

# THE JOURNAL of the Acoustical Society of America

Vol. 106, No. 4, Pt. 1

October 1999

**EDITORIAL**

Current criteria for selection of articles for publication Allan D. Pierce 1613

**SOUNDINGS SECTION**

**ACOUSTICAL NEWS—USA** 1617

USA Meetings Calendar 1618

**ACOUSTICAL NEWS—INTERNATIONAL** 1621

International Meetings Calendar 1621

**REVIEWS OF ACOUSTICAL PATENTS** 1625

**TUTORIAL REVIEW [10]**

The precedence effect Ruth Y. Litovsky, H. Steven Colburn, William A. Yost, Sandra J. Guzman 1633

**GENERAL LINEAR ACOUSTICS [20]**

Errors due to sensor and position mismatch in planar acoustic holography Kyoung-Uk Nam, Yang-Hann Kim 1655

A new approach to reflection-transmission between viscoelastic half-spaces G. Caviglia, A. Morro 1666

Backscattering enhancements due to reflection of meridional leaky Rayleigh waves at the blunt truncation of a tilted solid cylinder in water: Observations and theory Karen Gipson, Philip L. Marston 1673

An acoustic model of a laminar sand bed Nicholas P. Chotiros, Dennis J. Yelton, Morris Stern 1681

On the phase speed and attenuation of an interface wave in an unconsolidated marine sediment Michael J. Buckingham 1694

Applications of waveguide-type base functions for the eigenproblems of two-dimensional cavities Yoon Young Kim, Dong Kyun Kim 1704

**NONLINEAR ACOUSTICS [25]**

Nonlinear Stoneley and Scholte waves G. D. Meegan, M. F. Hamilton, Yu. A. Il'inskii, E. A. Zabolotskaya 1712

**UNDERWATER SOUND [30]**

The stabilized self-starter Michael D. Collins 1724

(Continued)

## CONTENTS—Continued from preceding page

An efficient parabolic equation solution based on the method of undetermined coefficients	Dalcio K. Dacol, Michael D. Collins, Joseph F. Lingeitch	1727
A phase-compensated distorted wave Born approximation representation of the bistatic scattering by weakly scattering objects: Application to zooplankton	Dezhang Chu, Zhen Ye	1732
Low grazing angle bistatic sea floor scattering on the Florida Atlantic coastal shelf	Christopher M. Day, Tokuo Yamamoto	1744
Monostatic and bistatic reverberation statistics west of the Mid-Atlantic Ridge	Yevgeniy Y. Dorfman, Ira Dyer	1755
The underwater sounds produced by impacting snowflakes	Lawrence A. Crum, Hugh C. Pumphrey, Ronald A. Roy, Andrea Prosperetti	1765
<b>ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND [35]</b>		
A Gaussian finite-element method for description of sound diffraction	Dehua Huang, M. A. Breazeale	1771
Nonlinear propagation of narrow-band Rayleigh waves excited by a comb transducer	D. C. Hurley	1782
An experimental verification of Rayleigh's interpretation of the Sondhauss tube	Guadalupe Huelsz, Eduardo Ramos	1789
Thermal diffusion and mixture separation in the acoustic boundary layer	G. W. Swift, P. S. Spoor	1794
<b>NOISE: ITS EFFECTS AND CONTROL [50]</b>		
A theoretical study of duct noise control by flexible panels	Lixi Huang	1801
<b>ARCHITECTURAL ACOUSTICS [55]</b>		
Experimental investigation of the acoustical characteristics of university classrooms	Murray Hodgson	1810
On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility	J. S. Bradley, R. D. Reich, S. G. Norcross	1820
<b>ACOUSTIC SIGNAL PROCESSING [60]</b>		
Time-frequency filters for target classification	P. Chevret, N. Gache, V. Zimpfer	1829
Matched field inversion of broadband data using the freeze bath method	Lothar Jaschke, N. Ross Chapman	1838
<b>PHYSIOLOGICAL ACOUSTICS [64]</b>		
A composite model of the auditory periphery for simulating responses to complex sounds	Arnaud Robert, Jan L. Eriksson	1852
The representation of pure tones and noise in a model of cochlear nucleus neurons	Jan L. Eriksson, Arnaud Robert	1865
Solution of the inverse problem for a linear cochlear model: A tonotopic cochlear amplifier	Emilios K. Dimitriadis, Richard S. Chadwick	1880
Modeling otoacoustic emissions by active nonlinear oscillators	R. Sisto, A. Moleti	1893
<b>PSYCHOLOGICAL ACOUSTICS [66]</b>		
Detection and intensity discrimination of Gaussian-shaped tone pulses as a function of duration	Thomas Baer, Brian C. J. Moore, Brian R. Glasberg	1907
Intensity discrimination and detection of amplitude modulation	Magdalena Wojtczak, Neal F. Viemeister	1917
Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees	Astrid van Wieringen, Jan Wouters	1925
Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters	Steven van de Par, Armin Kohlrausch	1940
Azimuthal tuning of human perceptual channels for sound location	Susan E. Boehnke, Dennis P. Phillips	1948

(Continued)

CONTENTS—*Continued from preceding page*

<b>Auditory localization of nearby sources. II. Localization of a broadband source</b>	Douglas S. Brungart, Nathaniel I. Durlach, William M. Rabinowitz	1956
<b>Determination of optimal data placement for psychometric function estimation: A computer simulation</b>	Chan F. Lam, Judy R. Dubno, John H. Mills	1969
<b>Methods for estimating the sound pressure at the eardrum</b>	H. Hudde, A. Engel, A. Lodwig	1977
<b>SPEECH PRODUCTION [70]</b>		
<b>Geometry, kinematics, and acoustics of Tamil liquid consonants</b>	Shrikanth Narayanan, Dani Byrd, Abigail Kaun	1993
<b>Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results</b>	Roger W. Chan, Ingo R. Titze	2008
<b>Interarticulator phasing, locus equations, and degree of coarticulation</b>	Anders Löfqvist	2022
<b>Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words</b>	J. Sean Allen, Joanne L. Miller	2031
<b>SPEECH PERCEPTION [71]</b>		
<b>A model of auditory perception as front end for automatic speech recognition</b>	Jürgen Tchorz, Birger Kollmeier	2040
<b>Acoustical and perceptual study of gemination in Italian stops</b>	Anna Esposito, Maria Gabriella Di Benedetto	2051
<b>Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception</b>	Andrew Faulkner, Stuart Rosen	2063
<b>Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors</b>	Ann R. Bradlow, David B. Pisoni	2074
<b>Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children</b>	Ann R. Bradlow, Nina Kraus, Trent G. Nicol, Therese J. McGee, Jenna Cunningham, Steven G. Zecker, Thomas D. Carrell	2086
<b>SPEECH PROCESSING AND COMMUNICATION SYSTEMS [72]</b>		
<b>On the number of channels needed to understand speech</b>	Philipos C. Loizou, Michael Dorman, Zhemin Tu	2097
<b>BIOACOUSTICS [80]</b>		
<b>Subharmonic backscattering from ultrasound contrast agents</b>	P. M. Shankar, P. D. Krishna, V. L. Newhouse	2104
<b>ERRATA</b>		
<b>Erratum: “Monosyllabic word recognition at higher-than-normal speech and noise levels” [J. Acoust. Soc. Am. 105, 2431–2444 (1999)]</b>	Gerald A. Studebaker, Robert L. Sherbecoe, D. Michael McDaniel, Catherine Gwaltney	2111

CONTENTS—*Continued from preceding page*

## ACOUSTICS RESEARCH LETTERS ONLINE

**Full-field mapping of ultrasonic field by light-source-synchronized projection**

Gang Yao, Lihong V. Wang

L36

## CUMULATIVE AUTHOR INDEX

2113

**NOTES CONCERNING ARTICLE ABSTRACTS**

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

***Document Delivery:*** Copies of journal articles can be ordered from the new *Articles in Physics* online document delivery service (URL: <http://www.aip.org/articles.html>).



# CONTENTS

	Page
Technical Program Summary .....	A8
Schedule of Technical Session Starting Times .....	A9
Map of Greater Columbus .....	A10
Map of Downtown Columbus .....	A11
Map of Meeting Rooms .....	A12
Calendar—Technical Program .....	A14
Calendar—Other Events .....	A22
Meeting Information .....	A24
Guidelines for Presentations .....	A36
Dates of Future Meetings .....	A38
Exhibitors Program .....	A41
Technical Sessions (1a_), Monday Morning .....	2115
Technical Sessions (1p_), Monday Afternoon .....	2119
Tutorial Session (1eID), Monday Evening .....	2128
Technical Sessions (2a_), Tuesday Morning .....	2129
Technical Sessions (2p_), Tuesday Afternoon .....	2160
Technical Sessions (3a_), Wednesday Morning .....	2188
Technical Sessions (3p_), Wednesday Afternoon .....	2218
Plenary Session, Business Meeting and Awards Ceremony, Wednesday Afternoon .....	2226
Silver Medal in Biomedical Ultrasound/Bioresponse to Vibration Award Encomium .....	2226
Silver Medal in Noise Award Encomium .....	2226
Trent-Crede Medal Award Encomium .....	2226
Technical Session (3eAO), Wednesday Evening .....	2226
Technical Sessions (4a_), Thursday Morning .....	2227
Technical Sessions (4p_), Thursday Afternoon .....	2254
Technical Sessions (5a_), Friday Morning .....	2283
Technical Sessions (5p_), Friday Afternoon .....	2299
Author Index to Abstracts .....	2304
Sustaining Members .....	2311
Application Forms .....	2313
Regional Chapters .....	2316
Index to Advertisers .....	2320

# Full-field mapping of ultrasonic field by light-source-synchronized projection

Gang Yao and Lihong V. Wang\*

*Biomedical Engineering Program, Texas A&M University  
233 Zachry Engineering Center, College Station, TX 77843-3120  
gyao@tamu.edu lwang@tamu.edu*

**Abstract:** A simple method for imaging ultrasonic fields in clear media is introduced. A modulated laser source is used to project the ultrasonic field onto a CCD camera. By use of the source-synchronized lock-in detection scheme, 2D images of the amplitude and phase distributions can be determined simultaneously. This technique is experimentally demonstrated with a 1-MHz and a 3.5-MHz ultrasonic transducer operated in continuous-wave mode. This method is very straightforward to implement and can be combined with the traditional tomographic reconstruction technique to obtain the 3D distribution of an ultrasonic field.

© 1999 Acoustical Society of America

**PACS numbers:** 43.58.-e, 43.20.Ye, 43.35.Sx

## Introduction

Because of the advantage of noninvasive measurements, various optical techniques, such as the Bragg diffraction,<sup>1</sup> Schlieren imaging,<sup>2</sup> Raman-Nath diffraction,<sup>3</sup> and TV holography,<sup>4</sup> have been used to map ultrasonic fields. Most of these methods except TV holography must use some scanning mechanisms to acquire an image. Using a speckle interferometer with a video camera in TV holography, one can measure the amplitude and phase information of an ultrasonic field simultaneously. TV holography has been successfully applied to the investigation of ultrasonic fields in clear media and water.<sup>5</sup>

In this letter, we demonstrate a simple method for mapping the projection of an ultrasonic field by using a source-synchronized parallel lock-in detection scheme.<sup>6</sup> The illuminating light is modulated at the frequency of the ultrasonic field. The pressure and phase distributions of a 1-MHz and a 3.5-MHz ultrasonic transducer operated in continuous-wave (cw) mode are obtained by this method. A comparison of the experimental and theoretical results is also provided.

