were applied through a dashpot element to model the frictional contact. As expected the soundboard received negligible energy from longitudinal modes by comparison to transverse modes. This is a direct consequence of the coupling between the vibrating elements and the soundboard. Hence in practice the relative attenuation of the longitudinal modes with respect to the transverse modes is even more pronounced than that computed here.

B. Tuning

The mbira is in essence a keyboard instrument (and it is sometimes called a "thumb piano" by Westerners), and therefore must have a well-defined relationship among the pitches (i.e., fundamental frequencies) of the different keys. The pitch of an individual key can be adjusted by shifting it relative to the bridge so that the vibrating length of the key is increased (to lower the pitch) or decreased (to raise the pitch). Since in its indigenous use the tuning of each key is done "by ear" (or by comparison to another such instrument) rather than by reference to an absolute standard of pitch, it is reasonable to expect that the fundamental frequencies to which the higher-pitched keys are tuned will bear some relationship to the overtone frequencies of the lower-pitched keys, so that when played together the overtones of the different keys will coincide to produce a harmonious sound.

Simple stretched strings have overtone frequencies that are in pure whole-number ratios to the fundamental frequency, and so tuning successive strings such that their fundamental frequencies are 2, 3, 4, 5, etc., times the fundamental frequency of the lowest string allows the frequencies of the overtones of the strings to overlap to produce a harmonious sound. This tuning produces musical intervals between the pitches produced by successive strings of an octave, a perfect fifth, a perfect fourth, and a pure major third. This type of tuning is known as *just* intonation, and was widely practiced in Western Europe in the 16th and early 17th century. It has been largely replaced by the modern system of *equal temperament* in which the frequency ratios for successive pitches (semitones) are fixed at $\sqrt[12]{2}$, which allows a keyboard instrument to play in different musical key signatures without retuning.

Instruments such as the mbira that produce their sound by the vibration not of simple strings but of cantilevered rods have ratios of their overtone frequencies that are different from those of strings. It is therefore reasonable to expect that the tuning of such instruments would differ from just intonation, and this is indeed what has been observed by ethnomusicologists. In a 1932 study Hugh Tracey observed[11] that in some cases mbira players tuned some keys using an overtone in preference to the fundamental to produce the desired effect. Tracey also noted[12] that the accepted tuning of the instrument varied by region, with each local group of musicians agreeing upon a "correct" local manner of tuning. In 1978 Paul Berliner made a detailed study of the tunings used by Shona mbira players in Zimbabwe.[1] The intervals between the pitches of the mbira keys in five tunings that he observed are displayed in columns 3–7 in Table III (labeled by the names of the musicians from whom he recorded them), and the intervals expected for just intonation are included in column 2 for comparison. As expected, these tunings all differ from the just intonation that would be produced in tuning simple stretched strings "by ear."

A sixth tuning included in Table III is that specified by Kevin Volans, a South African composer of classical music in the Western European tradition. In his composition *White Man Sleeps* for two harpsichords, viola da gamba and percussion, he specifies[13] that the instruments are to be tuned as listed in column 8 of Table III, which he labels "a tuning system derived from Shona Mbira music."[14] As can be seen from Table III, it deviates from the tunings recorded by Berliner from Shona musicians no more than they deviate from one another.

The tuning of the mbira studied in this work is also presented in Table III. While the above-described measurement procedure allows confidence in the accuracy of the frequencies listed there, their relationship to a "proper" tuning is less certain. It is possible that the tuning of this instrument may have changed since it left the hands of its musician-maker, and thus that the tuning recorded here may not accurately represent the tuning it was intended to have. However, the fact that the keys are quite firmly fixed in place so that increasing or reducing the vibrating length requires considerable force (applied with a hammer to one end of the key) suggests that the tuning has probably not changed significantly since it was last adjusted. Inspection of Table III reveals that the intervals between almost all the keys of the mbira studied here are very similar to those used by Volans and by the musicians recorded by Berliner.

The question that originally stimulated this study was

TABLE III. Mbira tuning intervals in cents (an interval of one cent corresponds to 1/100 of an equal-tempered semitone, or a frequency ratio whose logarithm equals 2.5086×10^{-4}). Column 1 labels the mbira keys, in order of increasing fundamental frequency. Column 2 ("Just") represents the intervals between the pitches sounded by adjacent keys for tuning in just intonation (see the text). Column 3–7 are tunings used by Zimbabwean musicians recorded by Paul Berliner (Ref. 1), and column 8 is a tuning used by composer Kevin Volans (Ref. 10). Column 9 contains the intervals between the pitches of adjacent keys of the mbira measured in this work.

Key	Just	Gondo	Mude	Mujuru	Kunaka	Bandambira	Volans	Mbira
B1								
	386	323	174	126	455	355	360	323
B2								
	112	157	115	117	92	199	154	130
B3								
	204	164	286	156	210	96	171	204
B4								
	182	186	37	314	178	179	175	192
B5								
	112	151	158	105	154	153	100	181
B6								
	204	198	206		171		140	171
L1								
	204	151	170		241		171	182
B7								
	182	187	180		136		189	177
R1								
	112	180	180		126		154	127
L3								
	204	191	211	263	209	118	171	225
L2								
	182	137	170		181		175	181
L4								
	112	164	139	97	164	169	100	119
L5								
	204	182	128	264	161	193	140	213
L6								
	0				15		0	4
R2								
	204	179	231	190	196	185	171	177
R3								
	182	218	180	156	181	204	189	193
R4								
	112	136	118	143	129	204	154	125

TABLE III. (*Continued.*)

Key	Just	Gondo	Mude	Mujuru	Kunaka	Bandambira	Volans	Mbira
R5								
	204	182	214	219	170	163	171	190
R6								
	182	151	172	151	201	158	175	195
R7								
	112	161	173	128	173	137	100	135
R8								
	204	192	260	240	98	251	140	185
R9								

this: *Is the tuning of the mbira determined by the overtones of the vibrations of its keys?* From the evidence here presented, it would appear that the answer is "no." Because the first overtone is approximately five times the frequency of the fundamental (corresponding to a musical interval of approximately two octaves plus a major third), the overtones of only the lowest few keys overlap with the fundamental frequencies of any of the upper keys, and the matching of overtone frequencies with fundamentals is not particularly good. The large frequency differences involved presumably make the matching of the frequencies sufficiently difficult to hear that they do not strongly influence the tuning. The lower intensity of the overtones compared to the fundamental may also make their use in tuning impractical. For the mbira the first overtone is typically 20–30 dB lower in intensity than the fundamental. In stringed instruments or Western European keyboard instruments the first and second overtones can often be comparable in intensity to (or even louder than) the fundamental.[15] However, Berliner[2] describes how master mbira maker John Kunaka deliberately constructed the lowest key (B1) on his mbira to have "two voices," i.e., the fundamental and the first overtone, with the overtone sounding two octaves plus either a fifth or a third above the fundamental. Kunaka stated that this overtone "helped the music," but that the overtones of the other keys did not and were ignored. Andrew Tracey[16] noted that in some mbiras the fundamentals of the lower-pitched keys are almost inaudible, and the maker has tuned them so that the prominent overtone, rather than the fundamental, gives the desired note. In other cases the fundamental is used, but the overtone is wildly discordant, giving the instrument a "tinkling, metallic effect." He experimented with methods of tuning the overtone to a pitch two octaves above the fundamental by removing material from the key at appropriate locations. He was successful in doing so, but remarked that he had encountered only one instrument by an African maker in which this had been done.

V. CONCLUSION

We have measured the acoustic spectrum of the keys of an African mbira. We find that the most prominent overtones present in the spectrum have frequencies that are approximately 5 and 14 times the lowest frequency. A finite-element model of the vibration of the key that takes into account the acoustic radiation efficiency of the various normal modes reveals that the far-field power spectrum is dominated by modes involving predominately transverse motion of the key. A procedure to quantify the acoustic radiation produced by each normal mode has compared favorably to experimental results.

The finding that the most prominent overtones in the sound spectrum have very high frequencies relative to the fundamental makes it unlikely that the tunings of the mbira that are used by expert musicians are determined by matching the fundamental frequencies of the upper keys with the overtones of the lower keys.

ACKNOWLEDGMENTS

L.E.M. would like to thank Brent Wissick of the UNC-CH Department of Music for introducing her to *White Man Sleeps*, and for sparking the original question that this work attempts to answer. We would also like to thank Louise Meintjes of the Duke University Department of Music for lending us the mbira upon which the experiments were performed, and Paul Berliner of the same department as well as mbira master Cosmas Magaya for providing helpful information about the musical context of the mbira. Fred Brown of the UNC-CH Department of Physics and Astronomy provided valuable assistance and equipment for the recording and data processing, Phillip Thompson of the same department provided the dimensional data, and Elliott Moreton of the UNC-CH Department of Linguistics graciously allowed us to use the sound chamber. We are grateful to all three.

[1]P. F. Berliner, *The Soul of Mbira* (University of California Press, Berkeley, CA, 1978).
[2]P. Berliner, "John Kunaka, mbira maker," African Arts **14**, 61–67 (1980).
[3]D. H. Hodges, "A mixed variational formulation based on exact intrinsic equations for dynamics of moving beams," Int. J. Solids Struct. **26**, 1253–1273 (1990).
[4]V. L. Berdichevsky, "On the energy of an elastic rod," J. Appl. Math. Mech. **45**, 518–529 (1982).
[5]V. L. Berdichevsky and S. S. Kvashnina, "On the theory of curvilinear Timoshenko-type rods," Prikl. Mat. Mekh. **47**, 809–817 (1983).

[6]K. J. Bathe, *Finite Element Procedures* (MIT, Cambridge, MA, 2006).
[7]J. E. Ffowcs Williams and D. L. Hawkings, "Sound generation by turbulence and surfaces in arbitrary motion," Philos. Trans. R. Soc. London, Ser. A **264**, 321–342 (1969).
[8]S. Arendt and D. C. Fritts, "Acoustic radiation by ocean surface waves," J. Fluid Mech. **415**, 1–21 (2000).
[9]R. H. MacNeal, *Finite Elements: Their Design and Performance* (Dekker, New York, 1993).
[10]See for example N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, New York, 1991), Sec. 2.15.
[11]H. Tracey, "The mbira class of African instruments in Rhodesia (1932)," African Music: Journal of the African Music Society **4**, 78–95 (1969).
[12]H. Tracey, "Measuring African scales," African Music: Journal of the African Music Society **4**, 73–77 (1969).
[13]K. Volans, *White Man Sleeps* (Chester Music, London, 1990).
[14]Composer's statement, ⟨http://kevinvolans.com/kv_arti_whit.shtml⟩ (last viewed 1 June 2007).
[15]Examples of such spectra can be found in Ref. 10.
[16]A. Tracey, "The tuning of mbira reeds," African Music: Journal of the African Music Society **4**, 96–100 (1969).

Influence of microarchitecture alterations on ultrasonic backscattering in an experimental simulation of bovine cancellous bone aging

K. N. Apostolopoulos and D. D. Deligianni[a)]
Biomedical Engineering Laboratory, Department of Mechanical Engineering & Aeronautics, University of Patras, Rion 26500, Greece

(Received 19 June 2007; accepted 16 November 2007)

An experimental model which can simulate physical changes that occur during aging was developed in order to evaluate the effects of change of mineral content and microstructure on ultrasonic properties of bovine cancellous bone. Timed immersion in hydrochloric acid was used to selectively alter the mineral content. Scanning electron microscopy and histological staining of the acid-treated trabeculae demonstrated a heterogeneous structure consisting of a mineralized core and a demineralized layer. The presence of organic matrix contributed very little to normalized broadband ultrasound attenuation (nBUA) and speed of sound. All three ultrasonic parameters, speed of sound, nBUA and backscatter coefficient, were sensitive to changes in apparent density of bovine cancellous bone. A two-component model utilizing a combination of two autocorrelation functions (a densely populated model and a spherical distribution) was used to approximate the backscatter coefficient. The predicted attenuation due to scattering constituted a significant part of the measured total attenuation (due to both scattering and absorption mechanisms) for bovine cancellous bone. Linear regression, performed between trabecular thickness values and estimated from the model correlation lengths, showed significant linear correlation, with $R^2=0.81$ before and $R^2=0.80$ after demineralization. The accuracy of estimation was found to increase with trabecular thickness.

[DOI: 10.1121/1.2822291]

PACS number(s): 43.80.Jz, 43.80.Ev, 43.80.Qf, 43.20.Fn [CCC] Pages: 1179–1187

I. INTRODUCTION

Osteoporosis is a condition of reduced mass and increased fragility with a more pronounced effect on cancellous bone. Current clinical prediction of fracture risk is based primarily on bone mass alone. The primary method currently used for clinical bone assessment is based on x-ray absorptiometry and measures total bone mass at a particular anatomic site. Because other factors, such as architecture, also appear to have a role in determining an individual's risk of fracture, ultrasound is one alternative that has generated much attention (Luo *et al.*, 1999).

Ultrasound speed of sound and broadband ultrasound attenuation (BUA) determination has been explored as an alternative method for estimating bone density. Both ultrasonic parameters have demonstrated strong correlation with bone mineral density (BMD) in human cancellous bone (Han *et al.*, 1996; Nicholson *et al.*, 1994; Serpe and Rho, 1996). Transmission ultrasonic techniques are capable of explaining about 70% of the variability in observed bone density. Part of the reason for this is the fact that ultrasound seems to be affected, not only by how much bone is contained along the propagation pathway, but also by how that bone is organized in terms of its microstructure or architecture (Luo *et al.*, 1999; Gluer *et al.*, 1993). The variability can be up to 94% by including in the model a measure of the bone architecture (Turner *et al.*, 1990).

Correlations between BMD and backscatter properties have been found to be significant but lower than with transmitted ultrasound. Logarithm of backscatter coefficient exhibited a substantial correlation with BMD ($R^2=0.62$; Wear *et al.*, 2000). Variance of BMD accounted for 58% of the variance of broadband ultrasound backscatter in human femoral cancellous bone (Jenson *et al.*, 2006) and 79% in human calcaneous (Chaffaï *et al.*, 2002).

Ultrasonic backscattering properties may contain substantial information not already contained in transmission properties. Jenson *et al.* (2003) used various autocorrelation functions (Gaussian, exponential and densely populated medium) to compute backscatter coefficient from human bone and found good agreement between experimental data and theoretical predictions. Significant relationship was found between the estimated correlation length from these models and the mean trabecular thickness (Chaffaï *et al.*, 2002; Jenson *et al.*, 2003; Wear, 2003). However, the correlation between individual experimental and estimated trabecular thickness values is moderate in human bone ($R2=0.44$) (Padilla *et al.*, 2006).

A scattering model was proposed for the ultrasonic frequency-dependent backscatter in dense bovine cancellous bone, using a combination of two autocorrelation functions to describe the medium: one with discrete homogeneities (distribution of equal spheres) and another, which considers tissue as an inhomogeneous continuum (densely populated medium) (Deligianni and Apostolopoulos, 2007). The estimated correlation lengths were highly correlated with trabe-

a)Author to whom correspondence should be addressed. Electronic mail: deligian@mech.upatras.gr.

cular thickness, corresponding to groups of dominant trabecular sizes present in individual specimens. Thickness distributions for individual trabeculae for each bone specimen were obtained, and dominant trabecular sizes were determined instead of mean trabecular thickness, because of the large variation in bovine bone trabecular thickness. The reasons for the higher correlations in bovine than in human bone at the individual level, discussed in this work, include the different range of trabecular thickness.

Ultrasonic tissue characterization techniques are often based on the premise that disease processes alter physical characteristics of tissue and that these alterations can cause observable changes in acoustic scattering properties. It has been shown that changes in trabecular bone microstructure are related to increased incidence of osteoporotic fractures (Parfitt, 1987). The diminished number and thickness of trabeculae that are associated with increased fracture risk would be expected to change backscatter.

Demineralization models have been used to study the influence of bone mineral content on ultrasonic velocity and attenuation by measuring the acoustic properties at various stages of demineralization and to explore the utility of ultrasonic properties on bone tissue assessment (Duquette *et al.*, 1997; Hoffmeister *et al.*, 2002; Tavakoli and Evans, 1991; Wu *et al.*, 1998). In demineralized specimens decreases were produced in ultrasound velocity and BUA. The presence of organic matrix has been found that contributes very little to BUA (Duquette *et al.*, 1997).

Demineralization has been extensively used to modify cortical bone grafts to foster their osteoinductive properties and the process of demineralization has been investigated (Birkedal-Hansen, 1974; Broz *et al.*, 1995; Lewandrowski *et al.*, 1997). Results showed that in the demineralization of diaphyseal cortical bone specimens using hydrochloric acid, a uniformly thick circumferential band of demineralized bone matrix surrounds an inner intact bone core as the process of demineralization occurs. Immersion in aqueous acid chelates the mineral phase while preserving bone cellular detail, maintaining the molecular structure integrity and the mechanical properties of type I bone collagen. Discrete boundaries are observed between mineralized and demineralized parts of cortical bone specimens (Lewandrowski *et al.*, 1997).

The effects of aging on trabecular bone are decrease of trabecular thickness and removal of the thinner trabeculae. Based on the results of demineralization of compact bone, which is formation of a sharp interface between demineralized and mineralized section of bone, an experimental model was developed in this study which can simulate physical changes that occur during aging by acid demineralization. The used method was to reduce progressively the mineral content of each sample by acid treatment while monitoring the ultrasonic characteristics as a function of density changes. The use of bovine bone which possesses different structure and higher density than human is a weakness of the model. The aim of the study was to investigate experimentally the change of the ultrasonic transmission and backscatter properties with the diminished thickness of trabeculae, related to aging and osteoporosis, and to assess the sensitivity of backscatter to microarchitecture alterations. An objective was to identify the relative roles played by mass and architecture as reflected in ultrasonic propagation.

II. EXPERIMENTAL PROCEDURE

A. Bone specimens

Samples of cancellous bone were obtained from bovine femora. Transverse cuts, oriented parallel to the articular surface, were prepared using a band saw. From these sections with a slow speed diamond wafering saw, a total of 40 cubic specimens were prepared with a side length of 22–25 mm.

Bone marrow was removed by alternately water jetting and immersing the samples in trichloro-ethylene solution in an ultrasound bath. The apparent density (the ratio of the dehydrated, defatted tissue mass to the total specimen volume) of the samples was determined from separate measurements of mass and volume. Mass was measured with a balance, after centrifuging the specimen for 10 min to remove excess water. Volume was determined by measuring the exterior dimensions of the cubes with calipers.

B. Simulation of aging

Bone aging was simulated by successive demineralization of fresh cubes of bovine trabecular bone at five stages by forcing hydrochloric acid solution (0.25 M) to flow through the specimen for five total time periods: 40 s, 2 min, 4 min, 8 min and 15 min. The flow path was controlled by sealing all but two opposite faces of the cubic specimen using a thin impermeable rubber skin. The solution was forced to flow in and out through the interrogated opposite sides of the specimen with the aid of a syringe pump. In this way, mineral was removed from the surface of the pores leaving the organic collageneous structure largely intact and resulting in decrease of trabecular mineralized thickness or in full removal of mineral from some thin trabeculae. Subsequently, the specimens were placed in an alkaline solution (5% ammonium hydroxide solution) to neutralize the remaining acid and stop the demineralization process. The specimens were then thoroughly rinsed with circulating water.

Demineralization was determined by apparent density measurement. For most specimens two demineralization stages were performed. For a few specimens up to five successive demineralization stages were succeeded. Five specimens were left in hydrochloric acid until no more reduction of apparent density was obtained (more than 24 h). These specimens were considered as completely demineralized. Specimens with low initial apparent density were completely demineralized after two or three demineralization stages.

Following this step, the specimens were sectioned in slices of about 1 mm thick with a diamond saw of a microtome. The slices were stained with von Kossa technique which stains calcium with a dark brown color. The calcium deficient areas were visualized as light brown to white strips around pores and along trabeculae (Fig. 1). The disadvantage of this technique is that it was difficult to discriminate the light-color strips in very short times of demineralization (40 s). Another qualitative method of visualization of the calcium deficient areas was also used: it was the composition

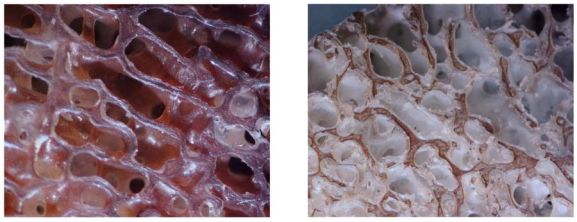


FIG. 1. (Color online) Von Kossa staining of acid-treated samples demonstrated the presence of a demineralized tissue layer surrounding pores whose thickness depended on immersion time. (A): Immersion in hydrochloric acid for 4 min. (B): Immersion for 15 min.

images technique provided by the scanning electron microscope (SEM). With this technique each element present on a surface can be displayed in a different color. In these images demineralized areas are illustrated by yellow and red dots and calcium by blue dots [Fig. 2(A)]. It can be observed that no detectable amounts of calcium were present in the demineralized tissue layer surrounding the core. The distribution of calcium content along trabecular thickness is displayed in Fig. 2(B).

C. Ultrasonic measurements

The ultrasonic measurements of the bone specimens were performed before and after acid treatment of the specimens in distilled water at room temperature, using methods previously reported (Deligianni and Apostolopoulos, 2007). The specimens were degassed in a vacuum container to remove air bubbles. A custom holder was used to position the specimen in the US beam. Focused (Panametrics, V303, d =1.27 cm (0.5 in.), center frequency 1 MHz, spherical focus F=20 mm) and unfocused (Panametrics, V303, d=1.27 cm (0.5 in.), center frequency 1 MHz) immersion transducers have been used in this study, connected to an ultrasonic pulser receiver (USD 10NF, Krautkraemer, Germany). Received rf signals were digitized at 35 MHz.

The measurements of ultrasonic speed of sound (SOS) and attenuation were performed in a through-transmission substitution method, using a pair of coaxially aligned unfocused transducers. The fast wave was analyzed. To minimize the effect of frequency-dependent attenuation on measured speed of sound, the pulse transit time was calculated at the first zero crossing of the signal (Haiat *et al.*, 2006). To determine broadband ultrasound attenuation (BUA), linear regression of ultrasound attenuation against frequency between 0.2 and 0.9 MHz was performed. The slope of the fitted line was divided by the specimen thickness to obtain the value of normalized broadband ultrasound attenuation (nBUA) with units of dB $MHz^{-1}cm^{-1}$.

For backscatter measurements, the same system was used in pulse-echo mode using a single focused transducer acting as transmitter-receiver. The frequency-dependent backscatter coefficient was derived following the method of Roberjot *et al.* (1996) and Chaffaï *et al.* (2002). A substitution technique was used, in which the signal scattered from the region being tested was compared to the signal from a standard reflecting target (a steel plate). The transducer output was gated appropriately to permit analysis of a 7-mm-long region. Measurements of backscatter coefficient on the specimens were performed along two-dimensional scans in steps of 1 mm. Thus, the volume of the interrogating area for spatial averaging was $20\times20\times7$ mm^3, centrally located within each specimen. The values of the backscatter coefficient were averaged over the measurement points and were compensated for attenuation (O'Donell and Miller, 1981), Hamming gate function and diffraction (Xu and Kaufman, 1993).

D. Image analysis and trabecular thickness estimation

A number of specimens (n=18, including specimens all over the density range) were subjected to a single demineralization stage, remaining in hydrochloric acid solution for 2 or 4 min, depending on their initial apparent density. Images of the von Kossa stained slices of the interrogated area of each sample were taken under a stereomicroscope. The images were captured as color images of a resolution of 2560 $\times$2048 pixels. Individual measurements of trabecular thickness were performed on each specimen. Each trabecula consisted of a dark brown, the undecalcified area, and a white, the calcium deficient area. Two values of trabecular thickness were measured from the images: one of the dark brown area, which corresponded to the thickness after demineralization

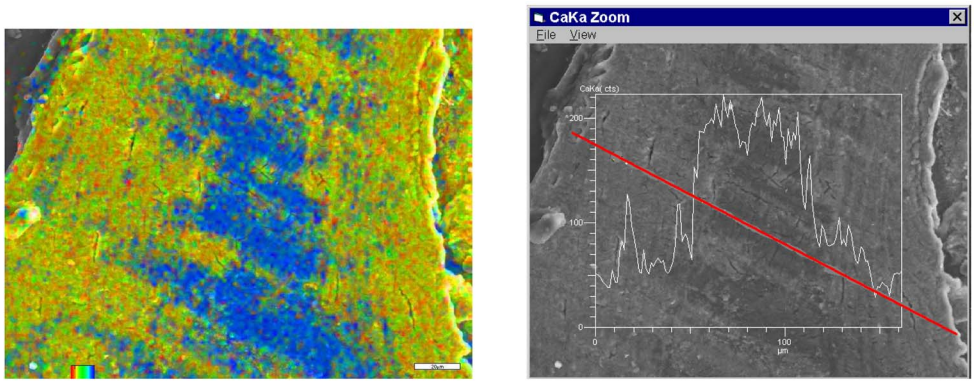


FIG. 2. (Color online) (A): SEM image from a specimen immersed in hydrochloric acid for 15 min. With the composition images technique each element present on a surface is displayed in a different color. The presence of calcium is illustrated by blue dots. (B): The elemental line scan displays the distribution of calcium content along trabecular thickness.

and the other of the dark brown and white parts together, which corresponded to the whole trabecular thickness.

Trabecular thickness was measured by an in-house code, based on a method developed by Kothari *et al.* (1999). Estimates of the thickness dimension were made in directions perpendicular to the trabecular orientation. They were measured at five equispaced points along the length of the trabecula. The used code was validated on simulated images. Thickness distributions for individual trabeculae were obtained for each bone specimen before and after acid treatment.

Trabecular thickness was compared to estimates of correlation length from backscatter coefficient measurements. Experimental and theoretical data were compared using a two-component model based on two autocorrelation functions, a densely populated and a spherical distribution function, presented elsewhere (Deligianni and Apostolopoulos, 2007). The best fit between experimental data and predictions, obtained by computing a least-square regression, yielded the value of an estimate of the correlation length a. The experimental data and the theoretical results fitted to within 5 dB, that is, the maximum accepted offset of the theoretical curve to the experimental was ± 5 dB. The calculation of the mean square fluctuation (velocity and density fluctuation) in medium properties $\langle \gamma^2 \rangle$ was based on the treatment of materials as immiscible mixtures and, for a two-component mixture, $\langle \gamma^2 \rangle$ was determined according to Sehgal (1993) as a function of porosity (Appendix). The numerical value of $\langle \gamma^2 \rangle$ has been estimated by taking $\rho_b = 1750$ kg/m^3, $\rho_w = 1000$ kg/m^3, $c_b = 3000$ m/s and $c_w = 1470$ m/s (Han *et al.*, 1996), constant throughout a given specimen as well as for different individuals (c_b, c_w, ρ_b, ρ_w are the velocities and densities in solid bone and water, respectively).

The predicted nBUA due to scattering by the densely populated model was also calculated (Fig. 4) and compared with the experimentally determined nBUA. Predicted backscatter coefficient and nBUA due to scattering were calculated as functions of porosity. Porosity was converted to apparent density by applying the formula: $\rho_{appden} = \rho_b - \rho_b * porosity/100$.

III. RESULTS

Von Kossa staining of all hydrochloric acid-treated samples demonstrated the presence of a demineralized tissue layer surrounding pores whose thickness depended on immersion time (Fig. 1). The position of the reaction front at a certain demineralization stage (related but not proportional to immersion time) was confirmed by evaluation of the corresponding calcium dot maps [Figs. 2(A) and 2(B)]. The elemental line scans showed that no detectable amounts of calcium were present in the demineralized tissue layer. Figure 2(B) displays the distribution of the calcium content along trabecular thickness. A centrally symmetric area with high calcium content can be observed, whereas the edges were sufficiently demineralized.

The apparent densities before demineralization of the 40 specimens ranged from 0.306 to 0.969 g/cm^3. After demineralization the minimum observed apparent density was 0.120 g/cm^3. Figures 3–5 display the correlation of transmission and backscatter ultrasound parameters with apparent density of undemineralized and demineralized specimens after a number of successive stages of demineralization at an interspecimen level. The precision of estimate of the ultrasonic properties resulted in a root-mean-square average coefficient of variation across the samples of 3.45% for backscatter coefficient at 800 kHz, 2.11% for nBUA and 0.65% for speed of sound. Ultrasonic speed of sound for apparent densities higher than 0.25 g/cm^3 (undemineralized and demineralized specimens) displayed a linear relationship with apparent density ($R^2 = 0.91$) (Fig. 3). In the final demineralization stage apparent densities lower than 0.2 g/cm^3 and values of speed of sound lower than 1500 m/s were observed. In these low densities, the slope of the linear relationship with apparent density decreased (Fig. 3). Completely demineralized samples displayed speed of sound approximating that of water (1470 m/s).

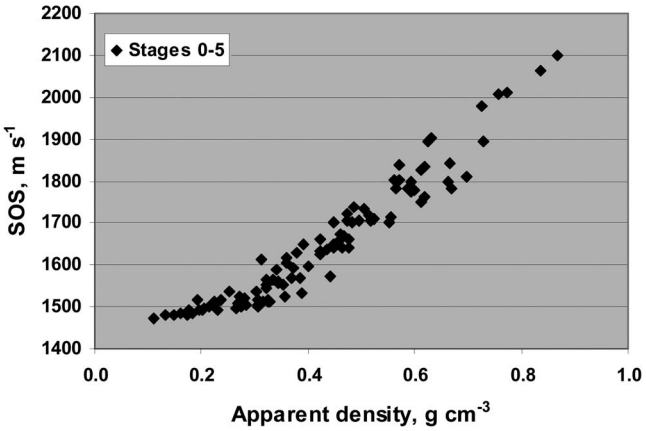


FIG. 3. Correlation of speed of sound (SOS) with apparent density including all specimens for all demineralization stages.

Figure 4 illustrates nBUA of undemineralized and demineralized specimens as a function of apparent density for all demineralization stages. The predicted nBUA due to scattering for the densely populated medium autocorrelation

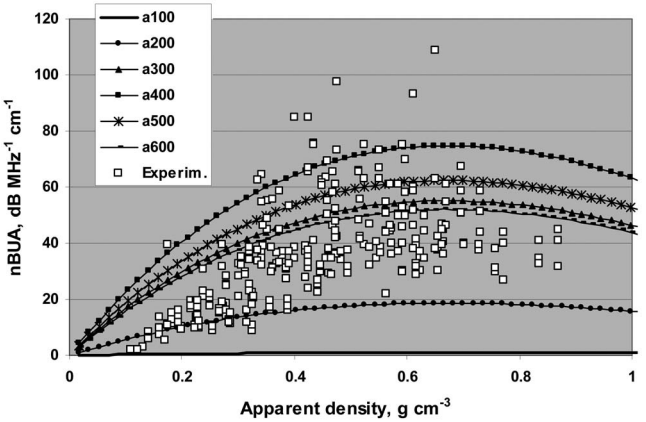


FIG. 4. Correlation of normalized broadband ultrasonic attenuation (nBUA) with apparent density including all specimens for all demineralization stages. Data were analyzed over the bandwidth 0.2–0.9 MHz. The line curves display the predicted nBUA due to scattering from the densely populated medium model and for correlation lengths from 100 to 600 μm.

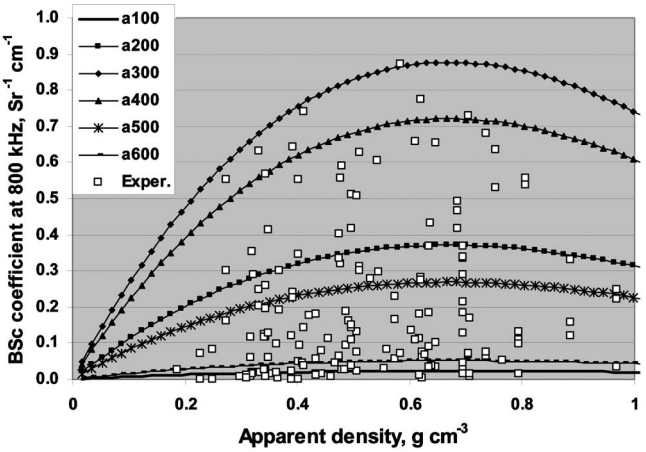


FIG. 5. Correlation of backscatter coefficient at 800 kHz with apparent density including all specimens for all demineralization stages. Data were analyzed over the bandwidth 0.4–1.2 MHz. The line curves display the predicted values of backscatter coefficient at 800 kHz from the densely populated medium model for correlation lengths from 50 to 600 μm.

function and for correlation lengths from 100 to 600 μm are also shown in this figure. The experimental nBUA values lie in regions where scatterers larger than 200 μm were predicted. However, scatterers with sizes of 100 and 150 μm were measured microstereoscopically in the specimens. The maximum predicted by the model attenuation due to scattering (for scatterer sizes of 400 μm, encountered in bovine bone) was lower than the maximum measured total attenuation (due to both scattering and absorption mechanisms) in bovine cancellous bone specimens, but still constituted a significant part of it (about 75%). Maximum experimental values of nBUA were observed at apparent density of approximately 0.60–0.65 g/cm^3 or 62–66% porosity. Predicted nBUA due to scattering displayed maximum value at the same apparent density. nBUA of completely demineralized samples tended to zero.

Figure 5 depicts the experimentally determined backscatter coefficient at 800 kHz (the middle of the usable bandwidth) for all demineralization stages. The predicted backscatter coefficient for the densely populated medium autocorrelation function, as a function of apparent density, for correlation lengths from 50 to 600 μm with constant scatterer concentration (porosity 70%) are also depicted. The model predicted the experimental backscatter coefficient magnitude. Both experimental and values and prediction displayed a maximum at apparent density of about 0.6–0.65 g/cm^3.

Figure 6 illustrates the change with apparent density of ultrasonic parameters SOS, nBUA and backscatter coefficient at 800 kHz, respectively, after five successive stages of demineralization of a representative specimen. All specimens displayed a linear correlation of speed of sound with apparent density in the demineralization process and indicated a correlation (R^2) greater than 0.97 in all cases [Fig. 6(A)]. Thus, ultrasonic speed of sound is the parameter most sensitive to changes in bone mineral density alone. Specimens with higher apparent densities at the base line demonstrated a nonlinear relationship of nBUA with apparent density. As apparent density decreased at the successive demineralization stages, nBUA increased up to apparent density of 0.65 g/cm^3 and then it decreased with apparent density. For specimens with values of apparent densities at the base line lower than 0.65 g/cm^3, this relationship was approximately linear [Fig. 6(B)]. Backscatter coefficient at 800 kHz demonstrated similar behavior with nBUA at the successive demineralization stages [Fig. 6(C)]. Specimens with higher apparent densities at the base line had a nonlinear relationship of backscatter coefficient at 800 kHz with apparent density. The maximum value occurred for an apparent density of 0.65 g/cm^3. In the low density region, backscatter coefficient at 800 kHz decreased as apparent density decreased.

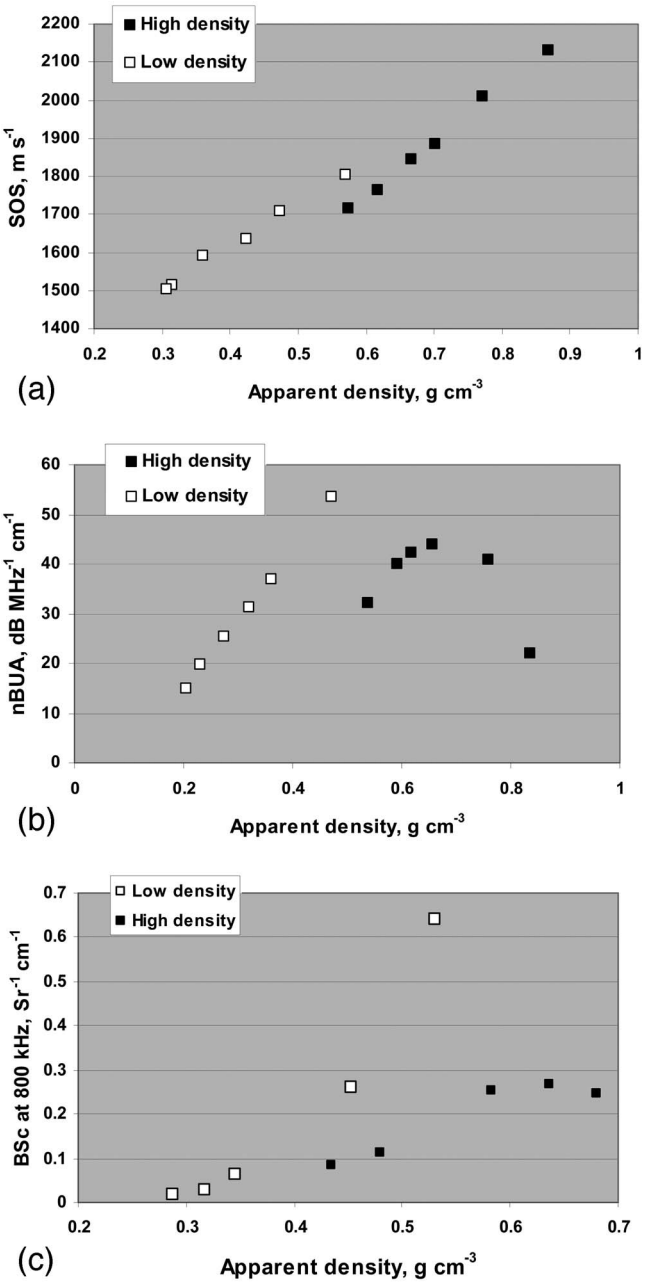


FIG. 6. Change with apparent density of ultrasonic parameters SOS (A), nBUA (B) and backscatter coefficient at 800 kHz (C), respectively, after five successive stages of demineralization of representative specimens. For specimens with higher base line apparent densities, nonlinear relationship of nBUA and backscatter coefficient at 800 kHz with apparent density was observed.

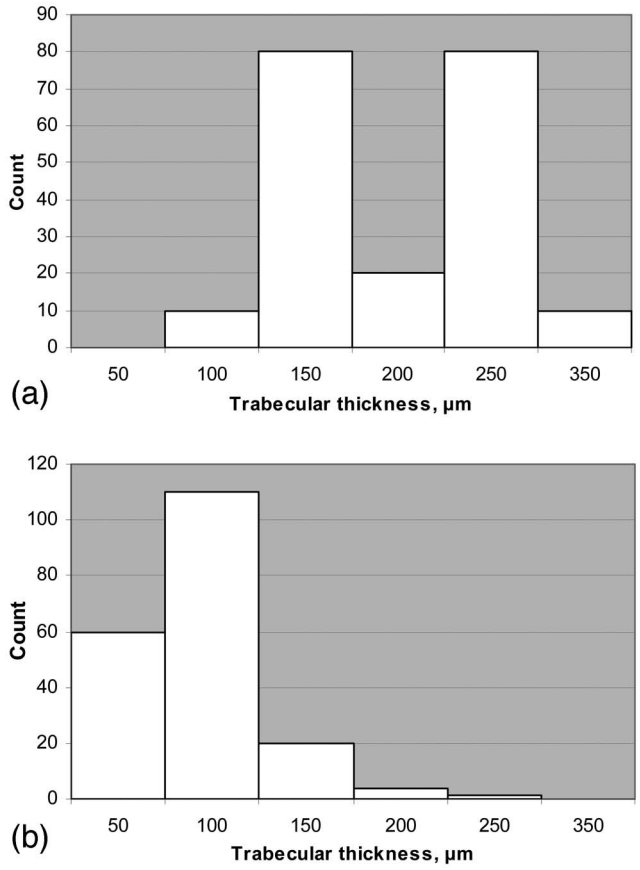


FIG. 7. Thickness distributions for individual trabeculae for a representative bone specimen before (A) acid treatment and after one demineralization stage (B).

Because of the high value of standard deviation of the mean trabecular thickness of the specimens (198 ± 102 μm), it was considered that a single mean value would not be characteristic of the architectural structure of the specimens. Thus, thickness distributions for individual trabeculae for each bone specimen before acid treatment and after one demineralization stage were obtained [Fig. 7(A)]. The thickness distribution graphs revealed the presence of one, two or, rarely, three dominant trabecular sizes for each specimen. Lower trabecular thickness was observed after demineralization [Fig. 7(B)].

Figure 8 shows the change of backscatter coefficient with frequency for a representative specimen after 40 s immersion in hydrochloric acid. It can be observed that the partial maximum values of the curves have shifted to the right in the frequency axis, indicating the decrease of trabecular thickness, according to the theoretical model (Deligianni and Apostolopoulos, 2007). For each correlation length, the maximum value of the backscatter coefficient appears at a certain frequency, increasing with decreasing correlation lengths. In the experimental curves, local maxima of backscatter coefficient (one or more for each specimen) are observed at certain values of frequency, which correspond to a correlation length.

Figure 9 represents the measured trabecular thickness (corresponding to one or more dominant scatterer sizes of trabeculae for each specimen) as a function of the estimated

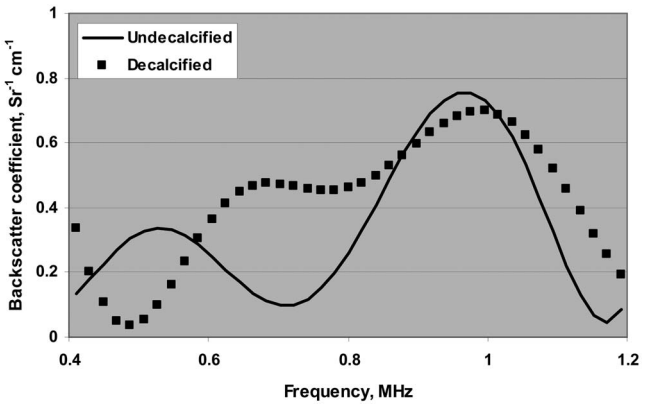


FIG. 8. Change of backscatter coefficient with frequency for a representative specimen after 40 s of immersion in hydrochloric acid.

by the two-component model correlation lengths, before and after acid treatment. Linear regression, performed between trabecular thickness values and estimated correlation lengths, showed significant linear correlation, with $R^2=0.81$ ($p<0.0001$) for undemineralized specimens and $R^2=0.80$ ($p<0.0001$) after one demineralization stage.