## Method

The experimental setup is shown in Fig. 1. The ultrasonic transducer is immersed in a 10-cm x 15-cm x 15-cm glass tank containing clear water. A 5-cm thick rubber slab is placed at the bottom of the tank to prevent reflection of ultrasound from the bottom of the tank. The laser beam from a diode laser (693 nm, Melles Griot 56IMS667) illuminates the ultrasonic field after beam expansion. The transmitted light is projected onto the CCD camera (Dalsa CA-D1-0256T) through an imaging lens. Two function generators (Stanford Research Systems, DS345) are used to excite the ultrasonic transducer and modulate the diode laser, respectively. The two function generators share the same time base to ensure synchronization. The function generators and the CCD camera are controlled by a personal computer. The ultrasonic

---

\* Correspondence should be addressed to the second author at Tel: (409) 847-9040, Fax: (409) 845-4450, Email: lwang@tamu.edu

transducer is mounted on a translation stage which is movable vertically (along the X axis). By moving the transducer along the X axis, different segments of the ultrasonic field can be imaged.

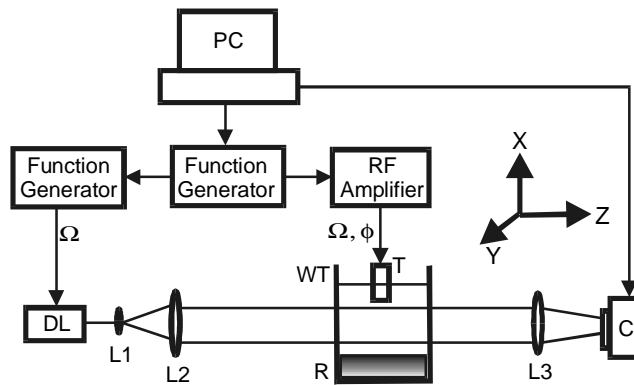


Fig. 1. Schematic sketch of the experimental setup for the measurement of ultrasonic fields in water. DL: diode laser; L1, L2, L3: imaging lens; WT: water tank; R: rubber slab; C: CCD camera; T: ultrasonic transducer.

A coordinate system is setup as shown in Fig. 1. The X axis is on the acoustic axis pointing to the ultrasonic transducer. The Y axis is perpendicular to both the acoustic and optical axes. The Z axis is along the optical axis pointing to the CCD camera.

The CCD camera is positioned before the focal plane of the imaging lens. In other words, we are working in the near-field region. In the Raman-Nath acousto-optic interaction regime, the near-field light intensity diffracted by an ultrasonic wave with frequency  $\Omega$  is:<sup>7,8</sup>

$$I(x, y, Z, t) = CI_i \left( 1 + 2 \sum_{n=1}^{+\infty} J_n [2v_{x,y} \sin(2n\pi Z)] \times \cos[n(\Omega t + \Phi_{x,y} + \varphi)] \right), \quad (1)$$

where  $I_i$  is the incident laser intensity,  $C$  is a constant,  $J_n$  is the  $n$ th order Bessel function of the first kind,  $v_{x,y}$  and  $\Phi_{x,y}$  are respectively the effective Raman-Nath parameter and the effective phase of the ultrasonic field along the projection path,<sup>5</sup>  $\varphi$  is the initial phase of the ultrasonic wave, and  $Z$  is a normalized parameter defined as:<sup>8</sup>

$$Z = (z - L) \times \lambda / (2n_0 \Lambda^2), \quad (2)$$

where  $L$  is the interaction length between the ultrasonic field and the light,  $\lambda$  and  $\Lambda$  are respectively the optical wavelength and the acoustic wavelength,  $n_0$  is the refractive index of the medium. The effective pressure  $P$  can be derived from the Raman-Nath parameter as:

$$P_{x,y} = \frac{v_{x,y} \lambda}{2\pi L (\partial n / \partial p)}, \quad (3)$$

where  $\partial n / \partial p$  is the piezo-optical coefficient of the medium.

The incident laser  $I_i$  intensity is also modulated at the frequency  $\Omega$ :  $I_i = I_0 + I_0 \cos(\Omega t)$ . The CCD camera acts as a low-pass filter because of its low running frequency relative to  $\Omega$ . Considering  $J_1(x) \cong x/2$  at small  $x$ , the signal detected by the CCD camera can be written as:

$$S_\varphi(x, y) = \langle I(x, y, Z, t) \rangle_t = CI_0 + C_1 I_0 v_{x,y} \cos(\Phi_{x,y} + \varphi), \quad (4)$$

where  $\langle I(x,y,Z,t) \rangle_t$  is the time average of the received light intensity in the exposure period of the CCD camera and  $C_1 = C \sin(2\pi Z)$ . In the experiment, the phase of the signal applied to the ultrasonic transducer is set sequentially to  $0^\circ, 90^\circ, 180^\circ, 270^\circ$  relative to the signal applied to the diode laser. The corresponding four frames of CCD images are acquired to calculate the amplitude and phase distributions of the ultrasonic field. The pressure-related Raman-Nath parameter  $v_{x,y}$  and the phase  $\Phi_{x,y}$  can be obtained as:

$$v_{x,y} = \frac{1}{2C_1 I_0} \sqrt{(S_{0^\circ} - S_{180^\circ})^2 + (S_{90^\circ} - S_{270^\circ})^2} \quad (5)$$

$$\Phi_{x,y} = -\tan^{-1} \frac{S_{90^\circ} - S_{270^\circ}}{S_{0^\circ} - S_{180^\circ}} \quad (6)$$

The Raman-Nath parameter can be converted to effective pressure by Eq. (3) if the piezooptical coefficient of the medium is known.

### Results and Discussion

In the experiment, a ruler placed on the acoustic axis is imaged by the CCD camera to calibrate the magnification of the system. Figure 2 shows the ultrasonic field of a 1-MHz transducer (Panametrics V314-1.0/0.75) operated in cw mode with a focal length of 2.54 cm. The gray-scale images are obtained by combining three images corresponding to different ultrasonic-field segments. In the amplitude image, white color refers to the higher intensity and black color the lower intensity. In the phase image, the phase is wrapped between  $[-\pi, +\pi]$ .

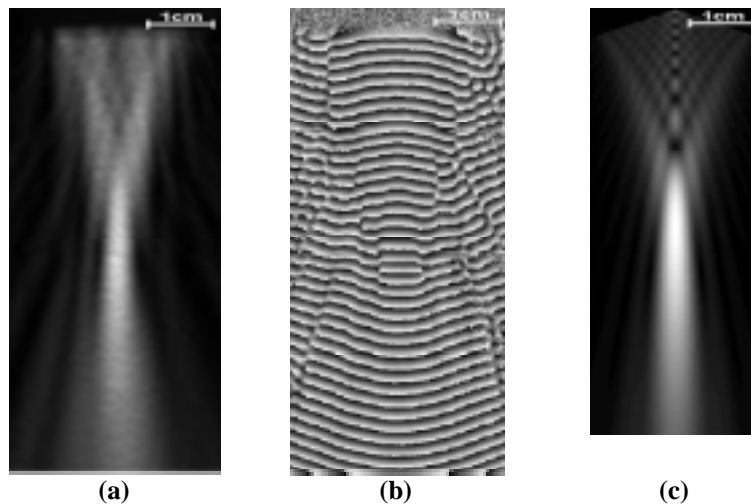


Fig. 2. 2D projection images of a 1-MHz ultrasonic field. (a) Experimental amplitude image; (b) experimental phase image; (c) theoretical amplitude distribution.

Figure 2(a) and 2(b) show the experimentally measured amplitude and phase distributions, respectively. Figure 2(c) shows the amplitude distribution calculated by the following equation:

$$p(\mathbf{r}, t) = C \int_s \frac{p(s) \exp[j(\Omega t - \phi)]}{d} ds \quad (7)$$

where  $p(\mathbf{r}, t)$  is the pressure as a function of the spatial position  $\mathbf{r}$  and time  $t$ ,  $C'$  is a constant,  $p(s)$  is the pressure distribution on the surface of the ultrasonic transducer,  $\Omega$  is the angular frequency of the ultrasonic wave,  $\phi$  is the phase delay between a point on the transducer surface and the observation point, and  $d$  is the distance between a point on the transducer surface and the observation point. The integration is over the surface of the transducer.

As shown in Fig. 2, the experimental result and the theoretically modeled result agree with each other very well in the far field. In the near field, the difference is caused by the projection nature of the experiment. In the projection measurement, the measured amplitude corresponds to an effective pressure along the optic projection line in the ultrasonic field. More accurate results can be obtained by reconstruction<sup>5</sup> as in the traditional tomographic measurements. Figure 3(a) shows the 1D amplitude profiles along the horizontal direction ( $Y$  axis in Fig. 1) at the focus point (2.54 cm) obtained experimentally and theoretically. The measured FWHM (full width at half maximum) of  $\sim 2.1$  mm is very close to the theoretical value of 2.2 mm. Figure 3(b) shows a plot of the phase along the acoustic axis of the transducer ( $X$  axis in Fig. 1). The measured period of the phase is  $\sim 1.43$  mm, which is close to the wavelength of 1-MHz ultrasonic wave in water (1.48 mm).

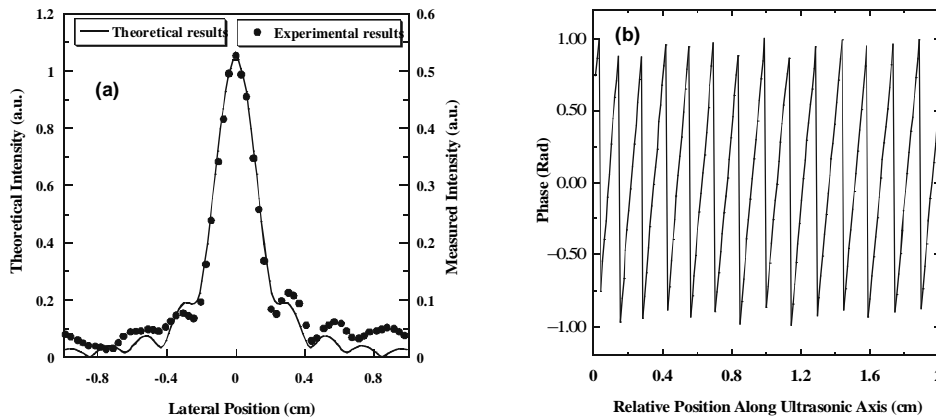


Fig. 3. 1D plots of the 1-MHz ultrasonic field. (a) Amplitude plot along the  $Y$  axis; (b) phase plot along the ultrasonic axis.