IV. DISCUSSION

The development of a tool to noninvasively monitor microarchitectural changes that occur during aging in cancel-

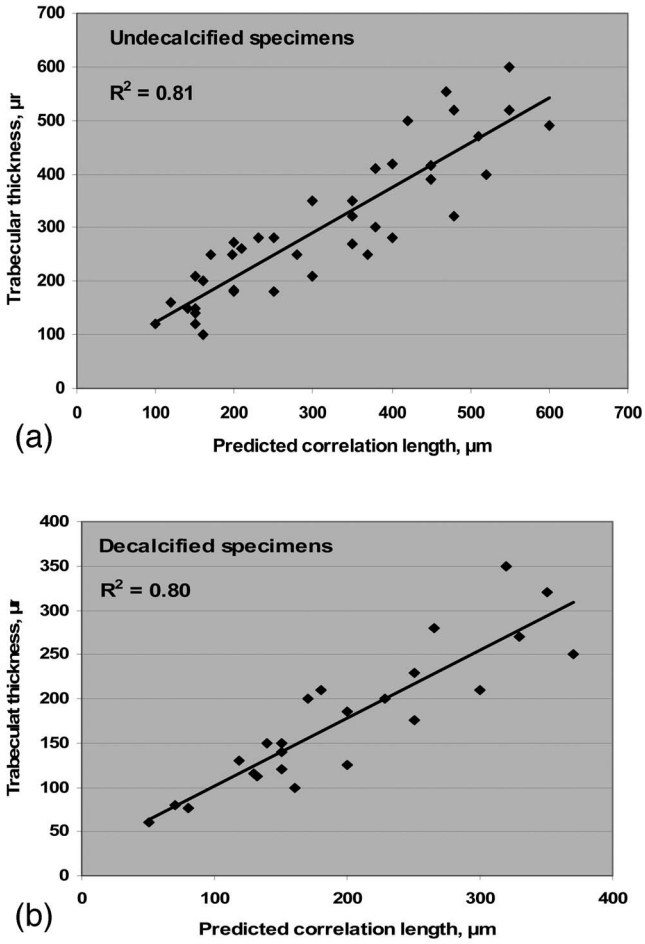


FIG. 9. Trabecular thickness, before (A) ($R^2=0.81$) and after (B) ($R^2=0.80$) hydrochloric acid treatment, versus correlation length estimated from ultrasonic measurements.

lous bone was a motivating factor in this study. An experimental model which can simulate these changes was developed in order to evaluate the effects of change of mineral content and microstructure on ultrasonic properties of bovine cancellous bone. Scanning electron microscopy and histological staining of the timed immersed in hydrochloric acid samples demonstrated a heterogeneous structure consisting of a mineralized core and a demineralized layer. The nature of the demineralization induced by the acid attack was that the mineral was removed leaving the organic collagen largely intact. When the acid enters the pores, the mineral is removed at the bone-fluid interface first (pore surface) and trabeculae are getting thinner, without considerably affecting the mineralized tissue beyond the interface. The bone mineral is totally removed from a thin layer of the surface. The density after demineralization is a mean value between approximately a totally demineralized and an intact area.

On the basis of the observations of the appearance of the reaction front, it was concluded that demineralization of bovine cancellous bone resulted in a sharp interface separating the outer demineralized portion from the inner undemineralized area. Similar observations have been done on cortical bone tooth enamel during demineralization process (Birkedal-Hansen, 1974; Broz *et al.*, 1995; Lewandrowski *et al.*, 1997). The discrete interface between the mineralized core and collagen layer was observed in SEM micrographs of acid-treated cortical samples. At this interface a gradient in the weight percent of calcium and phosphor occurred over approximately 20–25 μm width in each of the acid-treated samples. Within the core the mineralization, microhardness and acoustic properties were unaffected with respect to the control. The wave velocities of the mineralized cores were within 2% of the control value (Broz *et al.*, 1995).

The presence of organic matrix contributed very little to BUA and speed of sound. At complete demineralization, ultrasonic measurements were only 3–4% of the values at base line. Bone matrix that contains little mineral (as in the last stages of demineralization in this study) possesses ultrasonic properties which are similar to the water used as a coupling medium. The small contribution of organic matrix to BUA has also been reported by Duquette *et al.* (1997). A disadvantage of the model was that the resulting demineralization depth depended largely on specimen density and pore size and it was impossible to standardize immersion time. This behavior was dissimilar to that of ivory dentine specimens of certain geometrical shapes (plane sheet) for which the demineralization front distance during acid demineralization has been found to be directly proportional to the square root of the immersion time (Birkedal-Hansen, 1974). Another disadvantage was that the visualization methods which were used lacked the ability to discriminate the demineralized layer at very low immersion times. An improvement and refinement of the methods is necessary.

Simulated aging process led to trabecular thinning, loss of connectivity or removal of the thinnest trabeculae. Osteoporotic changes in trabecular bone are complicated and include bone mineral content, crystal composition and size, and matrix content and composition alterations (Faibish *et al.*, 2006). However, the gross changes are that both mineral and matrix are removed leaving the mineral/organic ratio essentially unchanged (Boskey *et al.*, 2005; Chappard *et al.*, 2007). Although, physiological redistribution of trabecular material occurs much more selectively in osteoporosis [preferential thinning of transverse trabeculae (McDonnell *et al.*, 2007)], the partial demineralization of cancellous bone at varying times of immersion can be considered as a good mimic of the changes found also in osteoporosis.

A clearer understanding of the role played by trabecular architecture in conjunction with bone mass and how this role is reflected in ultrasonic properties of bone is important for dealing with prognosis as effectively as possible with bone loss disorders. All three ultrasonic parameters were sensitive to changes of apparent density of bovine cancellous bone. If the analysis is confined to the lower density values where both nBUA and backscatter coefficient decrease with density, reduction of the apparent density at 10% produced a decrease of approximately 18% in speed of sound (considering the reduction relative to the speed of sound in water), 12.5% in nBUA and 35% in backscatter coefficient at 800 kHz.

The propagation of fast and slow longitudinal waves in bovine cancellous bone has been reported (Cardoso *et al.*, 2003; Hosokawa and Otani, 1997; Williams, 1992) when the acoustic wave propagates in parallel to the trabecular alignment. The presence of the slow wave depends on the orientation between the mean axis and the direction of main orientation of bone trabecular network. It tends to disappear at high angles to the trabecular alighment. In this work, measurements were taken from bovine femora, which do not present strong trabecular alignment, and the slow wave was not usually observed. Thus, only the fast wave was analyzed.

Strongly linear correlation was found in this study between speed of sound and apparent density at both intraspecimen and interspecimen level. Other studies comparing speed of sound to BMD in human calcaneous and femoral bone (Han *et al.*, 1996; Hoffmeister *et al.*, 2002; Nicholson *et al.*, 1994; Serpe and Rho, 1996) as well as in bovine bone (Cardoso *et al.*, 2003; Lee *et al.*, 2003; Tavakoli and Evans, 1991), have also demonstrated highly linear correlations over the entire range of density. The relationship of nBUA with apparent density was nonlinear. This is consistent with the results of other studies (Han *et al.*, 1996; Hodgskinson *et al.*, 1996; Serpe and Rho, 1996). The fast wave BUA has shown a parabolic behavior and reached a maximum for 75% porosity in human and bovine cancellous bone (Cardoso *et al.*, 2003). Lin *et al.* (2001), verifying their data with a stratified model, found that the maximum nBUA of sheep bone occured at porosity of 60% or at an apparent density of 800 kg m^{-3} and Hodgskinson *et al.* (1996) at 75% porosity. Potential differences in trabecular architecture might play a role in the peak porosity value, as well as the way of determination of porosity or density of compact bone.

Backscatter coefficient is also a nonlinear function of apparent density like nBUA. The nonlinearity of nBUA has possibly an impact on the estimation of backscatter coefficient. The frequency-dependent attenuation curves were used to compensate for attenuation when estimating the backscatter coefficient. Consequently, the same nonlinear behavior was observed for backscatter coefficient: there is a maximum

value at an apparent density of about 0.65 g/cm^3. In human calcaneous bone ultrasonic backscatter has been found to increase with bone mineral density and its logarithm showed a moderate linear correlation with bone mineral density (Wear and Armstrong, 2000). In the present study, during the demineralization process, at the individual level for low base line apparent densities of bovine specimens, a linear correlation of the logarithm of backscatter coefficient with apparent density has been found. Moreover, for bovine specimens with apparent densities at the base line lower than 0.65 g/cm^3, nBUA decreased approximately linearly with density [Fig. 6(B)], like human bone's nBUA (Han *et al.*, 1996; Hodgskinson *et al.*, 1996). Demineralized bovine bone at lower densities displayed ultrasonic behavior with similar trends to that of human bone from elderly subjects. Further experiments with the same experimental setup on human and bovine bone are necessary to verify whether demineralized bovine bone can be a model for human bone.

Significant linear correlation was found between predicted correlation lengths and measured trabecular thickness (R^2=0.81). Thus, a good prediction of the thickness of the dominant groups of trabeculae was obtained by the two-component backscatter model. Lower correlation coefficients (R^2=0.44–0.53) have been found between individual experimental and estimated trabecular thickness for human cancellous bone (Jenson *et al.*, 2003; Padilla *et al.*, 2006). A reason for better correlations obtained for bovine bone may be the consideration of the presence of a single dominant structure, the mean value of trabecular thickness, in human bone. There is a possibility that the scalloping, observed in some experimental curves of backscatter from human trabecular bone, is due to the presence of larger scatterers that produce undulated spectra. The distribution of scatterers sizes of bovine bone in a broader range might be another reason. In Fig. 10(A) the variation of backscatter coefficient at 800 kHz with trabecular thickness due to trabecular size distribution is illustrated. Over this range, the variation corresponding to the mean ± 25 μm of the values of trabecular thickness used in the thickness distribution graphs, depends on trabecular size. A change of 25 μm in trabecular thickness of 150 μm would correspond to a change of 0.40 in log magnitude of backscatter coefficient, in trabecular thickness of 200 μm to a change of 0.18 and in trabecular thickness of 250 μm to a change of 0.09. Similarly, a change of 25 μm in trabecular thickness of 275 μm would correspond to a shift of the maximum of backscatter coefficient at 96 kHz in the frequency axis, in trabecular thickness of 325 μm to a shift at 68 kHz and in trabecular thickness of 375 μm to a shift of 34 kHz [Fig. 10(B)]. The accuracy of estimation increases as the trabecular thickness increases. Nevertheless, this study along with others (Chaffaï *et al.*, 2002; Jenson *et al.*, 2003; Wear, 2003) documents the relationship of ultrasound backscatter to bone microarchitrecture and supports the view that this parameter provides direct access to microstructural parameters. Whether the conclusion drawn from this study can be generalized to bone affected by different diseases remains uncertain and additional studies will be useful to clarify this issue.

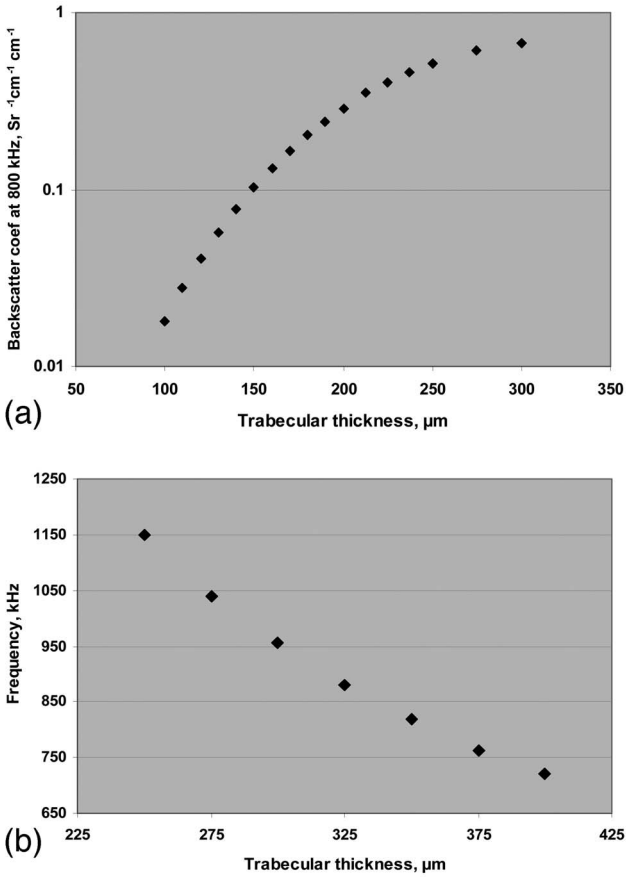


FIG. 10. (A): Predicted backscatter coefficient at 800 kHz as a function of trabecular thickness. (B): Predicted frequency dependence of the maximum value of backscatter coefficient. Over the range of trabecular thicknesses, the variation of the magnitude or the frequency where the maximum value of backscatter coefficient is observed, corresponding to the mean ± 25 μm of the values of trabecular thickness, depends on trabecular size.

This study addresses how relatively large changes in mineral content affect ultrasonic properties. The visualization methods, used in this study, cannot discriminate smaller changes, comparable to those produced by certain diseases or aging at early stages. A study that systematically investigated the effect of smaller changes would have improved clinical relevance.

The clinical implication of these data is not as obvious and they cannot be transferred to human bone. Although the results of the study of bovine bone ultrasonic behavior are not directly applicable for diagnostic purposes, the usefulness of this experimental model for the understanding of the ultrasonic behavior of cancellous bone during reducing mineral content may be significant. The specimens in most studies are obtained from elderly subjects. It would be of much interest to determine the relationships between BMD, microarchitecture and backscatter during demineralization of younger human bone tissue.

ACKNOWLEDGMENTS

This project was financially supported by the project "K. Karatheodori" of the Research Committee of the University of Patras. The authors are grateful to V. Cotsopoulos, Uni-

versity of Patras, for SEM images and E. Chota, MD, ASKLIPIOS, for BMD (QCT) measurements.

APPENDIX

The calculation of the mean square fluctuation in medium properties $\langle \gamma^2 \rangle$ is based on the treatment of materials as immiscible mixtures. For a two component mixture, $\langle \gamma^2 \rangle$ is defined as [Sehgal (1993)]:

$$\langle \gamma^2 \rangle = \langle \mu^2 \rangle + \langle \delta^2 \rangle \sin^4\left(\frac{\alpha}{2}\right)$$

$$\langle \mu^2 \rangle = \phi(1-\phi)\left(1 - \phi + \phi\left(\frac{c_b}{c_w}\right)^2\right)\frac{(c_b - c_w)^2}{c_b^2}$$

$$\langle \delta^2 \rangle = \phi(1-\phi)\left(1 - \phi + \phi\left(\frac{\rho_b}{\rho_w}\right)^2\right)\frac{(\rho_b - \rho_w)^2}{\rho_b^2},$$

where $\langle \mu^2 \rangle$ is the mean square of velocity fluctuation over scattering volume, $\langle \delta^2 \rangle$ is the mean square of density fluctuation, c_b, c_w, ρ_b, ρ_w are the velocities and densities in solid bone and water, respectively, and ϕ is the porosity of the specimen.

Birkedal-Hansen, H. (**1974**). "Kinetics of acid demineralization in histologic technique," J. Histochem. Cytochem. **22**, 434–441.

Boskey, A. I., DiCarlo, E., Paschalis, E., West, P., and Mendelsohn, R. (**2005**). "Comparison of mineral quality and quantity in iliac crest biopsies from high- and low-turnover osteoporosis: An FT-IR microspectroscopic investigation," Osteoporosis Int. **16**, 2031–2039.

Broz, J. J., Simske, S. J., and Greenberg, A. R. (**1995**). "Material and compositional properties of selectively demineralized cortical bone," J. Biomech. **28**, 1357–1368.

Cardoso, L., Teboul, F., Sedel, L., Oddou, C., and Meunier, A. (**2003**). "In vitro acoustic waves propagation in human and bovine cancellous bone," J. Bone Miner. Res. **18**, 1803–1812.

Chaffaï, S., Peyrin, F., Nuzzo, S., Porcher, R., Berger, G., and Laugier, P. (**2002**). "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microstructure," Bone (N.Y.) **30**, 229–237.

Chappard, D., Josselin, N., Rougé-Maillart, C., Legrand, E., Baslé, M. F., and Audran, M. (**2007**). "Bone microarchitecture in males with corticosteroid-induced osteoporosis," Osteoporosis Int. **18**, 487–494.

Deligianni, D., and Apostolopoulos, K. (**2007**). "Characterization of dense bovine cancellous bone tissue microstructure by ultrasonic backscattering using weak scattering models," J. Acoust. Soc. Am. **122**, 1180–1190.

Duquette, J., Honeyman, T., Hoffman, A., Ahmadi, S., and Baran, D. (**1997**). "Effect of bovine bone constituents on broadband ultrasound attenuation measurements," Bone (N.Y.) **21**, 289–294.

Faibish, D., Ott, S., and Boskey, A. (**2006**). "Mineral changes in Osteoporosis: A review," Clin. Orthop. Relat. Res. **443**, 28–38.

Gluer, C. C., Wu, C. Y., and Genant, H. K. (**1993**). "Broadband ultrasound attenuation signals depend on trabecular orientation: An in vitro study," Osteoporosis Int. **3**, 185–191.

Haiat, G., Padilla, F., Cleveland, R. O., and Laugier, P. (**2006**). "Effects of frequency-dependent attenuation and velocity dispersion on in vitro ultrasound velocity measurements in intact human femur specimens," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 39–51.

Han, S., Rho, J., Medige, J., and Ziv, I. (**1996**). "Ultrasound velocity and broadband attenuation over a wide range of bone mineral density," Osteoporosis Int. **6**, 291–296.

Hodgskinson, R., Njeh, C. F., Whitehead, M. A., and Langton, C. M. (**1996**). "The non-linear relationship between BUA and porosity in cancellous bone," Phys. Med. Biol. **41**, 2411–2420.

Hoffmeister, B. K., Whitten, S. A., Kaste, S. C., and Rho, J. Y. (**2002**). "Effect of collagen and mineral content on the high-frequency ultrasonic properties of human cancellous bone," Osteoporosis Int. **13**, 26–32.

Hosokawa, A., and Otani, T. (**1997**). "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562.

Jenson, F., Padilla, F., and Laugier, P. (**2003**). "Prediction of frequency-dependent ultrasonic backscatter in cancellous bone using statistical weak scattering model," Ultrasound Med. Biol. **29**, 455–464, 2003.

Jenson, F., Padilla, F., Bousson, V., Bergot, C., Laredo, J.-D., and Laugier, P. (**2006**). "In vitro ultrasonic characterization of human cancellous femoral bone using transmission and backscatter measurements: Relationships to bone mineral density," J. Acoust. Soc. Am. **119**, 654–663.

Kothari, M., Keaveny, T. M., Lin, J. C., Newitt, D. C., and Majumdar, S. (**1999**). "Measurement of intraspecimen variations in vertebral cancellous bone architecture," Bone (N.Y.) **25**, 245–250.

Lee, K. I., Roh, H. S., and Yoon, S. W. (**2003**). "Correlations between acoustic properties and bone density in bovine cancellous bone from 0.5 to 2 MHz," J. Acoust. Soc. Am. **113**, 2933–2938.

Lewandrowski, K.-U., Tomford, W. W., Michaud, N. A., Schomacker, K. T., and Deutsch, T. F. (**1997**). "An electron microscopic study on the process of acid demineralization of cortical bone," Calcif. Tissue Int. **61**, 294–297.

Lin, W., Qin, Y.-X., and Rubin, C. (**2001**). "Ultrasonic wave propagation in trabecular bone predicted by the stratified model," Ann. Biomed. Eng. **29**, 781–790.

Luo, G., Kaufman, J. J., Chiabrera, A., Bianco, B., Kinney, J. H., and Haupt, D., *et al.* (**1999**). "Computational methods for ultrasonic bone assessment," Ultrasound Med. Biol. **25**, 823–830.

McDonnell, P., McHugh, P. E., and O' Mahoney, D. (**2007**). "Vertebral osteoporosis and trabecular bone quality," Ann. Biomed. Eng. **35**, 170–189.

Nicholson, P. H. F., Haddaway, M. J., and Davie, M. W. J. (**1994**). "The dependence of ultrasonic properties on orientation in human vertebral bone," Phys. Med. Biol. **39**, 1013–1024.

O'Donell, M., and Miller, J. G. (**1981**). "Quantitative broadband ultrasonic backscatter: An approach to nondestructive evaluation in acoustically inhomogeneous materials," J. Appl. Phys. **52**, 1056–1065.

Padilla, F., Jenson, F., and Laugier, P. (**2006**). "Estimation of trabecular thickness using ultrasonic backscatter," Ultrason. Imaging **28**, 3–22.

Parfitt, A. M. (**1987**). "Trabecular bone architecture in the pathogeneses and prevention of fracture," Am. J. Med. **82**, 68–72.

Roberjot, V., Bridal, S. L., Laugier, P., and Berger, G. (**1996**). "Absolute backscatter coefficient over a wide range of frequencies in a tissue-mimicking phantom containing two populations of scatterers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 970–978.

Sehgal, C. M. (**1993**). "Quantitative relationship between tissue composition and scattering of ultrasound," J. Acoust. Soc. Am. **94**, 1944–1952.

Serpe, L., and Rho, J. Y. (**1996**). "The nonlinear transition period of broadband ultrasound attenuation as bone density varies," J. Biomech. **29**, 963–968.

Tavakoli, M. B., and Evans, J. A. (**1991**). "Dependence of the velocity and attenuation of ultrasound in bone on the mineral content," Phys. Med. Biol. **36**, 1529–1537.

Turner, C. H., Cowin, S. C., Rho, J. Y., Ashman, R. B, and Rice, J. C. (**1990**). "The fabric dependence of the orthotropic elastic constants of cancellous bone," J. Biomech. **23**, 549–561.

Wear, K. A., and Armstrong, D. W. (**2000**). "The relationship between ultrasound backscatter and bone mineral density in human calcaneous," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 777–780.

Wear, K. A., Stuber, A. P., and Reynolds, J. C. (**2000**). "Relationships of ultrasonic backscatter with ultrasonic attenuation, sound speed and bone mineral density in human calcaneous," Ultrasound Med. Biol. **26**, 1311–1316.

Wear, K. A. (**2003**). "The dependence of ultrasonic backscatter on trabecular thickness in human calcaneous: Theoretical and experimental results," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 979–986.

Williams, J. L. (**1992**). "Ultrasonic wave propagation in cancellous and cortical bone: Prediction of some experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112.

Wu, C., Gluer, C., Lu, V., Fuerst, T., Hans, D., and Genant, H. K. (**1998**). "Ultrasound characterization of bone demineralization," Calcif. Tissue Int. **62**, 133–139.

Xu, W., and Kaufman, J. J. (**1993**). "Diffraction correction methods for insertion ultrasound attenuation estimation," IEEE Trans. Biomed. Eng. **40**, 563–569.

An echolocation visualization and interface system for dolphin research

Mats Amundin
Kolmården Wild Animal Park, Kolmården, Sweden, and Biological Department, Institute of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden

Josefin Starkhammar,[a] Mikael Evander, Monica Almqvist, Kjell Lindström, and Hans W. Persson
Department of Electrical Measurements, LTH, Lund University, SE-221 00 Lund, Sweden

(Received 25 June 2007; revised 22 November 2007; accepted 4 December 2007)

The present study describes the development and testing of a tool for dolphin research. This tool was able to visualize the dolphin echolocation signals as well as function as an acoustically operated "touch screen." The system consisted of a matrix of hydrophones attached to a semitransparent screen, which was lowered in front of an underwater acrylic panel in a dolphin pool. When a dolphin aimed its sonar beam at the screen, the hydrophones measured the received sound pressure levels. These hydrophone signals were then transferred to a computer where they were translated into a video image that corresponds to the dynamic sound pressure variations in the sonar beam and the location of the beam axis. There was a continuous projection of the image back onto the hydrophone matrix screen, giving the dolphin an immediate visual feedback to its sonar output. The system offers a whole new experimental methodology in dolphin research and since it is software-based, many different kinds of scientific questions can be addressed. The results were promising and motivate further development of the system and studies of sonar and cognitive abilities of dolphins. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2828213]

PACS number(s): 43.80.Ka, 43.60.Qv, 43.80.Ev [WWA] Pages: 1188–1194

I. INTRODUCTION

Dolphins have gone through a long evolution which has resulted in a very advanced active sonar system based on the development of a unique sound generation and detection system (Amundin, 1991; Cranford *et al.* 1996; Au, 2004; Cranford and Amundin, 2004). Their sonar has been extensively studied over several decades, and much of their basic characteristics are already well known (Au, 1993). However, most of these studies have their foundation in an experimental setup where the dolphin has been trained to be voluntarily attached in a test rig, thus permitting its directional sonar beam to be recorded with fixed hydrophones. Although this setup allows for very exact measurements, it has most likely prevented the full dynamic potential of the dolphin's sonar to be revealed. Acoustic tag recordings carried out with free swimming dolphins have indicated that there obviously exists a very precise control of the frequency composition of individual clicks in a sonar click train (Sigurdson, 1997). However, this has not been fully reflected in the rigid setups. The system presented by Thomas *et al.* (2002) measured the sounds from free swimming captive dolphins with eight hydrophones distributed around walls of a 30 m by 45 m pool. The sounds were linked to the phonating dolphins which were also filmed with an overhead camera. Such a system allowed a spontaneous and social interaction between the dolphins while the sounds were recorded. However useful in its specific application, the system design did not enable real-time analysis, had a rather coarse spatial resolution and could not be used in studies of, for instance, dolphin target scanning behavior.

A dolphin's response to scientific questions, e.g., the in-target detection threshold or discrimination trials, has mostly constituted a "go/no go" response, pressing of a yes/no paddle (Au, 1993) or variations of match-to-sample studies. Such trials require trained dolphins, thus limiting the number of animals available for testing. Moreover, the go/no go or paddle responses are very rough indications of a choice, which is difficult to refine, and would thus be unpractical in a multiple-choice paradigm. In match-to-sample studies, multiple-choice paradigms are often required to be able to get clear and concise results. To make these test setups more practical, mostly in studies involving terrestrial animals, variations of symbol interfaces and touch screens have been developed.

In cognitive studies with primates, e.g., the chimpanzee, a computerized symbol interface, based on a finger operated touch screen, has been successfully used (Rumbaugh *et al.*, 1975). This approach has been employed even with birds, such as chickens and doves (Cheng and Spetch, 1995). So far, only two similar interactive tools with dolphins have been reported. Xitco *et al.*, 2001, reported the first use of an interactive keyboard for dolphins. It was designed to enable dolphins to activate symbols. The symbols were three-dimentional (3D) objects housed within circular tubes attached to one of four panels, or keyboards. The dolphins operated the keyboard by breaking an infrared light beam

[a]Author to whom correspondence should be addressed. Electronic mail: josefin.starkhammar@elmat.lth.se.

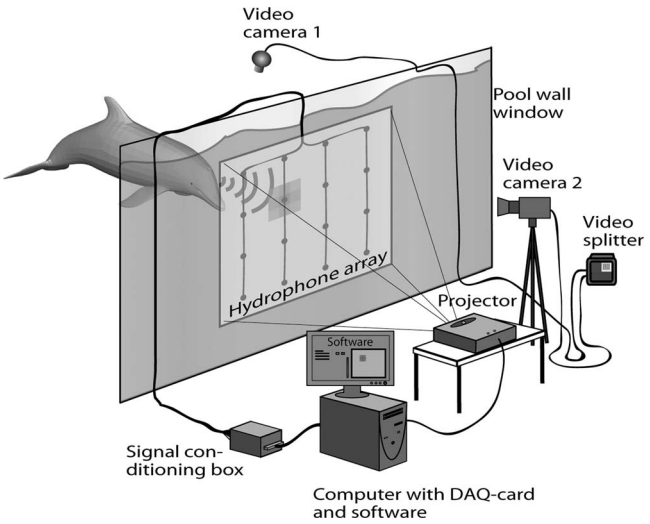


FIG. 1. The basic design of the EchoLocation Visualization and Interface System, ELVIS. The dolphin transmits a train of sonar pulses, focused in a narrow beam aimed at the hydrophone matrix. The relative sound pressure levels in the sonar beam are converted into light intensity variations on the PC screen. This dynamic image is projected back on the semitransparent hydrophone screen, offering an immediate visual feedback to the dolphin on its sonar output. The relative sound pressure levels can also be coded into, e.g., color variations only visible on the PC screen to the human observer.

projected in front of each symbol with its rostrum. When a symbol was activated the English word for the chosen symbol was played over an underwater speaker. The second tool within the interactive interface genre was reported by Delfour and Marten (2005). They investigated the ability of dolphins to associate sounds with visual stimuli. A custom-made touch screen, based on infrared light beams projected through an underwater viewing panel and guided by mirrors to create a grid in front of the panel, was used as the response interface. A TV screen placed in front of the panel displayed the visual test stimuli as well as various visual reward stimuli. The dolphin indicated its choice by breaking the light beams with its rostrum.

Although dolphins are known for using their rostrum to touch and manipulate objects, it was deemed of more interest to explore and study their main sensory system, i.e., their sonar. The present study thus describes the development of a new tool, called the EchoLocation Visualization and Interface System (ELVIS). An early version of ELVIS was presented by Nilsson (2003) and Nilsson *et al.* (2004). Since then the system has been further developed and tested in marine zoology applications. The aims of the investigation presented here are to demonstrate the basic principle of ELVIS and to test and evaluate the system in two applications: tracing the beam axis during a sonar target detection experiment and using it as an acoustically operated "touch screen" for dolphins.

II. MATERIALS AND METHODS

A. System configuration

The system consisted of a semitransparent screen with a matrix of 16 hydrophones, spaced 20 cm apart (Fig. 1). This screen was lowered into the dolphin pool in front of an underwater acrylic panel, and the system translated the variations of dolphin sound pressure levels in the sonar beam into variations in color and light intensity on the computer screen. Such an approach facilitates for the visually oriented human to observe the dynamics of the dolphin sonar. The computer screen image coded by light intensity was also projected back onto the hydrophone screen, by using a video projector, thereby offering an immediate visual feedback to the dolphins. ELVIS can be programmed to store all available data, such as the values of the measured sound pressure levels (SPLs), and the settings in the touch screen application. This allows for subsequent analyses and renders possible playback at any chosen speed. Two video cameras were used to document the measurement results and to help correlate acoustic and physical behavior (Ball and Buck, 2005). One video camera (video camera 1 in Fig. 1) was mounted directly over the location of the ELVIS screen so that the orientation of the dolphin in relation to the screen could be easily viewed. Also video recordings of the events on the screen were made (video camera 2) during both the testing of the beam axis tracing program and during the training of the dolphins to operate the acoustic touch screen. The obtained video sequences were then synchronized, using a video splitter and stored with a DVD recorder.

The interface was implemented as custom-made software, constituting the core of the interactive features of ELVIS. The maximum SPL in the sonar beam, i.e., the beam axis, was indicated by a round colored spot on the PC screen. The size of the spot was fixed and not related to the actual beam width in these applications. The position where the beam axis hit the screen was derived through interpolation of the signals from the hydrophones in the matrix, and even though the number of hydrophones was rather small, the spatial resolution of the tracing of the beam axis was sufficient for the present applications.

B. Hardware

The maximum sound pressure levels of the sonar pulses were measured by connecting a peak detector to each hydrophone. The block diagram in Fig. 2 shows the signal path through the system.

The hemispherical hydrophones (UD1, Ceram AB, Lund, Sweden) had peak sensitivities around 150 kHz, but were able to detect signals in the whole range where the dolphin echolocation signals have their peak energy, i.e., 40–120 kHz (Au, 1993). The hydrophone sensitivity at 120 kHz corresponds to −220 dB re 1 V/μPa.

A gain of 50 dB was used to amplify the received sonar signal levels up to a couple of volts, which is the optimal level for the DAQ-card (Data acquisition) (12 bit Adlink NuDaq PCI-9112, Adlink Technology Inc, Taipei, Taiwan). In order to keep a maximum amount of signal bandwidth, the gain was separated into two steps, i.e., first 20 dB and then 30 dB. After amplification, the signal was filtered with a 160 kHz first order low-pass filter followed by a 40 −130 kHz second order Butterworth band-pass filter.

The monolithic analog peak detectors with reset-and-hold mode (PKD01, Analog Devices, Norwood, MA, USA)

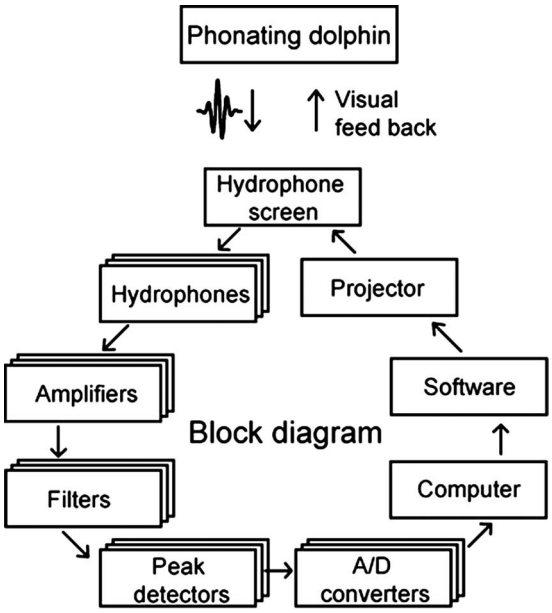


FIG. 2. Block diagram displaying the signal path through the system. The five blocks, labeled hydrophones, amplifiers, filters, peak detectors and A/D converters, represent 16 independent channels.

monitored the click amplitudes as detected by each hydrophone and retained the highest value until they were read and reset by a control signal from the DAQ card.

The small gray circles in Fig. 3 (marked by three arrows) on the peak detector output signal indicate the acquired amplitude readings that are stored by the computer software. According to the manufacturer, the sampling rate of the DAQ card is restricted by the hardware to a maximum of 110 kS/s. The Adlink DAQ card has 16 analog input channels and two analog output channels. Multiplexing the channels (i.e., acquiring data from one channel at a time) with an auto scanning function simplifies the design of the software. One of the analog output channels was used for resetting the peak detectors after reading.

The actual sampling rate was set by the software, and it turned out to be necessary to limit it to a total maximum rate of 70 kS/s, so as to eliminate the risk of overloading the hardware. This provided a sampling rate of 4.4 kS/s for each transducer, which was enough to catch every click even at very high click repetition rates. The normal click rate in a close distance sonar "buzz" is <500 clicks/s (Au, 1993). In Fig. 3 the theoretical maximum dolphin click rate of approximately 1000 clicks/s is used to illustrate the resulting sampling of the click amplitude.

C. Software

A software program was developed to control all the core functions of the system as well as the data acquisition via the DAQ card. Its function was to configure the card and sequentially read data from each of the peak detectors. The peak detectors were reset after each reading before a new series of data could be acquired.

Two main application programs were designed in order to render the system interactive to the dolphins, and both were extensions of the basic program. The first program (referred to as the "Beam axis tracing program") could be set to show either a light intensity graph on the screen as in Fig. 1 or a color light intensity spot at the location of the sonar beam sound pressure level maxima on the screen. It was possible to set the persistence time for the color spot, causing it to stay longer on the screen, thus also making it possible for the dolphins to "paint" on the screen by moving the beam axis over the hydrophone matrix. This also facilitated for the human eye to follow the trace of the sonar beam axis. The size of the spot was fixed, but its light intensity depended linearly on the sound pressure level of the click. A beam axis replay function, based on the amplitude measurements stored in a log file on the computer hard disk, was added to the program to facilitate the analysis of the dolphins' sonar behavior. From these log files the beam axis trace could be reconstructed on the computer screen in fast as well as slow motion.

The second program converted the system to a truly interactive, acoustically operated touch screen. This program (referred to as the "Acoustic touch screen" program) dis-

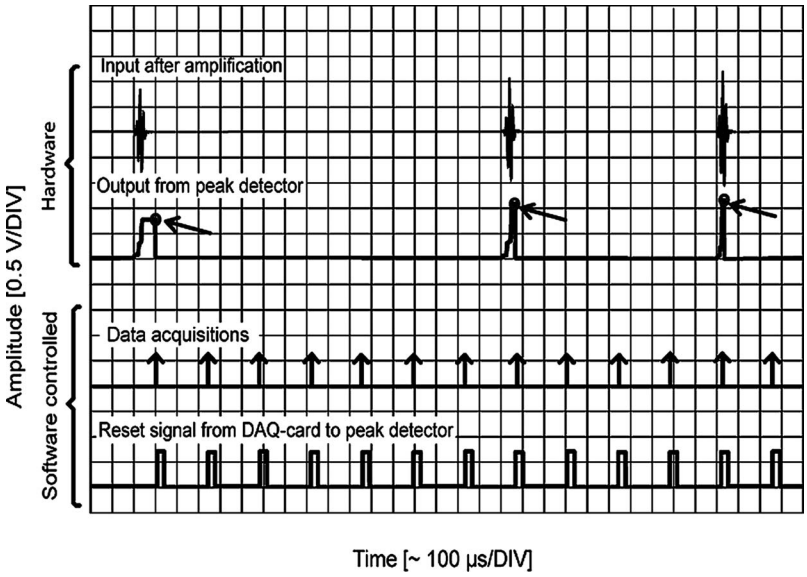


FIG. 3. The peak-hold detector output as it monitors the maximum amplitude of the input signal (dolphin click) from a hydrophone. After software-controlled storage to a hard drive, the detector is reset by a control signal from an output on the acquisition board.

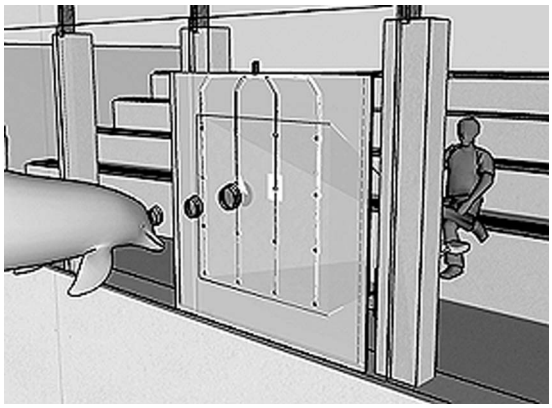


FIG. 4. The ELVIS, as seen from the dolphin's point of view, while being used as an acoustic "touch screen." The dolphin can "click" on the two white symbols by aiming its sonar beam axis at them. The symbols correspond to active buttons on a normal touch screen. The active area as well as the trig level can be set by the operator to adjust accordingly to the training level of the dolphin.

played a variety of white symbols at either random or fixed locations on the screen. These symbols functioned as active buttons to be "clicked" on by the dolphin. Figure 4 shows a dolphin operating the acoustic touch screen by pointing its sonar beam axis at one of the projected symbols.

Depending on the application, the buttons can be associated with, for instance, various actions of the program and/or a number of objects that can be chosen by the dolphin. In one study, the symbols represented different species of reward fish. A symbol was activated when the sound pressure level in the beam axis of a detected sonar click reached above a preset trigger level. This led to the symbol flashing shortly and a reward whistle (the bridging stimulus) to be played to the dolphin through speakers connected to the computer. The dolphin then returned to the trainer at the pool side to receive a fish of the species represented by the chosen symbol.

It was also possible to set a time criteria in the program, forcing the dolphin to echolocate on the symbol a fixed number of seconds before the trigger level was accepted. However, in the initial study, only one click above the trig level was required to indicate a choice. The purpose of this first application was to test the acoustic touch screen and to introduce the dolphins to this concept. Initial results of this study are presented in Starkhammar *et al.* (2007) and Olsén (2007).

To help evaluate the functionality of the touch screen, a tracing function of the beam axis was built into the program of the acoustic touch screen. This rendered it possible to continuously monitor the beam axis as the dolphin selected a symbol to click on, and consequently to study how the dolphins, through trial and error, learned how to handle the program. The beam axis spot was displayed in dark red to make it inconspicuous to the dolphins (Madsen and Herman, 1980) and thus not distract them while operating the touch screen. Moreover, a click detector (ECD-1, NewLeap Ltd, UK) with a speaker was connected to one of the hydrophones in the center of the screen, making it possible to hear the echolocation during the trials. This was done to provide additional

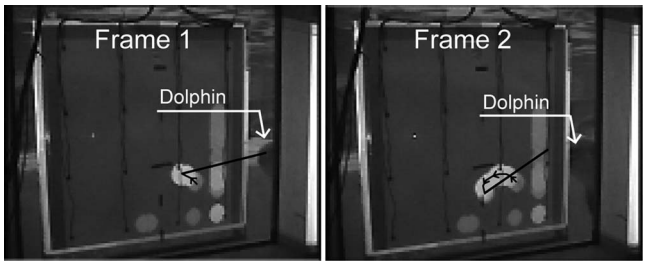


FIG. 5. These video frames span over a total time of 0.2 s. The picture shows the beam axis trace on the screen and a dolphin echolocating as it swims up to the right side of the screen. The video recordings were made with camera 2. The black line was added to indicate the beam axis orientation. The small black arrows illustrate the motion path of the beam axis across the screen.

feedback to the operator. Seen from the pool side of the screen the click detector was connected to the second hydrophone from the left in the second row from the top.

Several different settings were possible in the acoustic touch screen program. The active area around the symbols could be increased so as to comply with the level of training of a specific dolphin. For instance, an inexperienced dolphin could be aided by being provided with a larger active area around the symbol, hence making it easier for it to "hit" it with its sonar beam axis.

III. RESULTS AND DISCUSSION

The beam axis tracing program was found to perform according to the objective of the design and visualized the location and sound pressure levels of the beam axis as intended. When the screen was first presented to the dolphins, showing all sound pressure level variations within their sonar beam, they spontaneously and deliberately explored it with their echolocation. Their reactions were interpreted as them being intrigued and stimulated by the visual feedback of their sonar beams. With the inter-hydrophone distance set to 20 cm, the system was found to operate best when the dolphins were at a minimum of 1 m distance from the screen. If the dolphins were too close, the entire sonar beam might fall in between the hydrophones and hence not be registered at all. Such a scenario resulted in gaps in the tracing of the beam axis. Furthermore, if only one hydrophone was hit by the sonar beam, the interpolation function might result in a positioning error of a maximum of half an inter-hydrophone distance. An array of 64 elements, i.e., an inter-hydrophone distance of 10 cm, should make the interpolation operate properly for a dolphin at a distance as short as 0.5 m from the screen.