This method can also be applied to mapping ultrasonic fields of higher frequencies. The amplitude and phase distributions of a 3.5-MHz ultrasonic transducer are shown in Fig. 4 to demonstrate this point.

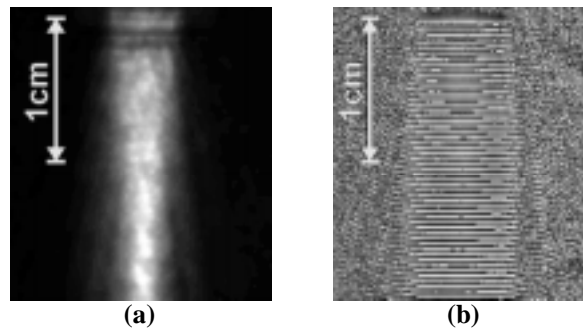


Fig. 4. 2D projection images of a 3.5-MHz ultrasonic field. (a) Amplitude image; (b) phase image.

## Conclusion

A simple method for investigating the ultrasonic field has been introduced in this letter. This method is based on a source-synchronized parallel lock-in detection scheme. The illumination light is modulated at the same frequency as the ultrasonic transducer and detected by a CCD camera after passing through the ultrasonic field. The amplitude and phase information can be measured simultaneously. This method can also be extended for tomographic measurements by rotating the transducer, which will furnish more accurate information. As with all other projection measurements, the application of this method is limited to weak acousto-optic interaction in which ray bending is negligible. Further theoretical studies as well as more accurate experimental calibration will be helpful for quantitative studies of ultrasonic fields by this method.

## Acknowledgments

This project was sponsored in part by the National Institutes of Health grants R29 CA68562, R01 CA71980 and R21 CA83760 and by the National Science Foundation grant BES-9734491.

## References and links:

- <sup>1</sup> A. Korpel, L. W. Kessler, and M. Ahmed, "Bragg diffraction sampling of a sound field," *J. Acoust. Soc. Am.* **51**, 1582-1592 (1972).
- <sup>2</sup> B. R. Barnes and C. J. Burton, "Visual methods for studying ultrasonic phenomena," *J. Appl. Phys.* **20**, 286-294 (1949).
- <sup>3</sup> R. Reibold and W. Molkenstruck, "Light diffraction tomography applied to the investigation of ultrasonic fields. I: continuous waves," *Acoustica* **56**, 180-192 (1984).
- <sup>4</sup> S. Ellingsrud and G. O. Rosvold, "Analysis of a data-based TV-holography system used to measure small vibration amplitudes," *J. Opt. Soc. Am. A* **9**, 237-251 (1992).
- <sup>5</sup> R. Rustad, "Acoustic field of a medical ultrasound probe operated in continuous-wave mode investigated by TV holography," *Appl. Opt.* **37**, 7368-7377 (1998).
- <sup>6</sup> P. Gleyzes, F. Guernet and A. C. Boccara, "Picometric profilometry. II. Multidetector approach and multiplexed lock-in detection," *J. Opt.* **26**, 251-265 (1995).
- <sup>7</sup> B. D. Cook, "Measurement from the optical nearfield of an ultrasonically produced phase grating," *J. Acoust. Soc. Am.*, **60**, 95-99 (1976).
- <sup>8</sup> E. Blomme and O. Leroy, "Plane-wave analysis of the near field of light diffracted by ultrasound," *J. Acoust. Soc. Am.*, **91**, 1474-1483 (1992).

## Current criteria for selection of articles for publication

Allan D. Pierce

Acoustical Society of America, Office of the Editor-in-Chief, P.O. Box 323, East Sandwich, Massachusetts 02537

(Received 3 August 1999; accepted for publication 4 August 1999)

Principal publication selection criteria currently used by the Editors of the *Journal* are reviewed and discussed. Emphasis is given to those criteria that may be less well known and which could be controversial. One example is the criterion that research papers should represent a significant advance in our understanding of acoustics. © 1999 Acoustical Society of America.

[S0001-4966(99)05410-7]

PACS numbers: 43.10.Gi [ADP]

### INTRODUCTION

In recent years, the instructions that Daniel Martin sent to associate editors of this journal included the following sentences:

“The *Journal's* continuing growth has made it necessary for us to be more selective in article acceptance. There *must* be significant extension of previous acoustical knowledge. One “just noticeable difference” is not enough. Articles must be acoustical, or have a clearly defined relationship to...acoustics. If your initial judgment is that the manuscript obviously does not meet these criteria, you may save reviewer time and effort by returning the paper to the author with thanks but without review.”

The purposes of the present editorial are to affirm the present Editor-in-Chief's concurrence with the overall spirit of these remarks, to elaborate on them in some depth, and to give an open airing of the criteria that the *Journal* follows in selecting articles for publication.

In regard to the open airing, the rationale for so doing is that readers of this journal may naturally ask as to what criteria have been used in the selection of the articles that are published. Similarly, prospective authors of research papers may raise questions such as:

1. How does one choose research topics, the reporting of which would be appropriate for publication in a journal such as JASA?
2. Given a completed piece of research, what journal or other publication outlet would be most appropriate for the initial submission for publication?
3. Given that the *Journal of the Acoustical Society of America* is one of the contemplated answers to the question above, how does one present the results so that there is a high likelihood of the written work being selected for publication?
4. Given the large expenditure of time and effort required to write a research article, what is the likelihood that a suitably directed effort will eventually result in the work being selected for publication in JASA.

Analogous questions could be asked by prospective authors of review articles, by prospective authors of tutorial articles,

and by prospective authors of papers that discuss practical applications of acoustics. The discussion here, however, is limited to research articles.

In all such cases, the answers to such questions strongly depend on the criteria that the editors use in article selection and on the success or rigor with which the editors apply such criteria. For the most part, these criteria, if in existence, are rarely openly publicized, and good arguments can be given as to why they should not be so publicized. One may argue, for example, that the criteria are invariably subjective and that the actual selection, as borne out by what appears in print, contains many examples that seem to be strongly at variance with the criteria. Also, there is a possible fear that the airing of the criteria will discourage many prospective authors whose work would be at least as worthy of selection as a substantial fraction of the articles that have been published in the past. The present writer has considered such issues, and has concluded that the potential good of an open discussion of selection criteria outweighs the potential evil.

### I. HOW MUCH SHOULD JASA PUBLISH?

The *Journal of the Acoustical Society of America* does not publish all of the world's literature in acoustics, and it never will. In the most recent calendar year (1998), JASA published 718 papers, including letters to the editor and technical notes, but excluding comments and responses to papers previously published. To put this number of 718 in perspective, one should ask what fraction does it represent of the world's literature in acoustics. A rough indication of such can be determined from the total number of items that are listed in the *Journal's* annual listing of contemporary papers on acoustics. This is given in part 2 of the August 1998 issue. There are 605 pages that list such references, and approximately 29 references per page, yielding an approximate total of 17,500 references. Thus JASA publishes approximately 4% of the world's literature in acoustics. Most readers will nevertheless agree that JASA publishes a much larger fraction of the world's *significant* papers on acoustics, although they may possibly not agree on just what constitutes a significant paper.

A more relevant question is just how many papers should JASA be publishing per year. It clearly should pub-

lish no more than the overall income possibilities (membership dues, nonmember subscriptions, and subsidies) allow. It also should publish enough to achieve a creditable accomplishing of the Acoustical Society's stated purpose "to increase and diffuse the knowledge of acoustics and to promote its practical applications." Another reasonable objective in the present age is that the *Journal* should be an instrument for encouraging (1) the carrying out of significant research in acoustics, (2) the lucid and concise writing of papers that effectively convey significant new ideas and new research accomplishments, (3) an orderly and scholarly archiving of work in acoustics, and (4) innovation and professional excellence in the practical applications of acoustics. The readers and the members will ideally desire that the size of the *Journal* be constrained by a conscientious attempt to publish only those papers that will be read and whose content will be used in either future research or in applications. The latter arguments strongly dictate that the *Journal* should not publish everything that is submitted to it, even if the economic circumstances would allow it. If it did publish so liberally, then insignificant research would be encouraged, sloppy and wordy writing would pervade, and the archival edifice would crumble.

As for a reasonable constraint on the numbers of papers to be published, the present writer suggests the simple rule that the number of published papers should be roughly a constant times the number of members of the Acoustical Society of America. Long term data indicate that this criterion has been tacitly met over the years, the rate being roughly one paper every ten years per member. For example in 1960, the number of papers published was 250 and the number of members was 2795, yielding 0.089 papers per member. In 1998, the corresponding numbers were 718 and 7253, yielding 0.099 papers per member. (The number of published pages per member is a slightly different story, as the number of pages per article has increased substantially in more recent years, suggesting that the editors may have been slack in the encouragement of concise writing.)

The rule just stated is not hard-and-fast, and we should not slavishly seek to keep to it. If we did deviate substantially from the rule, however, then we should understand why and be able to convince ourselves that the deviation was a good thing.

## II. THE CLEARLY ACOUSTICAL, NOT SOMETHING ELSE, CRITERION

With a limited size of the *Journal*, and given that some selection is desirable, an obvious criterion is that the subject matter should be acoustics, rather than something else. Different people may have different definitions of acoustics; a practical definition, insofar as the *Journal* is concerned, is that acoustics includes whatever has been regularly discussed within its pages and whatever has been regularly discussed within the technical sessions at the Society's semi-annual meetings. This goes far beyond the textbook definition of acoustics as the science of sound, but a gray area exists for papers which are partly or peripherally connected with acoustics. In other documents that this writer has seen, Daniel Martin, the recent Editor-in-Chief, requested that as-

sociate editors not accept papers unless they were clearly acoustical, and the present writer concurs with this.

A number of examples are paraphrased here to help authors recognize when their papers may not meet the *clearly acoustical* criterion. (Some of these examples crystallized during readings of various unpublished writings of Daniel Martin.)

1. *Useful mathematical techniques.* Much acoustical research makes use of mathematics, and the researchers often have a need for more useful mathematical techniques. However, a paper whose sole purpose is to describe such a technique, especially if there is no mention of actual acoustical topics, should not be published in the *Journal*. A more appropriate place for publication would be in a journal on applied mathematics.

2. *Generic signal processing.* Authors might submit a paper in which a signal processing procedure is described which applies to many different types of fields and signals. The procedure draws on no principles which are specific to acoustic fields, and the paper has no examples or applications which are unambiguously acoustical. Here a more appropriate place for publication would be a journal on signal processing.

3. *Solution of generic boundary-value problems.* Certain partial differential equations, such as the wave equation and the Helmholtz equation, pertain to many branches of physics, and are extensively discussed in standard textbooks on applied mathematics. A paper which discusses and solves (presumably approximately) a boundary-value problem involving such a partial differential equation without a strong reference to an acoustical context would also be more appropriate to a journal on applied mathematics, or perhaps to a journal on numerical methods.

4. *Routine use of acoustical instrumentation.* Authors might submit a paper in which off-the-shelf acoustical instrumentation is being used in a standard manner in research where the issues and investigated topics are not acoustical. The subject of the research may be important and the results may be significant, but the paper is not clearly acoustical.

5. *Work more strongly identified with another discipline.* Many research topics in acoustics are interdisciplinary, and some subtasks are carried out where the issues and techniques have very little to do per se with the science of sound. A pertinent question is how much prior knowledge of acoustics and related acoustical research did the authors have to actually use in carrying out their study. An example where the clearly acoustical criterion would likely not be met would be the study of the grammatical structure of a spoken language; another would be the design of towing hardware for underwater acoustic arrays; yet another would be the modification of musical instruments to make them more convenient to play.

## III. THE SIGNIFICANT ADVANCE CRITERION

Daniel Martin's statement given at the beginning of this article suggests that in previous years the significant advance criterion either may not have existed or may not have been universally recognized. What is probably the case is that whether or not the criterion was applied depended critically



on the reviewers and the editor. What Daniel Martin was trying to achieve and what the present editorial is also trying to achieve is to make this criterion more explicit.

One may argue that significance is hard to measure, and the present writer concurs with this. Few of us who were around at the times of the birth of the digital computer, the laser, the transistor, and the integrated circuit could have fully appreciated the significance of these inventions. Nevertheless, there are many articles submitted and sometimes published in the *Journal* whose lack of substantial significance would be clearly recognized by most workers in the field. It is here that the editors and the reviewers should draw the line.

At the risk (or perhaps with the hope) of stimulating some controversy, the present writer lists below some examples of categories of papers that would possibly be viewed as lacking substantial significance:

- (1) *New derivations of known results.* An author may discover a new way of deriving a known mathematical result in acoustics and decide that it is preferable to the derivation that was originally published in the literature. Unless the initial assumptions are less restrictive or the original derivation was flawed in its logic, the new derivation is more appropriate to a textbook or to a review article, but not to an article that purports to be reporting original research.
- (2) *New computational algorithms.* In the present era, much of the published scientific literature is computational in nature. The old standard division of research into two parts (experiment and theory) has been replaced by a division into three parts, the third being computation. If prior literature adequately shows how one can calculate a result, then the publication of a new algorithm for doing the same calculation *may* be of questionable significance. However, if the authors can demonstrate that the new algorithm will result in a significantly more efficient and faster computation, thereby enabling researchers and practitioners to do such computations on a more routine basis, then the substantial significance criterion is met. [An example that comes to mind is the fast-Fourier-transform (FFT) algorithm, which had enormous influence on the practical spectrum analysis of signals.]
- (3) *New computational results.* The existence of the computer and of generic software has made it possible for researchers to churn out computational results for categories of acoustical problems that may be of widespread interest. Nevertheless, if the techniques involved in doing such calculations are already described in the literature, the publication of new computational results for cases similar to those that have already been dealt with in the prior publications would not normally be regarded as significant. What might be significant are results that tend to defy one's nominal intuitive expectations. Experimentation via numerical calculation has frequently led to the discovery of novel phenomena which in turn has stimulated laboratory or field experiments and analytical studies.
- (4) *Confirmation of accepted theory.* A portion of acoustics

has reached a maturity where there is a substantial body of theory which has been well-confirmed by experiment and which is accepted without much reservation by the workers in the field. A new paper reporting new experiments for a case that is covered by such theory and which reaffirms that the theory adequately agrees with the data would hold no surprises, and one might well argue that the significance of the paper is below the threshold for publication in JASA.

- (5) *Slight variations on previously reported experiments.* A new paper is submitted reporting experiments which differ only slightly in terms of circumstances, parameters, or subjects from an experiment previously reported in the literature. Unless the results of the new paper are at considerable variance with the results of the previously reported paper, it would be argued that the substantial significance criterion is not met. Perhaps the paper should be published somewhere, if only to archive the results, but it should not be published in JASA.
- (6) *New wrinkles on old theory or on old simulations.* Existing theory and simulations based on such theory invariably are for restricted or simplified cases. A new paper which extends or modifies the theory to remove the restrictions risks being of small significance. The authors would have to demonstrate that the modifications make a substantial difference on the predictions for cases that are of common interest. If the modifications do make a large difference for some circumstances, then the authors would have to effectively argue that such circumstances are of existing or potential interest to the acoustical community. Simply confirming that the modifications make little practical difference on the results would not meet the substantial significance criterion.
- (7) *Routine solution of new problems.* Not every conceivable problem in acoustics has been solved, and many which have been solved have not been reported in the literature. However, if a substantial fraction of the acoustical community could solve such a problem in a routine fashion without having to exercise any unusual ingenuity, then the substantial significance criterion would ordinarily not be met. An exception would be if the results were surprisingly at variance with what one's intuition would expect. An author might argue that the publication of the solution would possibly save future researchers the effort of redoing it. However, in many such cases, the existence of the published solution may be sufficiently hard to find or the write-up may be such that the details are sufficiently hard to follow that subsequent workers may find it more convenient simply to work it out for themselves from scratch. A relevant question is that, if a competent acoustician were to need the result in a hurry for a specific application, would such a person seek to find the result in the literature or to work it out directly. (The latter might possibly be more fun and more professionally rewarding.)
- (8) *Unrealistic problems.* Occasionally one encounters a paper in which the authors develop a new theory or present computational results for circumstances that are not known to exist in nature or in practice. The paper might

begin, for example, with a postulated model of a material that corresponds to no existing material, and then proceeds to derive acoustical consequences that are appropriate to that hypothetical material. Possibly, the model is being introduced with the thought of its being tested by some existing or imminent experiment or with the thought of its being realized by an experimental fabrication in the laboratory. Alternately, the model might be considered because it captures salient features of an actual material for which the theoretical analysis is prohibitively difficult. If the authors can give effective arguments along such lines, then the substantially significant criterion might be met. However, if the paper only asks "what if" and then reports the results, without giving any realistic reason as to why the question should have been asked, then the criterion is not met.

- (9) *Results of severely limited generality.* A paper might deal with a set of circumstances that are so specific that it is unlikely that they would be encountered by others in applications of interest. Unless the research has been carried out so that the results either apply to or are generalizable to a broad range of circumstances, encompassing those of likely interest to others, the significance of the paper would be slight.

#### IV. THE ORIGINALITY CRITERION

For the most part, other criteria that the *Journal* applies in the selection of manuscripts need little discussion. Problems with writing, organization, conciseness, and completeness can usually be rectified by authors after such problems have been pointed out to them (but papers having excessive problems may be returned directly to the authors without much comment). However, there is one other criterion that can lead to misinterpretation and to controversy, this being the criterion of originality. The *Journal* will not publish any paper which is identical to or substantially the same as a paper that has been previously published. Authors sometimes will submit papers giving results which they believe are genuinely original. The work may be so out of the mainstream of current research that the editors and the reviewers cannot affirm without considerable effort that substantially similar work has not been done and reported *at some time* in the archival literature. One cannot expect the editors and the reviewers to do an extensive literature search, especially if the work is of such a classical nature that it could have easily been done by Lord Rayleigh or by one of his contemporaries. The criterion that is tacitly applied in such cases is that the authors should tangibly demonstrate by way of the discussion and of the citing of related literature that they have done a conscientious and thorough search of all likely sources. A paper which cites only standard textbooks and a few peripheral references but which purports to present original work of a classical nature is automatically suspect, and the editor may appropriately send the manuscript back to the author stating that the assertion of originality is insufficiently supported by the authors' manuscript.

Special problems arise when the work has been previously reported in whole or in part by the authors themselves.

Duplicate publication of material that has been published in another archival journal is clearly not appropriate. Originality insofar as journal publication is concerned is not lost if the author has sent out preprints of the paper to colleagues, or if the paper has been issued as an internal report, or if the preprint has been posted on an author's personal World Wide Web site. Publication of a lesser archival nature such as in conference proceedings is a gray area, and the present writer is reluctant to suggest hard and fast rules for such cases. Frequently, authors present "works in progress" at conferences whose organizers require written papers for the published proceedings of the conference. However, the proceedings may have very little circulation and may be inaccessible to the broader acoustical research community. If the work is rewritten with additional results and a greatly improved presentation of those results, and with a fresh title and abstract, the judgment of the editors may sometimes be that the criterion of originality is met. In all such cases, however, it is expected that the authors, at the time of submission, will give a full disclosure of the existence of prior publication and of how the present submission differs from what has been published before. Not doing so would be viewed as unethical.

#### V. CONCLUDING REMARKS

Good research and the writing of good articles that report one's research is intrinsically difficult, but both of these endeavors can be satisfying ways of spending one's time. There are many possible intangible rewards, but perhaps the one that we treasure the most is that of achieving excellence. One wants to do excellent research and one wants to write excellent papers. The *Journal*, with the criteria stated above, should be viewed as a stimulus and as an instrument for the achieving of such excellence.

The *Journal of the Acoustical Society of America* has long been the world's largest and most prestigious journal for acoustics research. There are nevertheless journals associated with other fields which many might argue are more prestigious than is JASA. Perhaps this is not so, but it should be a matter of concern to those who regard acoustics as a major part of their professional life. It is natural for all of us who work in acoustics to take some pride in our profession. We want *our* journal to be excellent and to be perceived as such by others outside our discipline as excellent. We want those who are stimulated to write for JASA to continue to be so stimulated and we do not want them to be discouraged by the *Journal's* aspirations for a higher degree of excellence. Rather, we encourage them to join in what is in reality a large common endeavor with each of us prodding, pulling, and cheering the others along. We want others of kindred spirit to join us, we want them to also write excellent papers, and we want them to submit such papers to *our* journal. And of course, with the *Journal* as an integral part of our common enterprise, we want to pursue, with an ever-increasing commitment to excellence, the ASA's enduring purpose: *to increase and diffuse the knowledge of acoustics and to promote its practical applications.*

# ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer

Physics Department, Georgetown University, Washington, DC 20057

## Silver Jubilee—India

The year 1999 marks the Silver Jubilee of the Ultrasonics Society of India. In order to celebrate the anniversary the society, in cooperation with the National Physical Laboratory, is organizing the Third International Conference and Exhibition on Ultrasonics (ICEU-99) to be held in New Delhi 2–4 December 1999. This year also marks the birth centenary of the founder and first director of the National Physical Laboratory, K. S. Krishnan.

The previous conferences were held 1980 and 1990, and like these conferences, the 1999 meeting will provide a forum for scientists and technologists working in the area of ultrasonics.

Further information about ICEU-99 is available from Janardan Singh, Head of the Ultrasonics Group, NPL, Dr. K. S. Krishnan Road, New Delhi-110 012, India.

## Papers published in JASJ(E)

A listing of Invited Papers and Regular Papers appearing in the latest issue of the English language version of the *Journal of the Acoustical Society of Japan*, JASJ(E), was published for the first time in the January 1995 issue of the Journal. This listing is continued below.