The software was developed under the assumption that only one dolphin at a time would echolocate towards the screen. Thus, if two dolphins were aiming their sonar beam at the screen simultaneously, the resulting maximum point, i.e., the beam axis, might appear on an incorrect location on the screen. However, the relatively fast scanning rate of the hydrophones made this a minor problem and even with several dolphins spontaneously echolocating at the screen at the same time, the system was able to monitor the beam axis. Figures 5 and 6 were created from a video sequence filmed

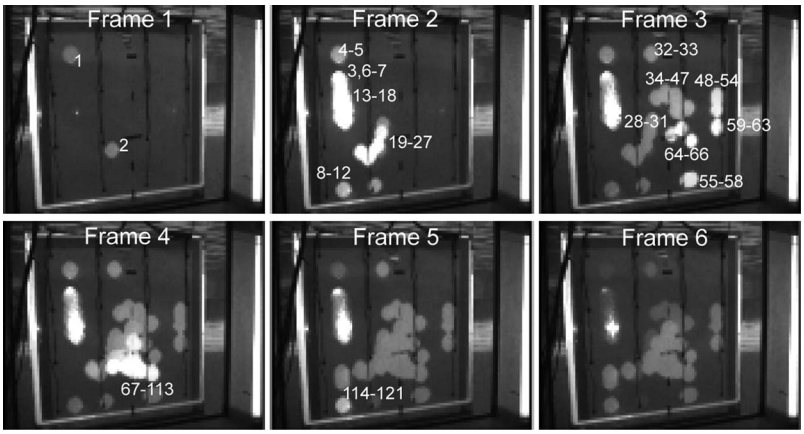


FIG. 6. These six frames show the beam axis tracing program while one dolphin is exploring the system for the first time. The frames are captured from a sequence filmed with video camera 2. The time between each frame is 1 s. All spots drawn by the system were numbered according to the order of which they appeared on the screen. The text and numbers in the pictures were added after the filming.

with camera 2 during the first time the dolphins encountered the full version of the beam axis tracing software. Unfortunately camera 1 was not used at the time of this initial testing of the newly developed software. The whole group of nine dolphins was allowed to swim freely in the pool and investigate the screen. In Fig. 5 an echolocating dolphin swims up to the screen from the lower right corner. The beam axis is visualized to both the dolphin and the human observer. The vertical trace at the very right side of the screen came from echolocation clicks a few seconds earlier, possibly from another dolphin. To help the readers to perceive the orientation of the dolphin in the pictures, a black line representing the beam axis was added in both frames. This shows that the beam axis tracing program visualizes the beam core location with acceptable accuracy. Small black arrows indicate the movement of the beam axis as it hits the screen.

During the screen shot sequence in Fig. 6, one dolphin (not visible to camera 2) explores the screen by echolocating at it. The time between each frame is 1 s. The dots drawn by the system, indicating the trace of the beam axis, are numbered according to the order of which they appeared on the screen. Note that the light intensity of the trace represents the maximum peak sound pressure levels of the measured click. If there are multiple clicks at the same location the light intensity sums up until it reaches a maximum level (white). The fade-out time was set to 6 s. A possible way of making the time for each spot more intuitively clear to the operator would be to let the diameter of the dots decrease over time, instead of letting the dots' intensity fade out and then number them afterwards.

The beam axis tracing program has been employed in a study of benthic foraging in bottlenose dolphins, carried out at Dolphin Encounters, in the Bahamas (Dahl, 2007). Here, the hydrophone matrix was buried under coral sand and the sonar beam axis was traced while trained dolphins searched the sea floor for buried sonar targets. In this application the dolphins received no visual feedback; this was restricted to the human observer operating the computer. The replay function proved to be a very useful tool in the analysis process, and Fig. 7 shows a screen dump from such a replay sequence.

ELVIS fulfilled all our expectations as a visualization tool for tracing the dolphin sonar beam axis. It also proved that the acoustic touch screen concept can be operated by the dolphins and thus function as a computerized interface between man and dolphin. The acoustic touch screen software rendered it possible for the dolphins to activate the buttons (the projected symbols) by pointing their sonar beam axis at them.

Figure 8(A) shows the position of the dolphin relative to the acoustic touch screen while operating it. The orientation and position of the screen is marked in white. Figure 8(B) displays the white projected symbols and the tracing of the dolphin's beam axis. In order to make the shapes in B clearer to the reader, the boundaries of the symbols have been enhanced with a white line. The round dots represent the beam axis of individual clicks and are numbered in the order they appeared on the screen. All edits were made in the analysis process after the session. The time difference between echolocation click No. 1 and No. 4 in the figure was less than 0.4 s.

Sixteen dolphin training sessions, each lasting approx. 10 min, were carried out in which three dolphins were subjected to the task of "clicking" on the buttons of the screen with their sonar beams (Olsén, 2007). All three dolphins

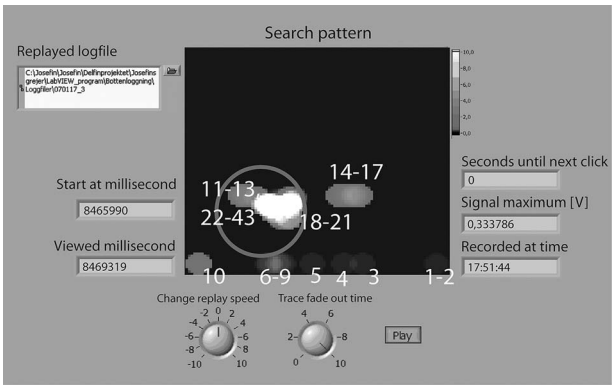


FIG. 7. A screen shot from running the replay program interface with a short sequence of a sonar beam axis trace of a dolphin searching for an object partially buried in coral sand over the hydrophone matrix. The object's position within the "Search pattern" area is indicated by the painted gray circle. The light intensity of the spots indicates the sound pressure levels of the clicks. It spans from dark red for weak clicks to white for strong clicks in a linear scale. The fade-out time of the trace can be set during replay. The intensity of each click hitting the same location was summed in this particular visualization mode. This figure has been converted into grayscale, thus representing dark red as dark gray. Consecutive clicks are numbered from 1 to 43 and were recorded during 3.5 s.

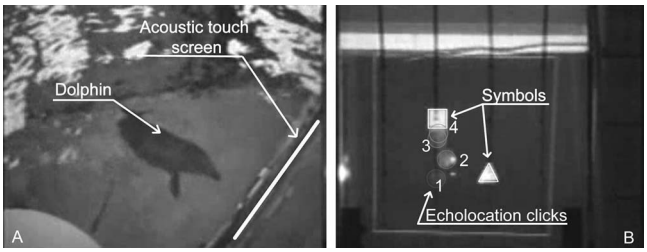


FIG. 8. Time synchronized video frames of a dolphin operating the acoustic touch screen, where A shows the dolphin from above, and B shows the acoustic touch screen from the dry side of the pool, with the trace of the dolphin's sonar beam axis as it hits the square symbol.

were previously trained to echolocate on a 3D sonar target lowered into the pool. During the very first training session all three dolphins learned the task of echolocating on a white square, in each run appearing at the same position on the screen. This was accomplished by holding the sonar target in front of the projected symbol for three runs. Then the target was removed. During the following 13 sessions, the dolphins learned to deliberately "click" on a symbol when up to three symbols were simultaneously projected onto the screen on random locations. Only one dolphin at a time was sent to the screen to perform the task. The dolphins were rewarded for clicking on any of the symbols on the screen. Each dolphin learned the task in its individual pace. The youngest subject (born 2001) learned the task quickest and clicked on a symbol with sufficient accuracy and sound pressure level within 10 s in 100% of the trials after the first session with the sonar target. The oldest subject (born 1973) seemed to be the slowest learner and clicked on a symbol within 10 s in 75% of all trials. All three dolphins required a mean time of less than 4 s to perform the task correctly, calculated using data from all trials during the 13 sessions.

Since sound and sonar dominate the lives of dolphins, it was assumed that they might intuitively understand and adopt the acoustic touch screen without extensive training. This assumption was supported by the obtained results. With the present system, a whole range of unexplored research areas regarding dolphin cognition and sonar skills can be studied. One cognitive study can be to investigate the ability to understand and interpret contents in pictures and videos. Regarding dolphin sonar skills the hypothesized internal beam forming ability in dolphins can be further explored with this system. The promising results of the tests highly motivate a further development of the system in order to render it even more potent. In addition to improving the resolution by increasing the number of hydrophones, the system should be expanded from merely measuring the peak sound pressure levels of the clicks to high frequency digital sampling of the clicks in each hydrophone position. Such a development would allow for the full frequency bandwidth of the sonar clicks and its dynamic frequency variations to be studied on-line. It would push the limits of fast data streaming to the hard drives but should nevertheless be realistic considering the fast development of DAQ cards and PC buses.

The flexibility of the software-based system makes it possible to match each application in an optimal way to the task. If one application requires only, e.g., the variations of peak frequency and not sound pressure levels to be visualized, this can easily be implemented to be set on line by the operator. In some applications, only some parameters would be presented to the dolphins and others only available to the operator. If more than one property at a time are to be shown (peak frequency, click interval, sound pressure level, beam width, etc.), the operator must have the possibility to control the presentation of these parameters. As an example, the change in peak frequency and sound pressure levels can be visualized by a variation of shape and light intensity, respectively. Click interval can, e.g., be visualized by a change in fade out time of the projected color spot, e.g., a long click interval giving a long fade out time and a short click interval a fast fade out time. This would make it easier to see each new click even in a fast click train. It should be emphasized that these are only suggestions to how the system can be set up to match a number of future applications.

IV. CONCLUSIONS

The preliminary tests of ELVIS described herein displayed that the system functioned as intended both as a tool for visualizing the sound pressure level dynamics in the dolphin sonar beam and as an interactive acoustic "touch screen" operated by the dolphins' sonar. The promising results motivate additional studies of dolphin behavior, echolocation and cognition, as well as a further development of the system. As a result of the system being software based, numerous types of ethological and cognitive studies (e.g., pictorial competence) can be implemented.

ACKNOWLEDGMENTS

Many thanks are expressed to Caroline Persson and her dolphin trainer colleagues at the Kolmården Dolphinarium for training the dolphins and always finding time for the experimental trials, often at awkward hours. We also thank Dolphin Encounters, Bahamas, for allowing us to conduct experiments at their excellent facilities, and Kathleen Dudzinski for making this expedition possible. Annette Dempsey at Dolphin Encounters directed the training of the dolphins during the trials, and she and the trainers are acknowledged for their training skills and constant good mood. We thank Lennart Nilsson, Lund University, for the excellent mechanical construction of the screen. Josefin Starkhammar is funded by Dean Gunilla Jönson, the Faculty of Engineering, Lund University, Lund, Sweden, and research grants have been awarded by the Crafoord Foundation, the Carl Trygger Foundation and the ELFA research foundation.

Amundin, M. (**1991**). "Sound production in Odontocetes, with emphasis on the harbor porpoise, *Phocoena phocoena*," Doctoral dissertation, Stockholm University.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L. (**2004**). "Dolphin sonar detection and discrimination capabilities (A)," J. Acoust. Soc. Am. **115**, 2614.

Ball, K. R., and Buck, J. R. (**2005**). "A beamforming video recorder for integrated observations of dolphin behavior and vocalizations (L)," J. Acoust. Soc. Am. **117**, 1005–1008.

Cheng, K., and Spetch, M. L. (**1995**). "Stimulus control in the use of land-

marks by pigeons in a touch-screen task," J. Exp. Anal Behav. **63**, 187–201.
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.
Cranford, T. W., and Amundin, M. (**2004**). *Advances in the Study of Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago).
Dahl, S. (**2007**). "Target detection in coral sand by bottlenose dolphins (Tursiops truncates)." M.Sc. thesis, Biology Department, Institute of Physics, Chemistry and Biology, Linköping University, Sweden.
Delfour, F., and Marten, K. (**2005**). "Inter-modal learning task in bottlenosed dolphins (Tursiops truncatus): A preliminary study showed that social factors might influence learning strategies," Acta Ethol. **8**, 57–64.
Madsen, C. J., and Herman, L. M. (**1980**). *Cetacean Behaviour: Mechanisms and Functions*, edited by L. M. Herman (Wiley, New York), pp. 101–147.
Nilsson, M. (**2003**). "Echolocation visualization interface system," M. Sc. thesis, Dept. of Electrical Measurements, Faculty of Engineering LTH, Lund University, Lund, Sweden.
Nilsson, M., Lindström, K., Amundin, M., and Persson, H. W. (**2004**). "Echolocation and visualization interface system," Clin. Physiol. Funct. Imaging **24**, 171.
Olsén, H. (**2007**). "First application of ELVIS—an echo-location visualization and interface system," M.Sc. thesis, Biology Department, Institute of Physics, Chemistry and Biology, Linköping University, Sweden.
Rumbaugh, D. M., Gill, T. V., von Glasersfeld, E., Warner, H., and Pisani, P. (**1975**). "Conversations with a chimpanzee in a computer-controlled environment," Biol. Psychiatry **10**, 627–641.
Sigurdson, J. E. (**1997**). "Analyzing the dynamics of dolphin biosonar behavior during search and detection tasks," In: *Proceedings of the Institute of Acoustics' conference "Underwater bio-sonar and bioacoustics*," Loughborough University, Loughborough, UK, Vol. **19**, pp. 123–132.
Starkhammar, J., Amundin, M., Olsén, H., Almqvist, M., Lindstrom, K., and Persson, H. W. (**2007**). "Acoustic touch screen for dolphins, first application of ELVIS—an Echo-location visualization and interface system," in *Proceedings of the Institute of Acoustics*, Loughborough University, Loughborough, UK, Vol. **29**, pp. 55–60.
Thomas, R., Fristrup, K. M., and Tyack, P. L. (**2002**). "Linking the sounds of dolphins to their locations and behavior using video and multichannel acoustic recordings," J. Acoust. Soc. Am. **112**, 1692–1701.
Xitco, M. J., Gory, J. D., and Kuczaj II, A. A. (**2001**). "Spontaneous pointing by bottlenose dolphins (*Tursiops truncatus*)," Anim. Cogn. **4**, 115–123.

Extended three-dimensional impedance map methods for identifying ultrasonic scattering sites

Jonathan Mamou,[a] Michael L. Oelze, and William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

James F. Zachary
Department of Pathobiology, University of Illinois, 2001 South Lincoln, Urbana, Illinois 61802

(Received 26 June 2007; accepted 16 November 2007)

The frequency-dependent ultrasound backscatter from tissues contains information about the microstructure that can be quantified. In many cases, the anatomic microstructure details responsible for ultrasonic scattering remain unidentified. However, their identification would lead to potentially improved methodologies for characterizing tissue and diagnosing disease from ultrasonic backscatter measurements. Recently, three-dimensional (3D) acoustic models of tissue microstructure, termed 3D impedance maps (3DZMs), were introduced to help to identify scattering sources [J. Mamou, M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, "Identifying ultrasonic scattering sites from 3D impedance maps," J. Acoust. Soc. Am. **117**, 413–423 (2005)]. In the current study, new 3DZM methodologies are used to model and identify scattering structures. New processing procedures (e.g., registration, interpolations) are presented that allow more accurate 3DZMs to be constructed from histology. New strategies are proposed to construct scattering models [i.e., form factor (FF)] from 3DZMs. These new methods are tested on simulated 3DZMs, and then used to evaluate 3DZMs from three different rodent tumor models. Simulation results demonstrate the ability of the extended strategies to accurately predict FFs and estimate scatterer properties. Using the 3DZM methods, distinct FFs and scatterer properties were obtained for each tumor examined. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2822658]

PACS number(s): 43.80.Qf, 43.80.Ev, 43.20.Fn, 43.80.Vj [FD] Pages: 1195–1208

I. INTRODUCTION

There has been a clear need for developing methodologies to better understand how ultrasonic waves are scattered by tissue microarchitecture. Developing a strategy to identify and quantify scattering sites could lead to significant advances in ultrasonic imaging because it would connect ultrasound-derived parameters with histopathologic properties of tissues. Consequently, medical diagnoses could possibly be drawn without the need for invasive biopsies.

The goal of quantitative ultrasound (QUS) is to identify, quantify and display tissue features that are related to histopathologic properties. Many QUS techniques rely on the frequency-dependent information contained in the radiofrequency backscattered signals to estimate tissue properties.[1,2] From the frequency-dependent information, QUS methods quantify tissue properties based on assumptions about the nature of scattering structures. Because in many tissues the actual structures responsible for scattering remain unknown, there is no scientific basis for validating current hypotheses or choosing a model. However, the limitations in understanding ultrasonic scattering sites in tissue have not deterred QUS progress. Our group conducted several studies using a Gaussian scattering model because it led to efficient estimation schemes[3–5] and also developed a new scattering model that took into account the cell's internal structure.[6] Other groups have successfully demonstrated that QUS parameters of extravascular matrix patterns (EMPs) correlated with histologic EMP, thus allowing for the discrimination between lethal and less lethal primary malignant melanoma of the choriod and ciliary body.[7]

For more than two decades, QUS methods have been developed to ultrasonically quantify ocular, liver, prostate, renal, and cardiac tissues.[8–11] As this work progressed, the need for a more fundamental understanding of ultrasonic scattering became evident. Thus, QUS studies were conducted on simpler biological media wherein cells in pellets or in suspension were characterized.[12,13] These studies had some success at quantifying cell apoptosis and more generally at suggesting which structures might be responsible for scattering in specific cell types.

Recently, we proposed an original method to predict ultrasonic scattering from complex tissue structures.[14–17] Based on an aligned set of adjacent and stained histologic sections, a three-dimensional (3D) acoustic computational model, termed the three-dimensional impedance map (3DZM), was constructed and utilized to obtain quantitative information about tissue scattering. A 3DZM is a 3D matrix whose elements are acoustic impedance values corresponding spatially to tissue structures. These preliminary studies demonstrated the feasibility of the 3DZM method, and also assisted in the identification of advanced methodologies to further investi-

[a]Author to whom correspondence should be addressed. Current address: F. L. Lizzi Center for Biomedical Engineering, Riverside Research Institute, 156 William St., 9th floor, New York, NY 10038. Electronic mail: mamou@rrinyc.org

gate and fundamentally understand ultrasonic scattering. The aim of this paper is to introduce and evaluate these advanced 3DZM methodologies.

The paper is organized as follows. Section II briefly reviews the theoretical background of 3DZMs and discusses their effectiveness as tools for understanding tissue scattering by ultrasound. The processing strategies to derive 3DZMs in an automatic fashion from the adjacent set of two-dimensional (2D) histologic sections were improved and are presented in Sec. III. Section IV presents advanced methodologies based on 3DZMs to help in identifying the anatomic structures responsible for ultrasonic tissue scattering. Section V presents results from simulated and experimentally derived 3DZMs. Finally, Sec. VI proposes possible extensions of the methodologies.

II. THEORETICAL MOTIVATION

A. Weak scattering in an inhomogeneous medium

The theory of scattering of a propagating acoustic wave in a heterogeneous medium is briefly reviewed (see Refs. 15, 18, and 19 for more details). In the theory, weak scattering is defined as the case where the inhomogeneities that cause scattering have tissue property values (density, ρ, and compressibility, κ) very close to those of the rest of the medium. For plane wave incidence, the backscattered acoustic pressure from a weakly scattering medium is[18]

$$p_{bs} = \frac{e^{-ikr}}{r}\Phi(2k), \tag{1}$$

where bs denotes backscattered, k is the propagation constant ($k=\omega/c$ where ω is the angular frequency and c is the propagation speed) and the angle distribution function, which describes the frequency dependence of the backscatter, is

$$\Phi(2k) = \frac{k^2}{4\pi}\int\int\int_{V_0}\frac{\Delta z(r_0)}{z(r_0)}e^{-2ikr_0}d\upsilon_0, \tag{2}$$

where $\Delta z(r_0)/z(r_0)$ is the relative change in acoustic impedance at spatial location r_0.

From Eq. (2), the final expression for the backscattered intensity is

$$I_{bs} = Ak^4S(2k), \tag{3}$$

where

$$S(2k) = \frac{S'(2k)}{S'(0)}, \quad \text{and} \tag{4}$$

$$S'(2k) = \left|\int\int\int_{V_0}\frac{\Delta z(r_0)}{z(r_0)}e^{-2ikr_0}d\upsilon_0\right|^2 \tag{5}$$

and where A is a proportionality constant. $S'(2k)$ and $S(2k)$ are the power spectrum and normalized power spectrum of the medium, respectively. (Details about this derivation can be found in Ref. 15.)

B. Form factor

Form factors (FFs) are functions of frequency that describe the amplitude of the backscattered intensity from an inhomogeneous medium. FFs are defined as the ratio of the medium's backscatter cross section to that of the same medium but under the Rayleigh scattering assumption.[1] Following that definition, FFs can be interpreted as the ultrasonic signature of a stochastic medium under ultrasonic interrogation.

Theoretical FFs can be deduced from 3D spatial correlation models by assuming some form or shape for the scattering tissue structures.[1,19] Usually simple scattering shapes are assumed and in most cases they have a spherical symmetry. Specifically, the FF is the Fourier transform of the 3D spatial autocorrelation function of a 3D medium containing a single scattering structure, that is, the magnitude squared of the Fourier transform of the single scatterer's shape. The normalization to Rayleigh scattering removes the k^4 dependence from the backscattered intensity [see Eq. (3)] and consequently FFs are normalized to a value of 1 when $k=0$ and their derivative usually vanishes when $k=0$.[19] Therefore, FFs are readily comparable to the normalized power spectrum $S(2k)$ [Eq. (4)].

C. Relationships between scattering, FF and 3DZM

The previous sections have introduced the theoretical background needed to use the 3DZMs as a powerful tool to understand ultrasonic scattering. In what follows, it is assumed that an accurate 3DZM of a specific tissue type exists. This assumption means that each impedance value at each 3DZM voxel is the correct impedance value for the underlying tissue type.

Based on Eqs. (1)–(5), the backscattered intensity resulting from a plane wave incident on a scattering volume is

$$I_{bs}^{3DZM} = Bk^4FT[SAF(3DZM)] = Bk^4|FT[3DZM]|^2, \tag{6}$$

where the superscript denotes the 3DZM-deduced intensity (as opposed to that ultrasonically measured), $FT[\bullet]$ denotes the Fourier transform operator, B is a proportionality constant and SAF(3DZM) is the spatial autocorrelation function of the 3DZM. The second equality is obtained using the Wiener–Khintchine theorem. Equation (6) reveals that the 3DZM can be used as an independent estimator of the backscattered intensity at any frequency. This frequency-independent property of the 3DZM will be exploited in Sec. IV A.

The FF of the inhomogeneous medium can be deduced by

$$FF(ka)^{3DZM} = \frac{FT[SAF(3DZM)]}{\lim_{k\to 0}FT[SAF(3DZM)]}, \tag{7}$$

where a is the mean scatterer radius and is often a parameter to be estimated. Equation (7) provides a unique means of estimating the FF for a given medium from its 3DZM. This is a powerful tool to identify ultrasonic structures responsible for scattering because of the intimate relationship that exists between scatterers and FFs. (Section IV B presents an algorithm that exploits Eq. (7) to estimate the FF of a medium

from its 3DZM.)

III. ADVANCED 3DZM CONSTRUCTION METHODS

A 3DZM is an acoustic computational phantom, i.e., a 3D matrix whose elements are the values of the acoustic impedance of the medium. The 3DZMs are derived from a 3D histologic dataset. The tissue is fixed in 10% neutral-buffered formalin, embedded in paraffin, sectioned, mounted on glass slides, and stained with H&E (Hematoxylin and Eosin). Hematoxylin stains nucleic acids (chromatin in nuclei and ribosomes) blue; the greater the nucleic acid density, the darker the blue color. Eosin stains proteins such as cell cytoplasm, connective tissue, muscle, etc. pink; the greater the protein concentration, the darker the pink color. Each stained section is photographed with a light microscope [Nikon (Nikon Corporation, Tokyo, Japan) Optiphot-2 optical microscope], and the photographs are digitized with a Sony (Sony Corporation, Tokyo, Japan) charge coupled device, Iris/RGB color video camera as a bitmap image.

The 3D reconstruction of a 3DZM from two-dimensional (2D) histologic sections involves several processing steps. Since initially presenting the 3DZM strategy,[15] three significant improvements of the processing strategies to construct 3DZMs from adjacent photomicrographs of histologic sections have occurred. These improvements include contrast equalization, fine-tune registration, and interpolation of missing sections. First, the contrast of the thin sections is equalized because the uptake of stain, in general, is not uniform from section to section; nor is the video capture intensity necessarily the same from section to section. Second, a fine-tuning registration is conducted. Third, after registration, the sections lost during sectioning are interpolated. Finally, each voxel in the 3D histologic volume is assigned an acoustic impedance value. The resulting 3D matrix is the final 3DZM.

A. Contrast adjustment

Image contrast of each H&E stained bitmap image is equalized prior to registration in order to increase the robustness of the registration algorithm. Contrast equalization leads to more similar images increasing the likelihood for correct convergence of the registration algorithm.

Equalization is conducted for each of the three color components (i.e., red, green, and blue) of each image. Specifically, let H_c^I denote the cumulative histogram of image I (for one of the color components). Thus, $H_c^I(\alpha)$ is the number of pixels with intensity level $\leq \alpha$ in the image (α is assumed to be an integer in the range 0–255 because of the 8 bit precision for each color component). If $I(i,j)$ represents the intensity level of pixel (i,j) in image I, then the new pixel intensity in the equalized image I' at (i,j) is given by[20]

$$I'(i,j) = 255\frac{H_c^I(I(i,j))}{N}, \tag{8}$$

where N is the total number of pixels in the image and the factor 255 has been added so that the new pixel intensity is in the range 0–255 (because $0 \leq H_c^I/N \leq 1$).

B. Registration

Registration or alignment is an essential step in the 3D reconstruction of the tissue model from serially sectioned photomicrographs. If the consecutive sections are not aligned correctly, then the resulting 3D model yields misleading or incorrect results that lead to potentially serious artifacts. The goal is to determine the best transformation to apply to an image I_2 so that it is the most similar to the original and adjacent image I_1. Hence, the registration algorithm is composed of three parts: a set of transformations, a similarity measure and an optimization process.

1. Set of transformations

The goal of this part of the registration algorithm is to transform I_2 using a transformation T such that $T(I_2)$ is as similar as possible to I_1 (I_1 and I_2 are adjacent histologic section images). Several types of transformations T can be used. The set of transformations, S, usually falls into two main categories: rigid and nonrigid transformations.[21,22] The rigid transformations set, S_r, consists of three degrees of freedom (DOFs) (two translational and one rotational). The nonrigid transformations set, S_{nr}, is composed of just about any conceivable adjustable parameter.

In the present work, S_r is used because it is a reasonable choice considering the error sources in the initial manual alignment of the images. A subset of S_{nr} limited to the affine transformations (S_{af}) is used because this class of transforms can compensate for the small (likely in the 2–5% range), but unknown and inherent tissue shrinkage that occurs during tissue preparation. The shrinkage could affect the predicted form factor by preferentially shrinking some components more than others. However, because water comprises a majority of the molecules in the cells in these tumors and all of the tumors were processed under identical conditions, it is reasonable to assume that shrinkage, if it occurred, was evenly distributed throughout the entire tissue including tumor cells and other extracellular matrix components. At this stage of 3DZM development, even if a 2–5% shrinkage occurred, it would not significantly affect the quantitative measures; however, this is an area that will need to be investigated further.

The affine transform has six DOFs that consist of a rotation, a scaling along both axes, a shearing parameter, and a 2D translation vector.[23] Basically, an affine transformation is the composition of a translation with any invertible linear transformation. Therefore, S_{af} is the set of all the transforms that conserves parallelism. It is also important to notice that the set of rigid transforms is strictly included into affine transforms.

When the pathologist performs the initial manual registration, only a rigid registration is performed. Therefore, when rigid registration is considered herein, it is a fine-tuning registration. Affine transforms include DOFs (stretching and shearing) that are not available to the pathologist when aligning adjacent sections.

2. Similarity measure

Once the image I_2 is transformed, the similarity of $T(I_2)$ to I_1 must be quantified. Define $G(I,J)$ as a function that quantifies the similarity between image I and image J. Many choices are possible for G,[24] but the functions G based on information theory concepts are usually very robust.[25,26] A robust response is also observed in the present study. Therefore, the normalized mutual information (NMI), based on entropy concepts, is used as a similarity measure.[21,23,27]

For a gray-level image I, the probability distribution, p_I, is defined as the normalized histogram of I. Specifically, $p_I(i)$ is equal to the number of pixels with intensity level i in image I divided by the total number of pixels in image I. The notation $H(I)$ denotes the entropy of an image I using the above definition for the underlying probability distribution of the image. $H(I)$ is defined by:

$$H(I) = -\sum_i p_I(i)\log[p_I(i)]. \tag{9}$$

Given two images, the joint probability distribution of images I and J (where I and J are assumed to have the same dimensions), $P_{I,J}(i,j)$, is equal to the number of pixels that have intensity level i in I and j in J divided by the total number of pixels of image I (or J). Consequently, the joint entropy of images I and J is defined by

$$H(I,J) = -\sum_{i,j} p_{I,J}(i,j)\log[p_{I,J}(i,j)], \tag{10}$$

and the NMI by[21,23,27]

$$NMI(I,J) = \frac{H(I)+H(J)}{H(I,J)}. \tag{11}$$

The images are assumed to be aligned when the NMI value is maximum. In the present work, the NMI is evaluated on intensity (gray-level) images derived from the color images by averaging the color components. The resulting gray-scale image is coded using 24 bits per pixel.

3. Optimization process

The next step is to update the transform $(T \rightarrow T')$ so that $NMI(I_1, T'(I_2)) \geqslant NMI(I_1, T(I_2))$, where T and T' belong to the set of transforms. Many algorithms (e.g., gradient, Newton's method, Levenberg–Marquardt method) rely on the possibility of estimating the gradient and/or the Hessian of the similarity function (with respect to the transformation's DOFs).[28] These techniques are not readily applicable to the NMI measure even though expressions exist for the gradient of the mutual information.[29] Popular algorithms that bypass the need for derivatives are Powell's algorithm and the simplex method.[30,31] Powell's algorithm optimizes each of the transformation's DOFs one by one. The simplex method considers all DOFs simultaneously and is, therefore, used in the present work.[32]

C. Interpolation

Interpolation is required in two different steps of the 3D reconstruction: registration and missing sections.

1. Registration interpolation

When an image is transformed, there is a high probability that the new grid does not line up with a Cartesian grid. Therefore, an interpolation is needed to sample the transformed image onto the Cartesian grid to permit computer implementation. During registration optimization, nearest-neighbor interpolation is used to interpolate the transformed images. Once the best transform is found, a more accurate 2D bi-cubic interpolation algorithm is used for the reconstruction of the registered sections.[33]

2. Missing section interpolation

During the histologic preparation procedures of sectioning, mounting and staining, some of the sections are lost or destroyed. Therefore, interpolation is necessary to replace the missing sections to avoid further artifacts in the 3D tissue model. After alignment, the missing sections are interpolated. To interpolate the missing sections, cubic Hermite interpolating polynomials are used. Specifically, a third-degree polynomial (four unknowns) is found to match slope and values of the signals (images) on each side of a "hole." This process yields four equations from which four unknowns are determined. The Hermite polynomials are chosen because they are *shape preserving*, i.e., extrema present in the available data are also extrema in the reconstructed signal. This property is of particular interest when dealing with images, because it guarantees that the reconstructed signals will remain within the range 0–255, thus avoiding the need for postprocessing to compress values into the range 0–255.

D. Acoustic impedance assignment

After completion of the previous steps, a 3D histologic map (3DHM) is obtained. The last step is to convert each 3DHM voxel to an adequate impedance value. To accomplish this task a color-threshold algorithm is carefully designed under the supervision of a board-certified pathologist, with each color representing a specific impedance value, to yield the final 3DZM. This part of the algorithm has not been changed and has been extensively discussed previously.[15]

IV. 3DZM-BASED IDENTIFICATION OF ULTRASONIC SCATTERING SITES

A. Estimation of scatterer properties

After the 3DZM is divided into smaller volumes called regions of interest (ROIs), the scatterer size and acoustic concentration are estimated for each ROI by an estimation routine that fits a chosen FF to the power spectrum of each individual ROI (see Sec. II C). A frequency-independent optimization routine has been implemented and tested previously[15] and is used in the present study.

B. 3DZM-based form factor extraction

A strategy is proposed to extract a tissue-specific FF from 3DZMs. The theoretical model is first introduced, then the methodology is described and finally the implementation is demonstrated.

1. Theory

Assume that a given 3DZM contains a single type of scatterer but with variations in size and local number density. The 3DZM is divided into N ROIs. Each ROI is large enough to contain a number of random-sized scatterers of the same type from which the FF is deduced (assuming the largest scatterer is much smaller than the ROI).

The power spectrum of the lth ROI is denoted by PS_l (where $1 \leqslant l \leqslant N$). Assume further that the scatterers are isotropic and described by the to-be-determined form factor $F^*(ka)$ and that the lth ROI contains n_l scatterers with random spatial locations within this ROI. The scatterers are assumed to have an acoustic impedance of z and the background impedance is z_0. Using Eq. (7) and Fourier transform arguments, one obtains

$$PS_l(k) = E\left[\left|\sum_{p=1}^{n_l} \frac{z-z_0}{z_0} V_s(a_p)\sqrt{F^*(ka_p)}\exp(i[k_x x_p + k_y y_p + k_z z_p])\right|^2\right], \tag{12}$$

where E[•] denotes the expected value; a_p and $V_s(a_p)$ are the radius and the volume of the pth scatterer in the lth ROI, respectively; x_p, y_p, and z_p describes the random location of the pth scatterer in the lth ROI; and k_x, k_y, and k_z are the spatial frequency variables of the 3D Fourier transform ($k = \sqrt{k_x^2+k_y^2+k_z^2}$). The term inside the absolute value bars is the 3D Fourier transform of the lth ROI. To expand and simplify this term, assume that the scatterer radii inside the lth ROI have a narrow distribution, i.e., $\forall p a_p \approx \langle a\rangle_l$, where $\langle a\rangle_l$ denotes the mean radius of the scatterers contained in the lth ROI. This approximation also yields $\forall p\ V_s(a_p) \approx V_s(\langle a\rangle_l)$. Thus, Eq. (12) reduces to

$$PS_l(k) \approx \left[V_s(\langle a\rangle_l)\frac{z-z_0}{z_0}\right]^2 F^*(k\langle a\rangle_l) \times \mathrm{E}\left[\sum_{p,q=1}^{n_l} e^{i[k_x(x_p-x_q)+k_y(y_p-y_q)+k_z(z_p-z_q)]}\right], \tag{13}$$

where the double sum term can be further reduced to

$$n_l + 2\sum_{p>q} \cos[k_x(x_p-x_q) + k_y(y_p-y_q) + k_z(z_p-z_q)], \tag{14}$$

where $\Sigma_{p,q=1}^{n_l} = \Sigma_{p=q} + \Sigma_{p\neq q}$ was used. The first term of Eq. (14) is the incoherent scattering and the second term represents the amount of coherence between the scatterers. For a large number of randomly located scatterers, the expected value of the second term of Eq. (14) is zero (i.e., the sum of cosines is a random process with a zero expected value when $k>0$). Thus, when $k>0$, Eq. (12) reduces to

$$PS_l(k) \approx n_l\left[V_s(\langle a\rangle_l)\frac{z-z_0}{z_0}\right]^2 F^*(k\langle a\rangle_l). \tag{15}$$

Equation (15) suggests that PS_l can be approximated by the to-be-determined FF of the unknown variable $k\langle a\rangle_l$ scaled by the unknown factor $n_l\left[V_s(\langle a\rangle_l)\frac{z-z_0}{z_0}\right]^2$.

2. Methodology

Based on the theory, the function F^* can be estimated from the 3DZM. Assume that the 3DZM is divided into N ROIs for which the assumptions of Sec. IV B 1 are valid over a spatial frequency range defined from k_1 to k_2. F^* will then be estimated from PS_l for every l ($1 \leqslant l \leqslant N$).

Each PS_l is transformed to $T_l(PS_l)$ using a transformation (T_l) with two DOFs (magnitude and frequency axis scaling) to account for the two unknowns $n_l\left[V_s(\langle a\rangle_l)\frac{z-z_0}{z_0}\right]^2$ and $k\langle a\rangle_l$. In the ideal case where Eq. (15) is valid, it should then be possible to determine a set of N transforms such that the transformed power spectra are all the same. F^* is deduced by normalizing the common power spectrum to unity when $k=0$.

Based on the theoretical assumptions, it is unrealistic to expect that the transformed power spectra fit perfectly. Therefore, a fitting criterion is introduced to quantify with a single number the similarity of the N transformed power spectra. This criterion is the mean standard deviation (MSTD) of the transformed power spectra over the chosen spatial frequency range. The quantity that needs to be minimized is

$$\mathrm{MSTD} = \frac{1}{k_2-k_1}\int_{k_1}^{k_2}\frac{1}{N}\sqrt{\sum_{l=1}^{N}[T_l(PS_l)(k) - \langle T(PS)(k)\rangle]^2}\mathrm{dk}, \tag{16}$$

where $\langle T(PS)(k)\rangle$ is the mean of the transformed power spectra and is defined by

$$\langle T(PS)(k)\rangle = \frac{1}{N}\sum_{l=1}^{N} T_l(PS_l)(k). \tag{17}$$

The algorithm is complete when the N transforms (i.e., $T_1,\ldots,T_N$) that minimize the MSTD, i.e., Eq. (16), are determined.

3. Implementation

The implementation is not straightforward because the fitting algorithm is attempting to optimize simultaneously $2N$ unknowns. To reduce the number of unknowns, the power spectra are first log transformed. The log of Eq. (15) transforms the magnitude scaling term $\left(n_l\left[V_s(\langle a\rangle_l)\frac{z-z_0}{z_0}\right]^2\right)$ to a magnitude shift $\left(\log\left(n_l\left[V_s(\langle a\rangle_l)\frac{z-z_0}{z_0}\right]^2\right)\right)$. The effect of this unknown magnitude shift is mitigated after the mean of each PS_l is removed. Specifically, each log power spectra is modified according to the following:

$$\log[PS_l'] = \log[PS_l] - \frac{1}{k_2 - k_1}\int_{k_1}^{k_2} \log[PS_l(k)]\mathrm{dk}. \quad (18)$$

Matching the mean values reduces the optimized DOFs to the N spatial frequency scaling coefficients that are defined by $T_l(PS_l)(k) = PS_l(k\alpha_l)$. From Eq. (15) if $\{\alpha_l, \text{for} 1 \leqslant l \leqslant N\}$ is a set of optimal scaling parameters, then so is $\{\chi\alpha_l, \text{for } 1 \leqslant l \leqslant N\}$ for χ arbitrary, thus allowing for the further reduction of the DOFs by enforcing a mean value of unity for the set of scaling coefficients. This $N-1$ parameter optimization problem is finally implemented using the simplex algorithm.[32]

Once the MSTD is minimized, $\langle T^*(PS)(k)\rangle$ (asterisk denotes the optimal transforms) is a scaled version of the to-be-determined FF, F^*, over the chosen spatial frequency range. However, to obtain the actual FF one also needs to transform $\langle T^*(PS)\rangle$ that is a function of k into a function of ka. The question is how to chose a; an heuristic approach of doing this consists, for example, of picking a such that

$$\frac{k_1 + k_2}{2} a = 1. \quad (19)$$

An alternate approach is to assume that the chosen spatial frequency range is such that $k_2 = 4k_1$, thus leading to a value of a such that the ka range is 0.5–2.0. Specifically, a would be deduced from $k_1 a = 0.5$ or equivalently $k_2 a = 2.0$. Nevertheless, unless prior information is known regarding the average size of the scattering structures, no perfect way exists to choose a. One way to establish bounds for the possible ranges of a is to limit the frequency analysis bandwidth to $0.5 < ka < 2.0$. In this study, a was chosen to be the value found with Gaussian FF for the same 3DZM (i.e., simulated or experimental). Denote the resulting function by $F(ka) = \langle T^*(PS)(k)\rangle$. This function of ka is now a magnitude-scaled estimate of $F^*(ka)$, that is, $F(ka) \approx BF^*(ka)$ for some unknown constant B.

$F(ka)$ is not expected to be a smooth curve, because it is a mean of N random processes (the N transformed power spectra) and for any 3DZM, the second term of Eq. (14) does not exactly go to zero. Several possible approaches exist to deduce a usable FF from F.

The simplest approach consists of keeping $F(ka)$ as it is; even though noisy, $F(ka)$ should be a curve with a fairly high signal-to-noise ratio (SNR) because it is the mean of the transformed power spectra that minimizes the MSTD. Each of these power spectra represents independent samples from a random process with the same statistics.

The next option is to conduct a smoothing of F^* by either linear filtering (e.g., low-pass filtering) or nonlinear filtering (e.g., median filtering). The problem with these filtering approaches is that the resulting estimate of F^* yields a function that is known only at the discrete locations where ka is sampled in the first place, which is a problem, for example, for the estimation step described in Sec. IV A. To avoid the need for interpolation, an approach is to model F^* as a function, $P(ka)$, that has a certain number of DOFs and find the DOF parameters that minimize the mean-squared error between F^* and P. Because the expression for the function P is closed form, it can be evaluated at any spatial frequency exactly without the need for interpolation. Furthermore, because the ultimate goal is to fit this to-be-determined FF to the available spectra, an advantageous model for fast optimization is the 2-parameter (i.e., α and n) exponential model

$$P_{(\alpha,n)}(ka) = e^{-\alpha(ka)^n}. \quad (20)$$

The exponential model reduces to the Gaussian FF when $\alpha = 0.827$ and $n = 2$. The choice for this model was motivated by our experience from simulations, ultrasound data and also because the model reduces to the Gaussian FF for specific values of α and n. In particular, mean-squared errors smaller than 3×10^{-5} were found when we modeled the fluid-filled, spherical shell, glass sphere, or fat sphere FF with Eq. (20).[34] Furthermore, this model has been previously used to model FFs over a given frequency range.[35] The 2-parameter exponential model is used in the remainder in this study. Therefore, the outputs of the FF-estimation algorithm are estimates of parameters α and n.

C. 3DZM Simulation methods

Each simulation was conducted in the same fashion: we simulated the actual power spectrum of each 3DZM. We did not simulate the 3DZM and then compute its spectrum. Both of these approaches are rigorously equivalent from a mathematical standpoint, but using the power spectrum of each 3DZM was much faster to implement.

In each simulation, the scatterers were modeled with the Gaussian FF. The spatial coordinates of the center of each scatterer were obtained from a realization of a uniform probability distribution function. The scatterer size was Gaussian distributed and quantified in terms of the mean radius or diameter and standard deviation. After all the spatial locations and sizes were obtained, the simulated power spectra were obtained by summing in the Fourier domain the contribution of each scatter as in Eq. (12). This summation could be performed because the FF of each scatterer is known and because spatial shifts become phase shifts in the Fourier domain.

Two sets of simulation studies were conducted as described below.