The August issue of JASJ(E), Vol. 20, No. 4 (1999) contains the following contributions:

R. Makarewicz, J. Koga, and S. Tonda, "Ground effects on railway bridge noise"

E. B. Skrodzka, "An influence of an enclosure on modal behavior of loudspeakers"

I. Kinoshita and S. Aoki, "Representation of sound localization transfer function and psychoacoustical evaluation"

K. P. Markov and S. Nakagawa, "Integrating pitch and LPC-residual information with LPC-cepstrum for text-independent speaker recognition"

S. Emura and M. Miyoshi, "Precise estimation of multiple transmission paths in acoustic system"

Y. Shimada, Y. Nishimura, T. Usagawa, and M. Ebeta, "Active control for periodic noise with variable fundamental extended DXHS algorithm with frequency tracking ability"

P.-K. Choi, T. Ezawa, and G. Kudoh, "Acoustic nonlinearity parameter measurement using optical beam deflection by ultrasonic waves"

## International Meetings Calendar

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

### October 1999

9–10 **INCE/Japan Annual Meeting**, Tokyo. (Fax: +81 42 327 3847) 8/99

18–19 **Acoustics Week in Canada**, Victoria. (Fax: +1 250 472 4100; e-mail: sdosso@uvic.ca) 8/99

20–22 **Iberian Meeting of the Spanish and the Portuguese Acoustical Societies**, Avila. (Fax: +34 91 411 7651; e-mail: ssantiago@fresno.csic.es) 12/98

22–24 **\*1999 Auditorium Acoustics Conference**, Manchester, UK. (Institute of Acoustics, 77A St. Peter's Street, St. Albans, Herts AL1 3BN, UK; Fax: +44 1727 850 553; e-mail: ioa@ioa.org.uk)

28–29 **Swiss Acoustical Society Fall Meeting**, Biel. (Fax: +41 1 823 4793; e-mail: beat.hohmann@compuserve.com) 6/99

### November 1999

17–18 **Institute of Acoustics (UK) Autumn Conference on**

18–21

24–26

### December 1999

2–4

15–17

### February 2000

17–18

### March 2000

15–17

19–22

20–24

### April 2000

3–4

### May 2000

17–19

23–26

24–26

### June 2000

5–9

6–9

**Environmental Noise**, Stratford upon Avon. (Fax: +44 1727 850 553; e-mail: acoustics@clus1.ulcc.ac.uk) 8/99

**Reproduced Sound 15**, Stratford upon Avon. (Fax: +44 1727 850 553; e-mail: acoustics@clus1.ulcc.ac.uk) 8/99

**Australian Acoustical Society Conference**, Melbourne. (Fax: +61 3 9720 6952; e-mail: acousticsdes@bigpond.com) 6/99

**\*International Conference and Exhibition on Ultrasonics**, New Delhi, India. (J. Singh, Ultrasonics Group, National Physical Laboratory, New Delhi-110 012, India; Fax: +91 111 575 2678; e-mail: iceau99@csnpl.ren.nic.in)

**\*International Conference on Stochastic Volume and Surface Scattering**, Cambridge, UK. (Institute of Acoustics, 77A St. Peter's Street, St. Albans, Herts AL1 3BN, UK; Fax: +44 1727 850 553; e-mail: ioa@ioa.org.uk)

**\*Measuring Noise Outdoors**, Home Counties Venue, UK. (Institute of Acoustics, 77A St. Peter's Street, St. Albans, Herts AL1 3BN, UK; Fax: +44 1727 850 553; e-mail: ioa@ioa.org.uk)

**Acoustical Society of Japan Spring Meeting**, Tokyo. (Fax: +81 3 3379 1456; e-mail: kym05145@nifty.ne.jp) 8/99

**\*25th International Acoustical Imaging Symposium**, Bristol, UK. (25th IAIS, Medical Physics and Bioengineering Department, Bristol General Hospital, Bristol BS1 6SY, UK; Web: www.bris.ac.uk/depts/medphys)

**Meeting of the German Acoustical Society (DAGA)**, Oldenburg. (Fax: +49 441 798 3698; e-mail: dega@aku.physik.uni-oldenburg.de) 10/98

**Structural Acoustics 2000**, Zakopane, Poland. (Fax: +48 12 423 3163; Web: www.cyf-kr.edu.pl/ghpanusz) 8/99

**9th International Meeting on Low Frequency Noise and Vibration**, Aalborg. (Fax: +44 1277 223 453) 6/99

**\*Meeting of the Russian Acoustical Society**, Moscow, Russia. (N. N. Andrejev Acoustical Institute, 4 Shvernika ul., Moscow 117036, Russia; Fax: +7 095 126 8411; e-mail: ras@akin.ru)

**\*Joint International Symposium on Noise Control & Acoustics for Educational Buildings (24–25 May) and 5th Turkish National Congress on Acoustics (25–26 May)**, Istanbul, Turkey. (Turkish Acoustical Society YTU Mim. Fak., 80750 Besiktas-Istanbul, Turkey; Fax: +90 212 261 0549; Web: www.takder.org)

**International Conference on Acoustics, Speech and Signal Processing (ICASSP-2000)**, Istanbul. (Fax: +1 410 455 3969; Web: icassp2000.sdsu.edu) 6/99

**5th International Symposium on Transport Noise and Vibration**, St. Petersburg. (Fax: +7 812 127 9323; e-mail: noise@mail.rcom.ru) 6/99

- 14–17      **\*IUTAM Symposium on Mechanical Waves for Composite Structures Characterization**, Chania, Crete, Greece. (IUTAM 2000, Applied Mechanics Laboratory, Technical University of Crete, Chania 73100, Greece; Fax: +30 821 37438; Web: www.tuc.gr/iutam)
- July 2000**  
4–7      **7th International Congress on Sound and Vibration**, Garmisch-Partenkirchen. (Fax: +49 531 295 2320; Web: www.iiav.org/icsv7.html) 12/98
- August 2000**  
28–30      **INTER-NOISE 2000**, Nice. (Fax: +33 1 47 88 90 60; Web: www.inrets.fr/services/manif) 6/99  
31–2      **International Conference on Noise & Vibration Pre-Design and Characterization Using Energy Methods (NOVEM)**, Lyon. (Fax: +33 4 72 43 87 12; Web: www.insa-lyon.fr/laboratoires/lva.html) 6/99
- September 2000**  
3–6      **5th French Congress on Acoustics—Joint Meeting of the Swiss and French Acoustical Societies**, Lausanne. (Fax: +41 216 93 26 73) 4/99  
17–21      **Acoustical Society of Lithuania 1st International Conference**, Vilnius. (Fax: +370 2 223 451; e-mail: daumantas.ciblys@ff.vu.lt) 8/99
- October 2000**  
3–5      **WESTPRAC VII**, Kumamoto. (Fax: +81 96 342 3630; Web: cogni.eecs.kumamoto-u.ac.jp/others/westprac7) 6/98
- 3–6      **\*EUROMECH Colloquium on Elastic Waves in Nondestructive Testing**, Prague, Czech Republic. (Z. Převorovský, Institute of Thermomechanics, Dolejškova 4, 182 00 Prague 8, Czech Republic; Fax: +420 2 858 4695; e-mail: ok@bivoj.it.cas.cz)  
16–18      **2nd Iberoamerican Congress on Acoustics, 31st National Meeting of the Spanish Acoustical Society, and EAA Symposium**, Madrid. (Fax: +34 91 411 7651; e-mail: ssantiago@fresno.csic.es) 12/98  
16–20      **6th International Conference on Spoken Language Processing**, Beijing. (Fax: +86 10 6256 9079; e-mail: mchu@plum.ioa.ac.cn) 10/98
- August 2001**  
28–30      **INTER-NOISE 2001**, The Hague. (Web: internoise2001.tudelft.nl) 6/99
- September 2001**  
2–7      **17th International Congress on Acoustics (ICA)**, Rome. (Fax: +39 6 4424 0183; Web: www.uniroma1.it/energ/ica.html) 10/98  
10–13      **\*International Symposium on Musical Acoustics (ISMA 2001)**, Perugia, Italy. (Perugia Classico, Comune di Perugia, Via Eburnea 9, 06100 Perugia, Italy; Fax: +39 75 577 2255; e-mail: perusia@classico.it)
- October 2001**  
17–19      **\*32nd Meeting of the Spanish Acoustical Society**, La Rioja, Spain. (Serrano 144, Madrid 28006, Spain; Fax: +34 91 411 76 51; Web: www.ia.csic.es/sea/index.html)



# REVIEWS OF ACOUSTICAL PATENTS

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

## Reviewers for this issue:

KEVIN P. SHEPHERD, M.S. 463, NASA Langley Research Center, Hampton, Virginia 23681  
WILLIAM THOMPSON, JR., The Pennsylvania State University, University Park, Pennsylvania 16802  
ERIC E. UNGAR, Acentech Incorporated, Cambridge, Massachusetts 02138  
ROBERT C. WAAG, University of Rochester, Rochester, New York 14642

5,719,823

### 43.25.Lj GROUND PENETRATING SONAR

R. L. Earp, assignor to Lucent Technologies, Inc.  
17 February 1998 (Class 367/92); filed 14 February 1997

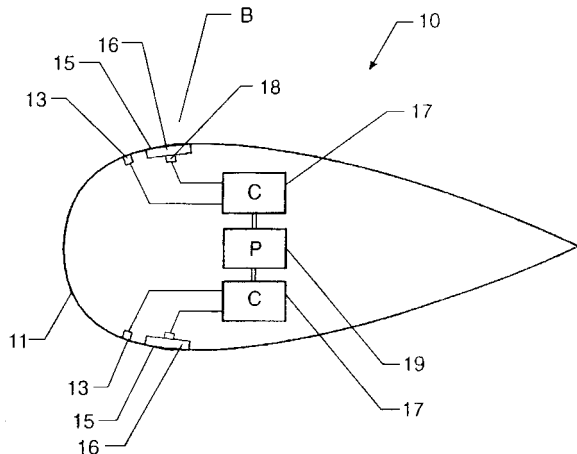
A ground-penetrating acoustic echo sounding device, for locating underground structures such as pipes, consists of a transducer which is excited at two frequencies that are generated by modulating some carrier frequency signal, e.g., 100 kHz, with a 5 kHz signal so as to create both 95 kHz and 105 kHz signals. These latter two signals then excite the transducer, and the resultant high amplitude sound waves interact parametrically in a nonlinear waveguide made of silastic rubber, whose length is about one wavelength at these frequencies. This creates a low frequency, i.e., 10 kHz, narrow beam of energy that can be directed into the ground. This low frequency beam suffers much less attenuation while propagating through the medium than would the original carrier or the modulated signals and, of course, much less attenuation than the 200 kHz sum frequency signal which is also generated by the nonlinear interaction.—WT