1. Simulation A

Previous simulation studies evaluated 3DZMs that contained either a single population of scatterers of the same size or contained two populations of scatterers of two different sizes for which the fluid-sphere FF was used.[15] The purpose of this simulation study is to evaluate further how the presence of a size distribution affects the estimates. The previous work is extended to a set of simulations involving 3DZMs containing single populations of Gaussian FF scatterers whose sizes follow a Gaussian distribution with a mean diameter of 40 μm. Nineteen 3DZMs are generated, each with a different standard deviation, σ, for Gaussian size distribution that varied between 2 and 20 μm in 1 μm increments.

TABLE I. Simulation parameters and results. For each simulation the size distribution and number density listed were used for each ROI, G and D denotes for Gaussian and deduced, respectively. The range values (i.e., symbol "–") mean that for each ROI, the corresponding parameter was chosen randomly and uniformly within this range.

Simulation Number	FF	Number density $10^3/\text{mm}^3$	G size distribution		N	Results			
			⟨diameter⟩ (μm)	σ (μm)		α	n	G FF Mean error (%)	D FF Mean error (%)
1	G	4–8	32	0	25	0.796	2.08	2.5	2.4
2	G	4–8	32	8	25	1.46	1.66	15	2.5
3	G	4–8	16–48	0	25	0.796	2.07	3.2	3.5
4	G	4–8	16–48	0–48	25	3.24	1.04	27	19

The scatterers have an impedance of 1.51 Mrayl with a background of 1.50 Mrayl, the same as the previous simulations.[15] The number density of the scatterers was $4\times10^3/\text{mm}^3$ for the 19 simulated 3DZMs. Each 3DZM was a cube with a side length of 500 μm.

2. Simulations 1–4

The purpose of these simulation studies is to assess the methodology to extract a FF from a 3DZM. Each of the four simulations consists in the generation of 25 ROIs (i.e., $N=25$). We found that using 25 ROIs yields adequate results and is not too computationally extensive for simplex optimization. Each of the 25 spectra of each simulation was generated using the parameters listed in Table I. Each ROI was a cube with a side length of 250 μm.

Each ROI is populated spatially by uniformly distributed scatterers. For each ROI the number density of scatterers is random and ranges between $4\times10^3/\text{mm}^3$ and $8\times10^3/\text{mm}^3$. The sizes of the scatterers within a given ROI follow a Gaussian distribution whose parameters for some simulations are constant and for others have a range indicated in Table I. The scatterers have an impedance of 1.51 Mrayl with a background of 1.50 Mrayl. The analysis frequency band is 7.66–30.7 MHz, from $ka=0.5$ to $ka=2$ when $c=1540$ m/s for a radius $a=16$ μm.

Each set of simulations consists of a three-step process: (1) 25 random ROIs are generated to model the 3DZM. (2) A 3DZM-deduced FF is extracted from the power spectra of each of the 25 ROIs using the 2-parameter exponential law, Eq. (20), yielding values for α and n. (3) The extracted FF is then used with the 25 power spectra to obtain estimates following the methodology presented in Sec. IV A. Estimates using the Gaussian FF (often used for QUS studies[3]) are also obtained, allowing for comparison of the estimates between the deduced FF and the Gaussian FF. Four different simulations assess quantitatively the FF-extraction methodology.

V. RESULTS

Simulated and experimental results obtained using the advanced methodologies are presented. Simulations evaluate the FF-estimation algorithm. The advanced 3DZM reconstruction strategies are illustrated with a mouse sarcoma dataset. Finally, results from three experimental 3DZMs are presented. The three experimental 3DZMs are a rat mammary fibroadenoma, a mouse mammary carcinoma, and a mouse sarcoma.

A. Simulation results

1. Simulation A: Gaussian size simulations

The larger scatterers in the distribution of sizes cause a bias in the estimates towards a larger value than the mean scatterer size when a Gaussian FF is used as the scattering model. This is observed in Fig. 1 by the increasing mean diameter estimate as a function of the increasing standard deviation. Even though biased, the estimates are precise because the standard deviation of the estimates is always small (<0.5 μm). The bias of the estimates is a problem if the absolute (mean) sizes of scattering structures are of importance. The size estimates represent a weighted average of the distribution of scatterer sizes in the population of scatterers. Scatterers with larger size distributions will have a volume-weighted average estimate of size that is much larger than the actual average size compared to smaller size distributions.

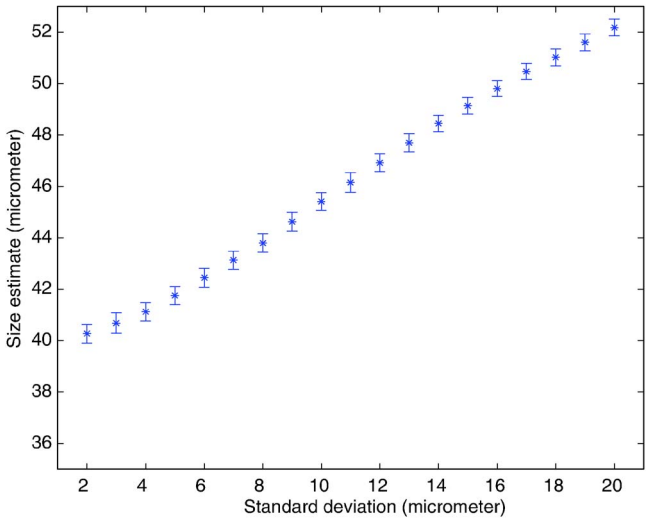


FIG. 1. (Color online) Size estimates obtained using a Gaussian FF. The simulated 3DZM contained Gaussian scatterers whose size distribution was Gaussian with a mean diameter of 40 μm. The STD of the size distribution varied from 2 to 20 μm by 1 μm increments. For each STD value, 64 3DZMs were generated. Error bars represent STDs of estimates.

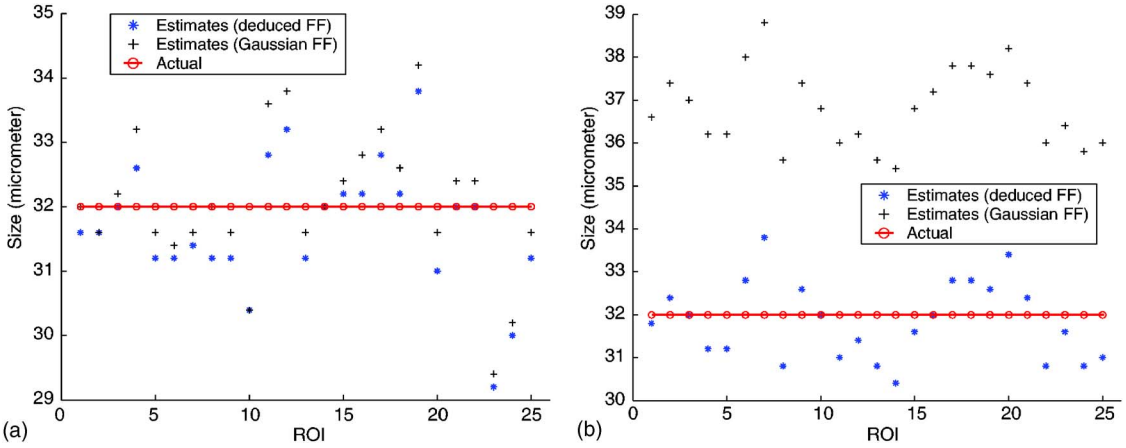


FIG. 2. (Color online) Estimates obtained from 25 ROIs using the Gaussian and the 3DZM-deduced FF. (a) Each ROI was filled with a random number of randomly located Gaussian scatterers. Each ROI contained scatterers of diameter 32 μm. (b) Each ROI was filled with a random number of randomly located Gaussian scatterers. Each ROI contained scatterers following a Gaussian size distribution with mean diameter of 32 μm and standard deviation of 8 μm.

2. Simulations 1–4: Validation of the 3DZM-FF estimation

a. Simulation 1. This simulation tests the FF-extraction methodology because in this case the optimal FF is the Gaussian FF [Fig. 2(a)]. The accuracy of the estimates is good; all the size estimates (Gaussian FF and deduced FF) are in the range 29–35 μm (i.e., within 10%). There is no visible difference between the Gaussian FF and the deduced FF estimates. The mean errors are 2.5% and 2.4% for the Gaussian and deduced FFs (Table I), respectively. These results validate that for this simulation the FF-extraction algorithm is capable of deriving an accurate FF. Specifically, the parameters found from the FF-extraction algorithm are $\alpha=0.796$ and $n=2.08$, which are near (within 4%) the Gaussian values ($\alpha=0.827$ and $n=2$).

b. Simulation 2. The estimation routine using a Gaussian FF is expected to overestimate the mean size because of the non-zero standard deviation [Fig. 2(b)]. The Gaussian estimates are greater than the actual mean scatterer size; the mean Gaussian error is 15%. The deduced FF estimates do not show any bias and the deduced FF mean error is 2.5%. The FF-extraction algorithm estimates $\alpha=1.46$ and $n=1.66$. These values are within 7% of the theoretical values ($\alpha=1.54$ and $n=1.57$) for this Gaussian distribution of Gaussian scatterers.[34] This agreement demonstrates the ability of the FF-extraction algorithm to estimate the correct FF.

c. Simulation 3. This simulation tests the ability of the algorithm to find the correct spatial frequency axis scaling coefficients (Table II). Furthermore, in this simulation ($\sigma=0$ for each ROI) the assumptions of Sec. B are valid [e.g., the transition from Eq. (12) to Eq. (13)]. The optimal FF for this simulation is the Gaussian FF because there is no variance in the size of the scatterers within a given ROI. Both FFs yield accurate estimates even though the actual diameters are randomly selected in the large range 16–48 μm. The mean errors are 3.2% and 3.5% for the Gaussian and deduced FFs, respectively. The FF-extraction algorithm estimates $\alpha=0.796$ and $n=2.07$, values that are within 4% of the actual Gaussian values. Thus, this simulation demonstrates the ability of the FF-extraction algorithm to determine accurate spatial frequency scaling coefficients.

d. Simulation 4. This simulation tests whether the FF-extraction algorithm is capable of extracting a FF from 25 ROIs that have different size distributions but contain the same (Gaussian FF) scatterers (Table II). For this simulation, the Gaussian estimates of size should be greater than the actual mean values, with a bias that depends on the (random) standard deviation of each ROI. From the results, the Gaussian estimates of size are greater than the actual mean values in all except one ROI. The deduced FF estimates of size are sometimes greater and sometimes smaller than the actual mean values. The mean errors are 27% and 19% for the Gaussian and deduced FFs, respectively. These larger errors indicate that neither of the two FFs are accurate scattering models. The FF-extraction algorithm estimates $\alpha=3.24$ and $n=1.04$. The fact that the Gaussian FF produces inaccurate estimates was expected. However, the fact that the deduced FF produces inaccurate estimates is unexpected. The problem may be due to the two-parameter exponential model that may not be capable of tracking the three DOFs: the random mean radius, the random standard deviation, and the random ROIs. One suggestion to mitigate this problem is to fit the extracted FF to a model with three or more DOFs. A three-parameter FF can be derived theoretically by extending further the theory of Sec. IV B.

These four simulations demonstrate the ability of the FF-extraction algorithm to accurately model complex scattering media. In particular, the 3DZM-deduced FF greatly outperformed the Gaussian FF for these advanced media in terms of estimate bias even though the media contain only Gaussian scatterers.

B. Experimental results

1. Sarcoma 3DZM reconstruction

The 3D reconstruction strategies are developed and demonstrated with a murine sarcoma tumor model. The EHS (Englebreth–Holm–Swarm) tumor cell line (CRL-2108, American Type Culture Collection, Manassas, VA) is a transplantable sarcoma in C57BL/6 mice. This tumor produces extracellular matrix (ECM) components such as laminin, collagen IV, entactin, and heparan sulfate proteoglycan. The cells are injected subcutaneously and the tumor was allowed to grow until it was approximately 1 cm in diameter. Details of cell maintenance and tumor growth with this animal model have been published.[36] The experimental protocol was approved by the Institutional Animal Care and Use Commit-

TABLE II. Actual (A), Gaussian (G) FF, and deduced (D) FF mean size estimates (μm) for each of the 25 ROIs of Simulations 3 and 4.

	Simulation 3			Simulation 4		
ROI number	A	G FF	D FF	A	G FF	D FF
1	18.2	20.1	19.5	38.0	41.8	37.8
2	20.8	17.9	18.1	46.5	45.0	45.0
3	16.9	14.8	14.9	37.8	42.5	39.9
4	30.1	29.9	29.9	28.1	38.8	34.1
5	29.4	29.4	29.4	19.5	36.5	29.8
6	42.1	42.1	41.7	42.5	41.5	37.5
7	37.5	37.0	36.9	29.1	39.1	34.1
8	36.0	36.0	35.9	20.8	31.8	20.6
9	24.9	22.1	23.1	28.0	32.0	30.0
10	25.9	24.9	24.9	38.5	43.1	40.2
11	42.0	41.0	40.5	18.9	33.8	26.9
12	25.3	25.2	25.2	24.9	41.9	36.9
13	23.9	27.1	27.1	30.0	39.0	34.0
14	42.1	42.3	42.1	20.8	24.9	12.9
15	35.6	34.9	34.8	22.2	28.1	16.8
16	33.8	34.8	34.7	23.2	27.1	16.4
17	42.5	42.4	41.6	33.8	38.9	33.8
18	26.1	24.9	24.9	21.1	38.1	33.2
19	36.5	36.7	36.5	45.2	45.5	45.6
20	22.5	21.5	21.2	22.5	31.2	20.8
21	45.1	46.1	46.2	42.5	43.3	42.0
22	32.0	31.8	31.8	23.1	23.7	12.1
23	29.1	29.1	29.1	26.1	27.8	18.2
24	25.1	24.9	24.9	33.7	40.0	25.5
25	44.8	44.4	44.1	30.0	37.2	30.5

tee, University of Illinois, Urbana-Champaign, and satisfied all University and NIH rules for the humane use of laboratory animals.

Sections of the sarcoma are fixed in 10% neutral-buffered formalin, embedded in paraffin, sectioned at 3 μm thickness, mounted on glass slides, and stained with H&E for evaluation. These processes cause a minor degree of inherent tissue shrinkage that is small, but indeterminate for each section. In addition, it is not possible to obtain 200 serial sections (one section after another) without the loss of some sections due to technical difficulties.

The sarcoma sections are examined and captured (see Sec. III). The optical (bitmap) images measure laterally 400 μm (640 pixels) by 300 μm (480 pixels) with 8 bit accuracy for each red, green and blue component. The EHS dataset contains 200 sections of which 54 sections (27%) were lost during tissue preparation.

The 3D reconstruction strategies are then applied to the EHS dataset. These 43 consecutive sections (Fig. 3) are 218 by 156 μm (i.e., 350 by 250 pixels) subimages of the original 400 by 300 μm EHS images. Among these 43 consecutive sections, seven are missing (including four consecutive sections). The sections are misaligned, and the contrast between sections is also slightly different. The misalignment can be observed by following the pink quasicircular structure (muscle) from one section to the next. In addition, sections 8 and 17 have slightly better contrast than the neighboring sections. In each section of the EHS dataset, the histopathologic characteristics of the tissue that are diagnostically consistent with an EHS sarcoma, that is, the islands of tumor cells intermixed with ECM, are clearly visible.

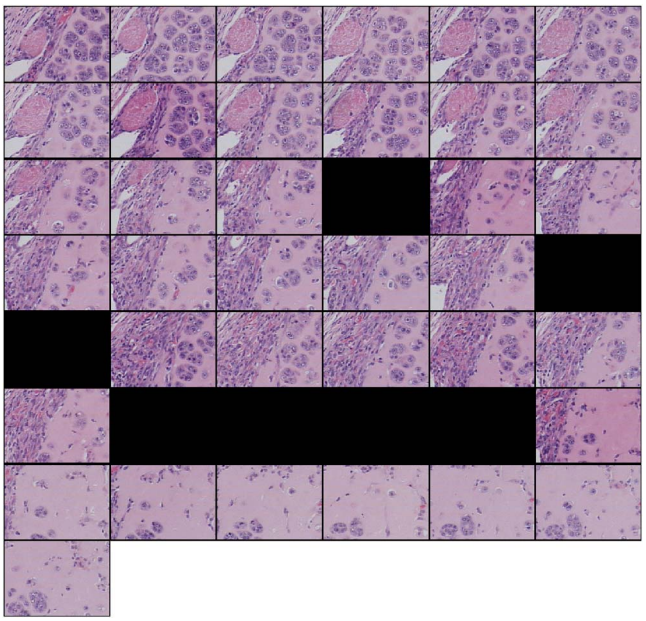


FIG. 3. (Color online) Forty-three-section dataset from the EHS dataset. Sections are of size 218 × 156 μm (i.e., 350 × 250 pixels). Sections are subimages of size 218 × 156 μm (i.e., 350 × 250 pixels) of the original 400 × 300 μm EHS sections. Subimages were extracted from the same location in each of the 43 original EHS sections. The sections were consecutive from left to right and top to bottom. Black fields symbolize missing sections.

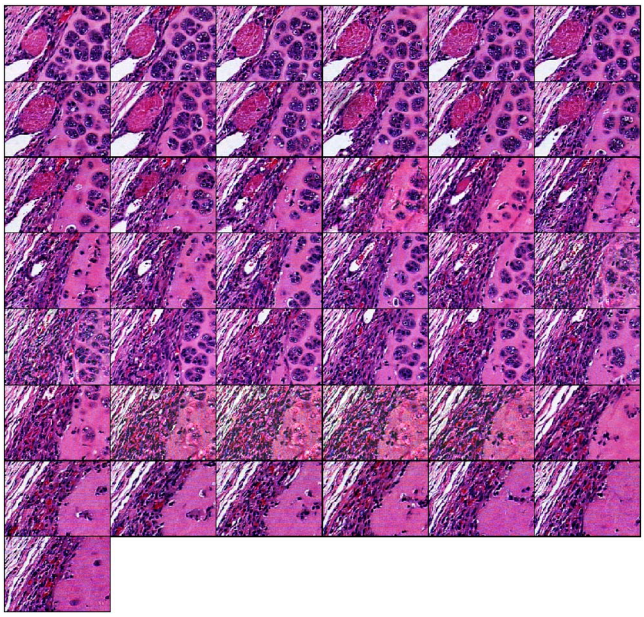


FIG. 4. (Color online) Reconstruction of the 43-section EHS dataset of Fig. 3. The contrast of the available sections was equalized. Each section was affine registered with the next available section. Missing sections were Hermite interpolated.

Figure 4 shows the reconstructed 43-section dataset. To align the images, affine registration was used. The reconstructed images have a similar contrast. The alignment can be observed by following the oblique structures in the top left corner. The interpolated missing sections are of lower quality than the others because the interpolated missing sections are not as sharp and contained some edge artifacts. This is particularly true for the four reconstructed images of the four consecutive missing sections in which the purple background in the bottom-right corners of the interpolated images is not as uniform as in the other sections. In addition, the background appears to be "noisy" and to contain random small structures with blurry edges. Except for the artifacts noted, in each section of the EHS dataset, the histopathologic characteristics of the tissue are diagnostically consistent with an EHS sarcoma.

Figure 5(a) shows a 3D rendering from the 43-section reconstructed EHS dataset. The interpolated section artifacts are noticeable on the bottom half of the left side of the volume where the rendering is blurry. To obtain a 3DZM of the EHS sarcoma, seven distinct impedance values are used: 2.00 Mrayl for the nuclear heterochromatin [black on Fig. 5(b)], 1.85 Mrayl for the nuclear euchromatin (blue), 1.70 and 1.65 Mrayl for the extracellular matrix (dark red/red), 1.60 for the vascular space/whole blood (white), and 1.58 and 1.55 for the cytoplasm (green/yellow). The 3DZMs and impedance values for the fibroadenoma and carcinoma were published previously.[15]

2. Scatterer property estimates

The frequency-independent estimation algorithm is conducted on four ROIs on the three 3DZMs (fibroadenoma, carcinoma and sarcoma) using the Gaussian FF and the 3DZM-deduced FF. The use of four ROIs was to reduce

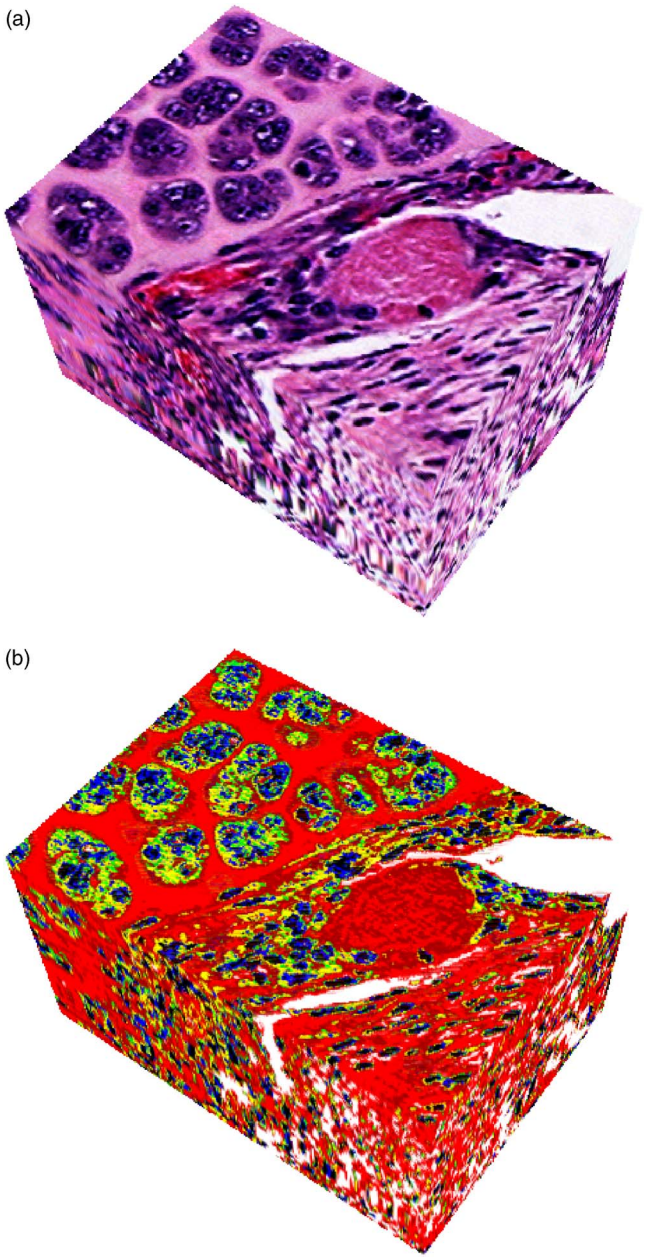


FIG. 5. (Color online) (a) Three-dimensional rendering from the 43-section dataset (Fig. 4). (b) Derived 3D impedance map. The volumes shown in (a) and (b) are of size $218\times156\times129$ μm (depth).

estimate variance through averaging. The dimensions of the 3DZMs and of the ROIs are displayed in Table III. We chose to use only four ROIs because if we decided to further divide the 3DZM, then each individual ROI would have been too small to possibly contain a sufficient number scatterers in the size range being investigated. For each experimental 3DZM, the 3DZM-deduced FF was obtained using 25 overlapping ROIs (i.e., $N=25$, adjacent ROI overlap by 75%). We use 25 ROIs because the same number was chosen for the simulation and yields acceptable results. Going beyond $N=25$ would increase further the redundancy of the data within two adjacent ROIs.

The 3DZM results along with the independent ultrasonic estimates are summarized in Table IV. Ultrasonic results were obtained by using standard *in vivo* backscattered

TABLE III. The 3DZM and ROI dimensions in μm used for scatterer property estimation and FF extraction.

	Rat fibroadenoma	Mouse carcinoma	Mouse sarcoma
3DZM width	800	200	218
3DZM length	600	150	156
3DZM depth	390	330	129
ROI width	400	100	109
ROI length	300	75	78
ROI depth	390	330	129

methodology.[4] (Some of the results presented in Table IV were reported before and are indicated by the superscript *,[15] and a topic for future research.)

Experimentally derived estimates were obtained by measuring the ultrasonic backscatter from the tumors using a Gaussian FF.[4] The experimentally derived estimates represent an independent measure from those of the 3DZM estimates. The mean scatterer size and acoustic concentration 3DZM-based estimates (Table IV) are obtained by computing the mean and standard deviation of the estimates over the four ROIs for each of the three 3DZMs. Results have good agreement (difference less than 10%) between ultrasonic and 3DZM estimates of size estimates for three tumors. However, the acoustic concentration values are significantly different (difference greater than 6 dB mm^{-3}). The difference in acoustic concentration has already been observed and discussed previously.[15]

The deduced FF study has yielded interesting results. (1) The exponential fit FFs deduced using the FF-extraction methodology are different for each tumor (Table V). This difference may indicate that each tumor has its own unique ultrasonic scattering signature (i.e., FF) that is not detected using a Gaussian FF. The results suggest that the tumor-specific FF may be used to characterize, distinguish and possibly diagnose disease with QUS techniques. (2) Using the deduced FF the mean diameter estimate is now statistically different for each of the three tumors (Table IV). The specificities of the estimates are potentially significant for diagnosing (classifying) tissue.

VI. DISCUSSION

The purpose of this study has been to report on continued developments of strategies to identify and characterize ultrasonic scattering sites in biological tissue. In our previous contribution,[15] a novel approach to identify the anatomical scattering sources was introduced. This approach uses volume sections (3D histologic maps) that correspond to *in vivo* scanned tumor volumes to generate a 3D impedance map (3DZM). In the previous contribution, 3DZMs were created by manually aligning serial photomicrographs of tumor sections. In this contribution, an automated capability is presented that enhances, aligns and interpolates serial photomicrographs. In the previous contribution, 39 10-μm-thick rat fibroadenoma sections were assigned three impedance values and 66 5-μm-thick mouse mammary carcinoma sections were assigned five impedance values. In this contribution, 200 3-μm-thick mouse sarcoma sections were assigned seven impedance values. In the previous contribution, the Gaussian FF was used to estimate QUS parameters (size and acoustic concentration) from the fibroadenoma and carcinoma 3DZMs. In this contribution, a tissue-specific FF using a two-parameter exponential model [Eq. (20)] is developed, validated, and used to estimate QUS parameters from the 3DZMs of all three tumors. Each of these incremental improvements represents a significant advance to yield an accurate tissue macrostructure- and microstructure-based computational phantom for identifying and quantifying the anatomical scattering sources. Thus, accurate 3DZMs enable several important issues to be resolved or better understood: (1) The comparison between tumor 3DZMs and ultrasound scatterer property estimates of the same tumors will allow precise connections to be made between the microanatomical and ultrasound echo features. (2) The 3DZMs will allow FFs to emerge directly from histological analyses by assigning impedance values to candidate structures and calculating the corresponding spatial autocorrelation. (We already have developed a "new cell model" based on this idea.[6]) (3) Tissue anisotropy can be deduced by examination of the 3DZMs. Determining the anisotropy of a tissue (deduced from 3DZMs) will be important for understanding the interaction of ultrasound with the tissue and formulation of models. (4) Periodicities in the tissue structures can be quantified through the 3DZMs because current backscatter models assume scattering from randomly located structures.

TABLE IV. The 3DZMs and ultrasound estimates. Ultrasound (US) estimates were obtained using the Gaussian FF. The 3DZM estimates were obtained using the Gaussian (G.) and the 3DZM-deduced (D.) FF.

	Rat fibroadenoma		Mouse carcinoma		Mouse sarcoma	
	Diameter μm	Conc. dB mm^{-3}	Diameter μm	Conc. dB mm^{-3}	Diameter μm	Conc. dB mm^{-3}
US	105 ± 25	-15.6 ± 5	30.0 ± 9.6	10.6 ± 6.9	33.0 ± 8.0	9.9 ± 5.3
3DZM G.	$91.5 \pm 25^*$	$-21.9 \pm 6.1^*$	$31.5 \pm 2.5^*$	$-1.4 \pm 6.1^*$	$32.9 \pm 6.1^*$	$-0.59 \pm 4.3^*$
3DZM D.	96.5 ± 14	-21.8 ± 5.1	56.0 ± 19	-6.9 ± 7.2	37.0 ± 12	0.46 ± 5.5

TABLE V. Parameters and frequency range of the exponential fit obtained using the FF extraction methodology. Exponential fit parameters were extracted from the three experimental 3DZMs.

	Frequency range (MHz)	α	n
Rat fibroadenoma	0.5–15	1.501	0.6145
Mouse carcinoma	1–25	2.359	0.4054
Mouse sarcoma	1–25	1.070	1.580

A. 3DZM validation

Within this paper, theories relating 3DZMs to scatterer property estimates are presented and tested through improved 3DZM construction, simulations, and experimental data. Both qualitative and quantitative assessments of the 3D reconstruction strategies have been conducted, but they extend beyond the scope of this paper.[34] In particular, on well-controlled cases it was quantitatively demonstrated that the registration algorithm was robust. Furthermore, interpolated sections were evaluated by a board-certified pathologist who was able to diagnose the tumor.

Nevertheless, work is still needed to explicitly validate 3DZMs. In particular, acoustic impedances were empirically assigned.[15] These assignments were conducted jointly between a board-certified pathologist (who described the tissue constituent and its amount) and an ultrasonic tissue property expert (who conducted and published many ultrasonic tissue measurements). Because scattering properties depend significantly on the assigned impedance values, efforts should be pursued to acquire independent and reliable ultrasonic tissue property measurements at the micron-scale level.

An alternate technique to be considered is to deduce ultrasonic tissue property values by iteration. The 3DZMs could be used as computational phantoms to simulate backscattered signals to be compared with real experimental backscatter data. At each step, the 3DZM would be updated until the simulated backscattered signals matched the signals obtained by ultrasonic measurements.

B. Gaussian FF limitations

Estimating scattering properties using a Gaussian FF can be misleading even in the case of Gaussian scatterers (Sec. IV A 1). Therefore, while the Gaussian FF allows fast optimization schemes,[15] the results might not be useful for identifying and quantifying the actual scattering structures.

The size estimates of the tumors do not allow for statistical distinction between the carcinoma and sarcoma (Table IV) using the Gaussian FF over the analysis bandwidth. The three tumors were chosen because they contained distinct histopathologic features (Fig. 6), features that are used by pathologists for diagnosis. An explanation may be that the Gaussian FF is an inadequate scattering model and is unable to track the specific ultrasonic scattering properties expected from these two tumors over the chosen bandwidth. Over the same bandwidth, improved FFs may yield greater sensitivities to the differences between the carcinoma and sarcoma using characterization techniques, as demonstrated successfully in Refs. 6 and 36.

C. Identification of the ultrasonic scattering sites

The initial motivation in the development of the 3DZMs was that they represent accurate morphological models of tissue microstructure that can aid in identifying and quantifying scattering sites. Identifying and quantifying the ultrasonic scattering sites would lead to significant improved capabilities for diagnosing pathologies using QUS techniques. Parameters estimated could then be chosen that actually describe histologic features of tissue microstructures.

1. Identification using size estimates

One approach to identify anatomic scattering sites is to utilize the information from the histologic images. For the fibroadenoma, all the size estimates are near 100 μm (Table IV). The circular white structures, acini with neighboring epithelial cells, have diameters near 100 μm [Fig. 6(a)]. Thus, it is likely that for the fibroadenoma the anatomic structures responsible for scattering are the acini.

The carcinoma size estimates obtained with either the Gaussian or the deduced FF (Table IV) do not correspond to any specific histologic structures. The histology is very dense with a lot of tumor cells (nuclei around 10 μm in diameter) and very little ECM [Fig. 6(b)]. These structures are much smaller than those obtained via QUS estimation techniques. Thus, the estimates suggest that the Gaussian FF and the deduced FF are not good scattering models for this tumor. A recent study also demonstrated that carcinoma cells in pellet form had the same ultrasonic size estimates using a Gaussian FF.[36] Thus, it is likely that individual cells combined with their inner constituents (e.g., cytoplasm, nuclei, organelles, and cytoskeleton) are significantly responsible for ultrasonic scattering.

For the sarcoma, all the size estimates (Table IV) are in the range 30–40 μm in diameter. This range is consistent with the diameters of the islands of tumor cells [Fig. 6(c)]. Furthermore, the ECM background is quite uniform, which adds confidence that the anatomic structures responsible for scattering may be the islands of tumor cells.

Figure 6 also raises the concern for having periodic scattering structures. Periodic structures would in particular violate the uniform random spatial locations assumption of scatterers used for the simulation and estimation methods. Our previous experience in which ultrasonic backscatter analysis of these tumors matched the 3DZM approach[15] and more recent study[37] reveal that the assumption of randomness was most likely not violated.

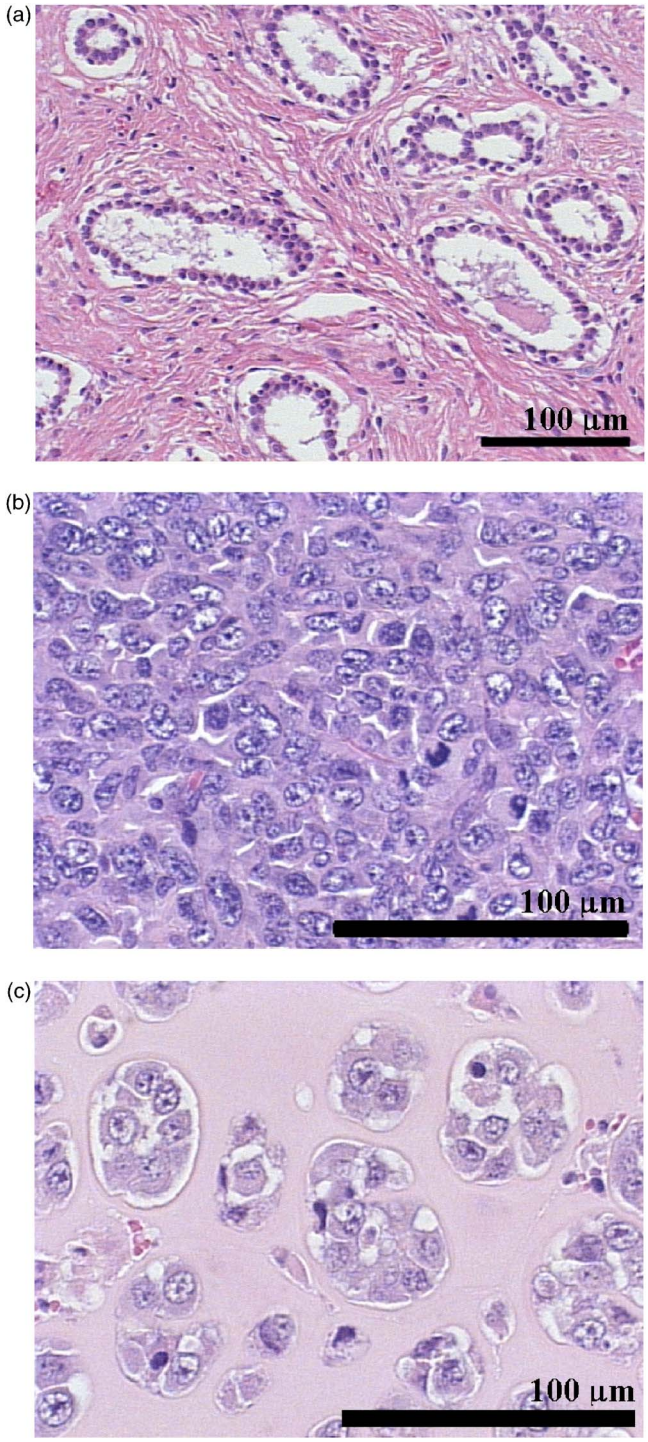


FIG. 6. (Color online) Typical H&E stained tissue sections for a rat mammary fibroadenoma (a), a 4T1 mouse mammary carcinoma (b), and a EHS mouse sarcoma (c). This figure demonstrates that the histopathologic properties of the tumor tissue investigated are very different.

2. Identification using form factors

Another approach to identify anatomic scattering structures is to consider the computational phantom capability (3DZMs) to deduce FFs. The 3DZM-deduced FFs could be iteratively compared to ultrasound backscatter signatures from either the same tissues from which *in vivo* ultrasonic echo data were acquired or from computational echo data acquired for the 3DZM, or both. When the ultrasonic echo data converges with the 3DZM-deduced FFs, the FF-SAF duality (Sec. III B) would help to identify the scattering sources. Further, from the inverse Fourier transform of the 3DZM-deduced FF, a 3D acoustic model of the SAF of a scattering structure can be determined. Then, it might be possible to identify the scattering sites by comparison of the SAF characteristics with the histology.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support by NIH Grant No. CA111289 and by the University of Illinois Research Board. J.M. thanks the Riverside Research Institute and the Lizzi Center for Biomedical Engineering for its support in enabling the preparation of this paper.

[1]M. F. Insana, R. F. Wagner, and D. G. Brown, "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192 (1990).
[2]E. J. Feleppa, F. L. Lizzi, D. J. Coleman, and M. M. Yaremko, "Diagnostics spectrum analysis ophthalmology: A physical perspective," Ultrasound Med. Biol. **12**, 623–631 (1986).
[3]M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., "Characterization of tissue microstructure using ultrasonic backscatter: Theory and technique for optimization using a Gaussian form factor," J. Acoust. Soc. Am. **112**, 1202–1211 (2002).
[4]M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., "Parametric imaging of rat mammary tumors in vivo for the purposes of tissue characterization," J. Ultrasound Med. **21**, 1201–1210 (2002).
[5]M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, "Differentiation and characterization of rat mammary fibroadenomas and 4tl mouse carcinomas using quantitative ultrasound imaging," IEEE Trans. Med. Imaging **23**, 764–771 (2004).
[6]M. L. Oelze and W. D. O'Brien, Jr., "Application of three scattering models to characterization of solid tumors in mice," Ultrason. Imaging **28**, 83–96 (2006).
[7]D. J. Coleman, R. H. Silverman, M. J. Rondeau, H. C. Boldt, H. O. Lloyd, F. L. Lizzi, T. A. Weingeist, X. Chen, S. Vangveeravong, and R. Folberg, "Noninvasive in vivo detection of prognostic indicators for high-risk uveal melanoma: Ultrasound parameter imaging," Ophthalmology **111**, 558–564 (2004).
[8]J. E. Perez, J. G. Miller, B. Barzilai, S. Wickline, G. A. Mohr, K. Wear, Z. Vered, and B. E. Sobel, "Quantitative characterization of myocardium with ultrasonic imaging," J. Nucl. Med. Allied Sci. **32**, 149–157 (1988).
[9]F. L. Lizzi, M. Greenebaum, E. J. Feleppa, M. Elbaum, and D. J. Coleman, "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1373 (1983).
[10]F. L. Lizzi, M. Ostromogilsky, E. J. Feleppa, M. C. Rorke, and M. M. Yaremko, "Relationship of ultrasonic spectral parameters to features of tissue microstructure," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 319–329 (1986).
[11]M. F. Insana, J. G. Wood, and T. J. Hall, "Identifying acoustic scattering sources in normal renal parenchyma in vivo by varying arterial and ureteral pressures," Ultrasound Med. Biol. **17**, 613–626 (1991).
[12]R. E. Baddour, M. D. Sherar, J. W. Hunt, G. J. Czarnota, and M. C. Kolios, "High-frequency ultrasound scattering from microspheres and single cells," J. Acoust. Soc. Am. **117**, 934–943 (2005).
[13]M. C. Kolios, G. J. Czarnota, M. Lee, J. W. Hunt, and M. D. Sherar, "Ultrasonic spectral parameter characterization of apoptosis," Ultrasound Med. Biol. **28**, 589–597 (2002).
[14]J. Mamou, M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., "Ultrasound scatterer size estimation technique based on a 3D acoustic impedance map from histologic sections," *Proceedings of the 2003 IEEE Ultrasonics Symposium*, Honolulu, HI, pp. 1022–1025 (2003).
[15]J. Mamou, M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, "Identifying ultrasonic scattering sites from 3D impedance maps," J. Acoust. Soc. Am. **117**, 413–423 (2005).
[16]J. Mamou, M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, "Ultrasound characterization of three animal mammary tumors from three-dimensional acoustic tissue models," *Proceedings of the 2005 IEEE Ultrasonics Symposium*, Rotterdam, Netherlands, pp. 866–869 (2005).

[17]J. Mamou, M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, "A view on quantitative biomedical ultrasound imaging," IEEE Signal Process. Mag. **23**, 112–116 (2006).
[18]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).
[19]K. K. Shung and G. A. Thieme, *Ultrasonic Scattering in Biological Tissues* (CRC, Boca Raton, FL, 1993).
[20]V. Caselles, J.-L. Lisani, J.-M. Morel, and G. Sapiro, "Shape preserving local histogram modification," IEEE Trans. Image Process. **8**, 220–230 (1999).
[21]J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," IEEE Trans. Med. Imaging **22**, 986–1004 (2003).
[22]J. Kybic and M. Unser, "Fast parametric elastic image registration," IEEE Trans. Image Process. **12**, 1427–1442 (2003).
[23]J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes, *Medical Image Registration* (CRC, Boca Raton, FL, 2001).
[24]G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," IEEE Trans. Med. Imaging **17**, 586–595 (1998).
[25]A. A. Cole-Rhodes, K. L. Johnson, J. LeMoigne, and I. Zavorin, "Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient," IEEE Trans. Image Process. **12**, 1495–1511 (2003).
[26]C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," Pattern Recogn. **32**, 71–86 (1999).
[27]Y.-M. Zhu, "Volume image registration by cross-entropy optimization," IEEE Trans. Med. Imaging **21**, 174–180 (2002).
[28]D. P. Bertsekas, *Nonlinear Programming* (Athena Scientific, Belmont, MA, 1999).
[29]F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," Med. Image Anal. **3**, 373–386 (1999).
[30]M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," Comput. J. **7**, 155–162 (1964).
[31]F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," IEEE Trans. Med. Imaging **16**, 187–198 (1997).
[32]J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J. **7**, 308–313 (1965).
[33]N. A. Dodgson, "Quadratic interpolation for image resampling," IEEE Trans. Image Process. **6**, 1322–1326 (1997).
[34]J. Mamou, "Ultrasonic characterization of three animal mammary tumors from three-dimensional acoustic tissue models," Ph.D. Dissertation: University of Illinois at Urbana-Champaign (2005).
[35]T. A. Bigelow, M. L. Oelze, and W. D. O'Brien, Jr., "Estimation of total attenuation and scatterer size from backscattered ultrasound waveforms," J. Acoust. Soc. Am. **117**, 1–10 (2005).
[36]M. L. Oelze and J. F. Zachary, "Examination of cancer in mouse models using quantitative ultrasound," Ultrasound Med. Biol. **32**, 1639–1648 (2006).
[37]M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, "Quantitative ultrasound assessment of breast cancer using a multiparameter approach," *Proceedings of the 2007 IEEE Ultrasonics Symposium*, New York City, NY (2007).