5,717,657

### 43.30.Nb ACOUSTICAL CAVITATION SUPPRESSOR FOR FLOW FIELDS

A. A. Ruffa, assignor to the United States of America  
10 February 1998 (Class 367/131); filed 24 June 1996

A set of sensor elements 13 is mounted on the inside surface of sonar dome 11. Located aft of the sensor elements, at the approximate low pressure region of the flow around the dome, is a set of high power, high frequency, acoustic source transducers 15, which are typically a piston-type transducer 18 surrounded by some baffle structure 16. When the sensors 13 detect high-level broadband noise characteristic of cavitation, the controllers 17 and amplifier 19 generate a high-amplitude, high-frequency, short duration, pulsed (300–500 kHz) signal to excite the source transducer 15. The



resultant acoustic waves are intended to increase the cavitation threshold in the flow field.—WT

5,724,305

### 43.30.Xm APPARATUS FOR ACOUSTIC NEAR FIELD SCANNING USING CONFORMAL ARRAYAL

J. A. Clark *et al.*, assignors to the United States of America  
3 March 1998 (Class 367/13); filed 30 June 1995

Various movable sets of hydrophones for scanning the acoustic near-field of large, submerged, vibrating structures are discussed.—WT

5,602,801

### 43.30.Yj UNDERWATER VEHICLE SONAR SYSTEM WITH EXTENDABLE ARRAY

F. Nussbaum and C. H. Beauchamp, assignors to the United States of America  
11 February 1997 (Class 367/165); filed 6 December 1995

The effective acoustic aperture of an underwater, cylindrically shaped, vehicle, viz., a torpedo, is increased by mounting a set of transducers on a series of arms that lie in longitudinal conformal channels positioned around the vehicle outer circumference. These arms are hinged and can be deployed radially outward from the vehicle axis like spokes of a wheel to create a relatively large diameter planar array. In the retracted positions, these additional transducers provide side-looking capability with 360 degree circumferential coverage.—WT

5,719,824

### 43.30.Yj TRANSDUCER ASSEMBLY WITH ACOUSTIC DAMPING

S. G. Boucher, assignor to Airmar Technology Corp.  
17 February 1998 (Class 367/176); filed 7 May 1996

A hull-mounted ultrasonic sonar transducer, which is designed to be insensitive to the flow noise associated with high-speed water flow over its housing, is described. The active element is a conventional piezoceramic disc mounted within a urethane-filled cavity of a housing fashioned from a thermoplastic material. One thin face of the housing serves as the acoustic window. The piezoceramic disc is surrounded on all but one face by an electrostatic shield. All but this active face of the piezoceramic disc and all surfaces of the electrostatic shield are covered with low impedance material such as neoprene foam or Corprene, the purpose of which is to acoustically shield the active element from responding to extraneous acoustic energy incident upon it from directions other than that of the active face.—WT

5,595,243

### 43.35.Zc ACOUSTIC WELL CLEANER

V. E. Maki, Jr. and M. M. Sharma, assignors to the University of Texas  
21 January 1997 (Class 166/177.2); filed 10 October 1995

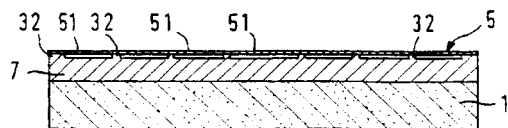
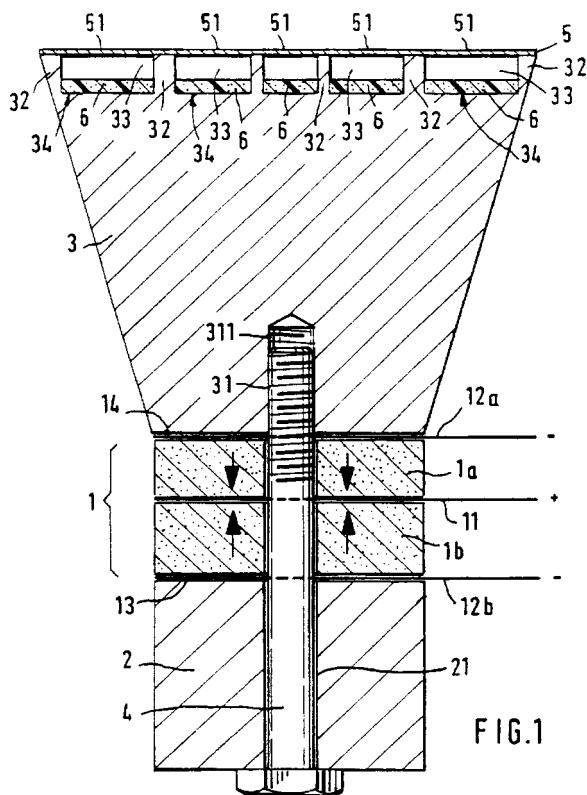
A cylindrical shaped array of conventional tonpizl-type transducers is lowered into an oil well hole. The high intensity acoustic field generated when these transducers are excited helps unplug the pores of the walls of the well hole that have become clogged with various sediments, thereby increasing the oil flow.—WT

5,726,952

### 43.38.Ar SOUND OR ULTRASOUND SENSOR

M. Eckert and K. Flögel, assignors to Endress Hauser GmbH, Germany  
10 March 1998 (Class 367/140); filed in Germany 18 May 1996

A tonpizl-type transducer, shown in cross section, has a series of concentric grooves or cavities 33 cut into the front face of its conical shaped radiating head mass 3 (in spite of the title, the primary discussion concerns the use of the transducer as a source). This face is then covered with a metal disc 5 which is rigidly attached to the interstices 32 creating a set of concentric annular membranes 51 with clamped boundary conditions, each of whose lowest resonance frequency is higher than the lowest fundamental length mode resonance frequency of the entire transducer. Hence when the



transducer is excited at the length mode resonance frequency, the membranes 51 deflect in phase with one another, but not necessarily with the same amplitude, and in phase opposition to the displacement of the interstices 32. This variation of the amplitude of displacement across the radiating face can presumably be adjusted, by the choice of the sizes of grooves 33 and hence the resonance frequencies of the membranes 51, to achieve a desirable beam pattern. In the case of pulsed excitation, damping material 6 is placed at the bottom of the grooves to minimize ringing.—WT

5,745,438

### 43.38.Bs ELECTROSTATIC TRANSDUCER AND METHOD FOR MANUFACTURING SAME

J. A. Hill and A. R. H. Goodwin, assignors to Gas Research Institute  
28 April 1998 (Class 367/181); filed 31 January 1997

The design and construction of two electrostatic transducers suited for use in extreme temperature and/or extreme pressure environments, and which should not experience polymer diaphragm degradation or decomposition, are discussed.—WT

5,724,315

### 43.38.Fx OMNIDIRECTIONAL ULTRASONIC MICROPROBE HYDROPHONE

M. B. Moffett *et al.*, assignors to the United States of America  
3 March 1998 (Class 367/153); filed 29 May 1996

A very small hydrophone, designed for use up to 10 MHz, is realized by depositing layers of PZT, containing numerous, internal, parallel, column-shaped air voids, onto both faces of a substrate member which serves as both mechanical support and an electrode surface. The voids are created by a constrained densification process during sintering. The voids ensure that the lateral stresses within the PZT are small compared to those in the thickness direction so that the desirable 3-3 mode of electromechanical coupling is dominant. A suitable waterproof encapsulant is applied to the assembly. Variations of this design are discussed.—WT

5,748,566

### 43.38.Fx ULTRASONIC TRANSDUCER

J. M. Goodson, assignor to Crest Ultrasonic Corp.  
5 May 1998 (Class 367/158); filed 9 May 1996

The design of a basic longitudinal vibrator, with a piezoceramic motor section, is modified to include an additional member of metal or ceramic, denoted a resonance enhancing disc, which is inserted between the motor section and the head mass. It is alleged that this additional member increases the intensity of the response at resonance (as well as shifting the resonance frequencies) and results in a more stable, i.e., temperature invariant, response.—WT

5,816,289

### 43.50.Gf DEVICE FOR DAMPENING VIBRATION AND NOISE IN HYDRAULIC INSTALLATIONS

Alfons Knapp, assignor to Masco Corporation  
6 October 1998 (Class 137/625.17); filed in Italy 22 June 1993

The device described in this patent has the purpose of suppressing the noise and vibration produced by water flow through faucets, mixing valves, and the like. The device consists in essence of a flexible element (e.g., of

rubber) that includes an air pocket. It is intended to be placed in an area where it is in contact with the water, but out of the direct flow stream. It may be embodied, for example, in the shape of a thick washer or of a cylindrical liner located within a faucet body or in a separate cartridge.—EEU

5,821,473

**43.50.Gf SILENCER AND A METHOD FOR FORMING AND ATTACHING A SILENCER TO A BLOWER PIPE**

**Kazunori Takahashi, assignor to Kioritz Corporation, Tokyo**  
 13 October 1998 (Class 181/224); filed in Japan 1 July 1996

A silencer for particular application to leaf blowers is described which consists of a perforated inner pipe and sound absorbing material between this inner pipe and the blower (exhaust) pipe. In order to reduce the cost and complexity of manufacture, a clever system which relies on flexible plates forming the inner pipe is described which eliminates the need for fasteners.—KPS

5,821,474

**43.50.Gf MUFFLER WITH VARIABLE DAMPING CHARACTERISTICS**

**Thomas Olszok et al., assignors to Heinrich Gillet Gmbh and Company, Germany**  
 13 October 1998 (Class 181/254); filed in Germany 2 November 1995

This patent relates to a muffler that incorporates a valve which alters the flow path of exhaust gases. The opening or closing is typically controlled by the difference in pressure between ambient and that inside the muffler such that the valve is closed when the vehicle is coasting. This patent describes an actuating mechanism that can incorporate a number of variables into the valve operation. It consists of several chambers, diaphragms, and control pressure lines, and is able to vary the muffler performance characteristics by opening and closing various valves based upon operating conditions such as engine air intake pressure, muffler pressure, and barometric pressure. Several geometrical arrangements with various numbers of chambers and diaphragms are described.—KPS

5,821,475

**43.50.Gf VENTURI MUFFLER WITH VARIABLE THROAT AREA**

**Clyde A. Morehead and John W. Henry, assignors to the United States of America**  
 13 October 1998 (Class 181/255); filed 8 May 1996

A muffler to attenuate low frequency noise generated at the gas intake of compressors and internal combustion engines is described in which one or more venturi nozzles are connected to a chamber which is in turn connected to the engine intake. Methods to adjust the venturi throat areas are described for both circular and rectangular cross sections. The throat area is sized to increase the average flow speed to between about 0.7 and 1.0 of the sound speed, and the chamber is sized to be greater than the volume of one engine intake cycle. Arrangements of multiple venturi nozzles that can be selectively activated are described.—KPS

5,821,472

**43.50.Nm SUPERSONIC AIRCRAFT EXHAUST NOISE SUPPRESSION SYSTEM**

**Gregory R. Zwernemann, assignor to Northrop Grumman Corporation**  
 13 October 1998 (Class 181/215); filed 21 August 1997

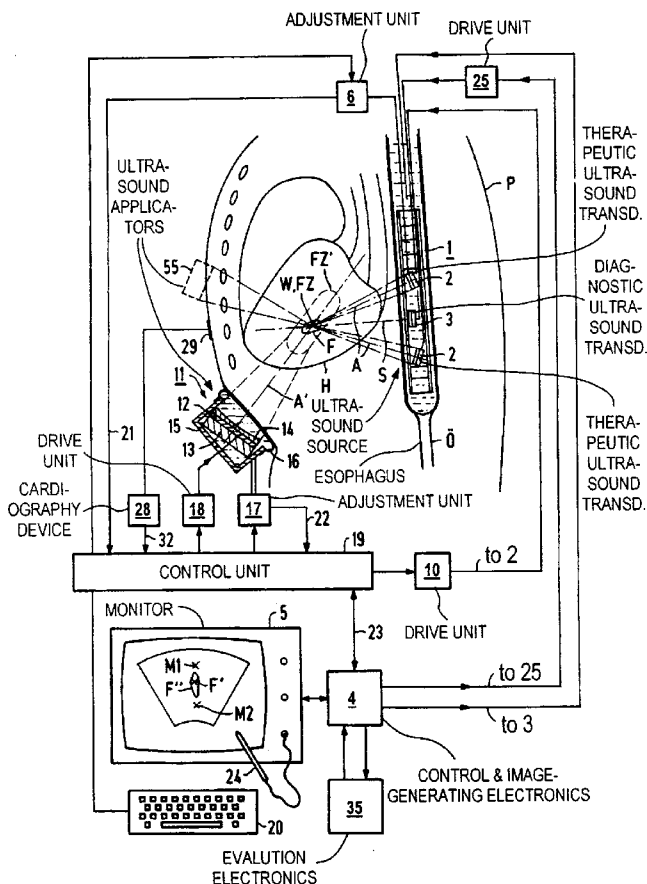
A noise suppression system for jet engine exhaust is described in which the exhaust gas flows through plates containing many small openings, thus forming a multitude of smaller jets and shifting the noise to a higher frequency which is more readily absorbed by the atmosphere. The openings in the plates may be narrow slots, round holes, or nozzles. A similar system is described for reducing the noise from the flow of ambient air which is entrained in the exhaust. Various geometrical arrangements are described which enable the suppression system to be deployed only when needed, such as on landing or takeoff.—KPS

5,817,021

**43.80.Qf THERAPY APPARATUS FOR TREATING CONDITIONS OF THE HEART AND HEART PROXIMATE VESSELS**

**Helmut Reichenberger, assignor to Siemens Aktiengesellschaft**  
 6 October 1998 (Class 600/439); filed 22 March 1994

This apparatus produces ultrasonic waves with an intensity that causes necrosis. The transducers that produce the high intensity ultrasonic waves are located in a transesophageal probe and in a probe applied subcostally or



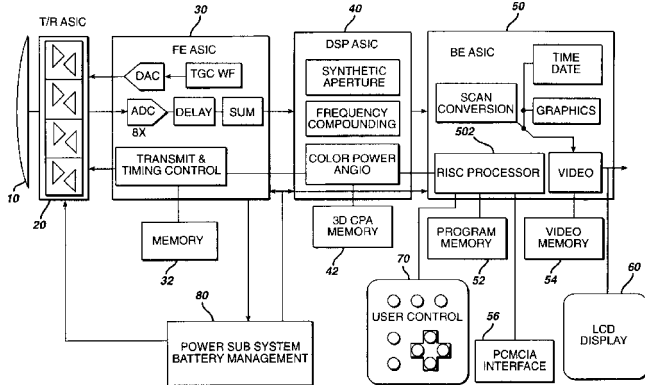
transthoracically. A diagnostic ultrasound imaging unit is included in the apparatus.—RCW

5,817,024

**43.80.Vj HAND-HELD ULTRASONIC DIAGNOSTIC INSTRUMENT WITH DIGITAL BEAMFORMER**

William R. Ogle *et al.*, assignors to SonoSight, Inc.  
6 October 1998 (Class 600/447); filed 27 May 1997

This ultrasonic instrument produces both b-scan and bloodflow images. A digital beamformer that forms ultrasonic scan lines from echo signals at the elements is contained in the enclosure. Images produced by the



instrument that is implemented with application specific integrated circuits (ASIC's) are viewed either on a standard monitor or a liquid crystal display.—RCW

5,820,554

**43.80.Vj ULTRASOUND BIOPSY NEEDLE**

Richard E. Davis and Garey L. McLellan, assignors to Medtronic, Inc.  
13 October 1998 (Class 600/431); filed 24 March 1997

The stylet in the tubular cannula of this needle has grooves to enhance echogenicity for visualization of the needle in an ultrasonic image.—RCW

5,820,564

**43.80.Vj METHOD AND APPARATUS FOR SURFACE ULTRASOUND IMAGING**

Michael H. Slayton, Young-Kwan Cho, and Peter G. Barthe, assignors to Albatross Technologies, Inc. and Medison, Co., Ltd.  
13 October 1998 (Class 600/459); filed 16 December 1996

In this method and apparatus, ultrasonic beams from a probe are dif-fused perpendicular to a scanning direction and thereby permit an ultrasonic image of a surface perpendicular to the scan plane to be produced while other ultrasonic beams are focused to permit formation of a conventional b-scan image. The probe utilizes a so-called 1.5-D array or a 2-D array. Conventional ultrasonic imaging apparatus and processing can be used.—RCW



# The precedence effect

Ruth Y. Litovsky<sup>a)</sup> and H. Steven Colburn

*Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215*

William A. Yost and Sandra J. Guzman

*Parnly Hearing Institute, Loyola University Chicago, Chicago, Illinois 60201*

(Received 20 April 1998; revised 9 April 1999; accepted 23 June 1999)

In a reverberant environment, sounds reach the ears through several paths. Although the direct sound is followed by multiple reflections, which would be audible in isolation, the first-arriving wavefront dominates many aspects of perception. The “precedence effect” refers to a group of phenomena that are thought to be involved in resolving competition for perception and localization between a direct sound and a reflection. This article is divided into five major sections. First, it begins with a review of recent work on psychoacoustics, which divides the phenomena into measurements of fusion, localization dominance, and discrimination suppression. Second, buildup of precedence and breakdown of precedence are discussed. Third measurements in several animal species, developmental changes in humans, and animal studies are described. Fourth, recent physiological measurements that might be helpful in providing a fuller understanding of precedence effects are reviewed. Fifth, a number of psychophysical models are described which illustrate fundamentally different approaches and have distinct advantages and disadvantages. The purpose of this review is to provide a framework within which to describe the effects of precedence and to help in the integration of data from both psychophysical and physiological experiments. It is probably only through the combined efforts of these fields that a full theory of precedence will evolve and useful models will be developed. © 1999 Acoustical Society of America.

[S0001-4966(99)01910-4]

PACS numbers: 43.10.Ln, 43.71.An, 43.71.Bp [ADP]

## INTRODUCTION

This paper reviews recent work and current thinking about a group of auditory phenomena that are thought to account for listeners’ abilities to function in reverberant spaces, and that for historical reasons have been attributed to the “precedence effect.” The most extensive review to date on this topic is in Blauert’s classic book on spatial hearing which was just recently expanded and reprinted (Blauert, 1997). Blauert reviews the classic data on precedence up to about 1982, and then in a newly added chapter he reviews some of the recent work on the buildup phenomenon (which is covered in Sec. II of this paper). Zurek (1987) also provides a review of the work through the mid-1980s, and related chapters can be found in a recent book edited by Gilkey and Anderson (1997). In recent years there has been a resurgence of interest in the precedence effect by psychoacousticians and physiologists, and a new body of literature has been produced which has led us, and many others, to re-evaluate our assumptions about the auditory and neural mechanisms involved. This review is our attempt to encompass all of these topics, especially those that have been studied since the mid-1980s.

When a sound is produced in a reverberant environment, it propagates in multiple directions and is subsequently reflected from nearby surfaces. The auditory system is thus faced with resolving competition between the first sound and its reflections for perception and localization. Despite this

clutter of information, we can localize sound sources and identify their meaning fairly accurately. Figure 1 illustrates a recording of a source click and its reflections. The recording was made in a “typical” classroom (approximately  $6 \times 11$  m). The sound source (*S*) was a brief click delivered 1.3 m in front of a blackboard and the recording measurement was made 4 ft in front of the source. Three reflections can be identified and these are labeled *R1*, *R2*, and *R3*. They occur approximately 8, 8.5, and 10 ms after the direct click and are attenuated relative to the source. For example, the first reflection occurs about 8 ms after the source and its level is attenuated by about 9.5 dB from the level of the source click. These reflections are sometimes referred to as “early reflections” to differentiate them from the total reverberation created by the interaction of all the reflections (i.e., the total acoustic clutter produced by a sound in a reflective environment). In general, a reflection is an attenuated, sometimes spatially separated, delayed and coherent copy of the originating sound.

As a simplification of a natural situation, consider an arrangement of two loudspeakers in an anechoic room such that the speakers are equally distant, and stimulated by identical sounds such that the onset of one sound is delayed relative to the onset of the other sound. This can be considered a model of a direct sound (the lead) with a single reflection (the lag). This situation for click stimuli is shown in Fig. 2(A), and an idealized sketch of the perceived location(s) of the image(s) as a function of the lead–lag delay is shown in Fig. 2(B). The lead is at  $45^\circ$  to the right, and the lag is at  $45^\circ$  to the left. When the delay is zero and the speakers are stimulated equally, the stimuli to the two ears of the

<sup>a)</sup>Address for correspondence: Boston University, Biomedical Engineering, 44 Cummington St., Boston, MA 02215; electronic mail: litovsky@bu.edu

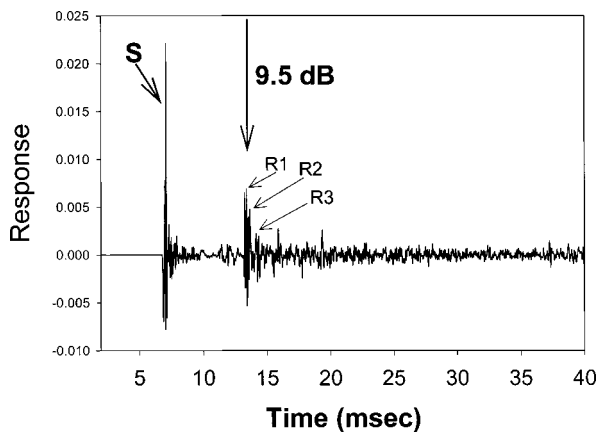


FIG. 1. The recording of the response to a 10- $\mu$ s electrical pulse in a college classroom that was approximately  $6 \times 11$  m in size. The source of the transient was placed 1.3 m in front of a blackboard attached to the front of the classroom and the recording was made 4 ft in front of the source. The first 40 ms of the recording shows the direct click (S) and several early reflections labeled as R1, R2, and R3. These reflections stand out from the overall reverberation.

listener are approximately equal and a single (fused) image is perceived in the plane of symmetry, approximately straight ahead of the listener. As the delay increases, the fused image moves toward the direction of the lead speaker, reaching this direction after about a millisecond. For delays between about 1 and 5 ms, the image is still unitary (fused) and remains located in the direction of the leading speaker. Finally, for large delays, the image breaks into two images, one at each

speaker position, as one would expect for sequential, well-separated stimuli.

This idealized situation can be used to provide the vocabulary used to describe these phenomena in general. “Summing localization” refers to a delay (0–1 ms) when the sounds from the lead and lag sources are perceptually fused and when both the lead and lag contribute to the perceived direction of the fused image (e.g., de Boer, 1940; Warncke, 1941; for review see Blauert, 1997, pp. 204–206). Note that the simplest case of summing localization, as illustrated in Fig. 2(B), assumes no temporal overlap between the direct and reflected signals and the perceived location is an average of the two directions. In cases where the stimuli overlap in time, perceived direction is mediated by more complex averaging that include the amplitudes and phases of the summed wave forms.

As the delays are increased beyond 1 ms several observations can be made. At relatively short delays (typically in the range of 1 to 5 ms, or more, depending on the stimulus wave form and room acoustics) the two sounds remain perceptually fused, hence we refer to this percept as “fusion.” As the delay increases, the lagging source becomes audible as a separate auditory event; this perceptual boundary between “one fused sound” and “two separate sounds” is often referred to as the echo threshold. Blauert (1997) emphasizes this definition for two spatially separated sources.

We emphasize the fact that echo threshold is not the threshold of detectability of the lag; lead–lag sounds and lead-only sounds can be distinguished easily based on overall sound quality (e.g., Blauert, 1997). The echo threshold estimates the delay at which the fused image perceptually splits into two images. This distinction is important since information contained in reflections is important for our ability to sense that the environment we are in is reverberant.

In addition to fusion, the finding that at short delays the image location is dominated by the location of the leading source has been called the “law of the first wave front” or “localization dominance.” Finally, “lag discrimination suppression” refers to the observation that at short delays stimulus parameters of the lag stimulus are less discriminable due to the presence of the lead stimulus. Discrimination improves as the delays increase [see Fig. 2(C) and (D)]. In part, the purpose of this review is to introduce the vocabulary of fusion, localization dominance, and lag discrimination suppression as a means to organize the various “effects” associated with what has been called the precedence effect.

Although the lead–lag stimulus paradigm is quite idealized compared with realistic stimuli in reverberant spaces, it has become widely used in psychophysical and (in recent years also in) physiological studies. The term “precedence effect” was originally coined by Wallach *et al.* (1949) in their classic study to describe the dominance of the lead stimulus characteristics in the determination of the spatial location of the fused image (localization dominance). However, in the past two decades this term has become popularized and is used to refer to most measurements made using lead–lag stimulus configurations, regardless of the psychophysical measurement that is made. Hence, fusion, localization dominance, and discrimination suppression have all

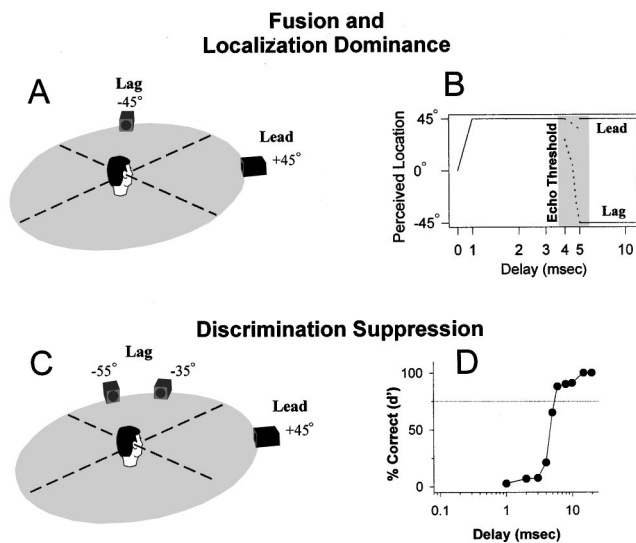


FIG. 2. Schematic diagram of spatio-temporal relation for a lead–lag click-pair. Panels A and C give examples of stimulus configurations for the fusion/localization and discrimination tasks, respectively. In panel A the lead is at  $45^\circ$  to the right, and the lag is at  $45^\circ$  to the left. Panel B shows changes in perceived locations of the auditory events as a function of the delay. With no delay, one fused image is heard at a “phantom” location between the lead and lag. Between 0–1 ms the image shifts toward the lead speaker. Between 1 ms and the “echo threshold” a fused image is heard at the lead location. When echo threshold is reached, a second image appears, initially near the lead location, and at longer delays at the lag location. In C the lead is at  $45^\circ$  to the right, and the two lag locations are at  $35^\circ$  and  $55^\circ$  to the left. Panel D shows sample data for the discrimination task, in which performance is poor at short delays and improves as delays are increased (panel B is modified from Blauert, 1997, with permission).

been used somewhat interchangeably. We view this as a potential problem, which can easily result in confusion when trying to understand the mechanisms underlying these perceptual phenomena. The primary goal of this paper is to carefully delineate between studies that claim to measure “the precedence effect,” but which may measure different perceptual effects. When appropriate, we will attempt to draw parallels and to link related findings in these areas. To date, few studies exist in which all three measurements have been made in the same listeners, hence to the extent to which fusion, localization dominance, and discrimination suppression are directly related is not clearly understood.

In addition to delineating between perceptual phenomena, we feel that it is important to keep in mind which method of stimulus presentation is used in the various studies. The situation depicted in Fig. 2 represents aspects of the precedence effect in a free-field environment. However, much of what is known comes from headphone studies, in which stimuli from loudspeakers at different locations are replaced by stimuli with different interaural time or level differences for the lead and lag sources (e.g., Wallach *et al.*, 1949; Zurek, 1980; Gaskell, 1983; Yost and Soderquist, 1984; Shinn-Cunningham *et al.*, 1993). There have also been a few experiments with headphones using virtual acoustic stimuli that attempt to recreate the same stimuli in the ear canals that arise from free-field stimulation (e.g., Dizon *et al.*, 1997; Litovsky *et al.*, 1999). These experiments allow manipulation of spectral, temporal, and level differences separately with the simulated free-field condition as the reference condition. We draw attention to this distinction because some measurements indicate that effects differ depending on the method of stimulus presentation. In Fig. 3, the stimulus configuration for the free field is shown in panel A, where the lead and lag are each shown as arriving from different locations with a delay between their onsets. The acoustic signals at the left and right earphones for the headphone stimulation case are shown in panel B, where the lead and lag each contain interaural time differences that result in images with different perceived lateralization. Finally, the resulting wave forms at the ears for the free-field case are sketched in panel C.

It is clear from Fig. 3 that stimuli that are often used in studies of the precedence effect differ dramatically from “realistic” stimuli. The lead/lag simulation of a sound source and its reflection differs in several ways from the acoustics in real rooms. Lead and lag stimuli are often clicks of equal amplitude and identical wave forms, while sound sources need not be transient and reflections are usually different from the sound source in amplitude and wave form. Nonetheless, click stimuli have been popular in precedence studies due to their transient nature, which avoids temporal overlap between the lead and lag. About 50 years ago there were some very interesting studies on the processing of complex stimuli in a source-reflection paradigm (e.g., Haas, 1949), and that work is reviewed in detail by Blauert (1997). We believe that studies on precedence, in which the stimuli are optimized and nonrealistic, represent the first step toward understanding basic auditory processes that are involved in resolving competition for perception and localization be-

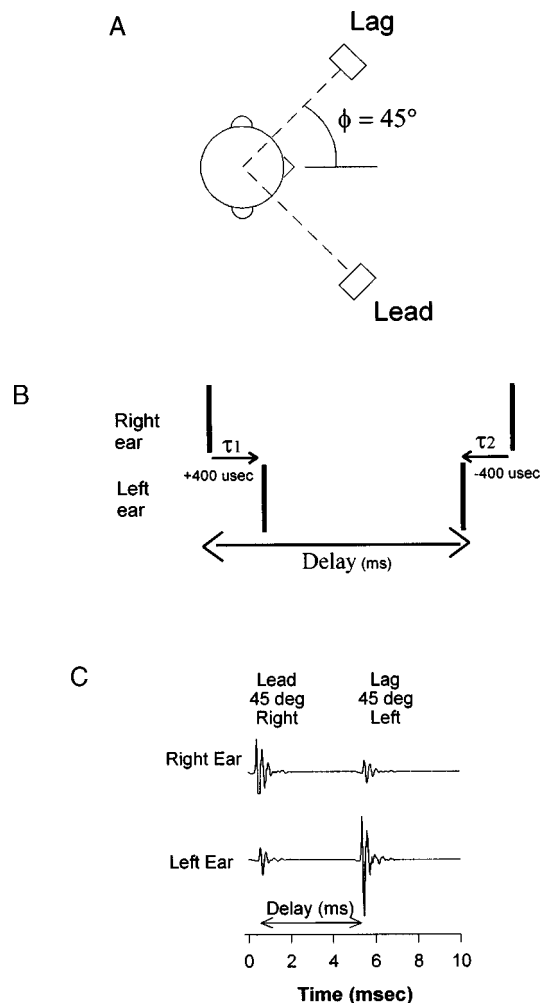


FIG. 3. Stimulus configurations commonly used in precedence studies. A: Free-field setup. Two sounds are emitted from locations  $45^\circ$  to the right and left with the right signal leading the left signal by several milliseconds. B: Dichotic headphone stimuli used by Wallach *et al.* (1949). Two click pairs with ITDs leading to the right ear and then the left ear, with a delay of a few milliseconds between the pairs. The individual wave forms at each ear for the lead and the lag pairs would be approximately the same levels and differ only in the interaural delay. C: Sketch of two impulse responses to click-pairs as the sound sources reach the ears in free field from  $45^\circ$  to the right and left. There are natural interaural differences in time and intensity between the stimuli at the two ears, as well as some differences in the spectral shape of the sounds. Finally, to simulate the precedence effect there is an interstimulus delay, simulating the echo delay.

tween a source and its reflection. Ultimately, experimenters should aim to use more realistic stimuli; however, while the mechanisms underlying precedence are not well understood, minimizing stimulus complexity remains essential.

In this paper we will use lead and lag stimuli to refer to conditions involving the types of simulations we have just described. We will reserve the terms “sound source” and “reflections” to discussions of real acoustic environments or applications of experimental work to the real world. We also will reserve the word “echo” for situations in which fusion breaks down and a reflection or lag stimulus is perceived as a separate sound source. The vocabulary we are applying to the study of precedence in this review is not necessary for understanding the literature. However, we believe that it will help the reader *better* understand the literature, and it will

TABLE I. Critical thresholds for fusion, discrimination suppression, and localization dominance.

Study	Stimulus	Thresholds	Criterion for threshold
<b>FUSION ECHO THRESHOLDS</b>			
Haas (1951)	speech	30–40 ms	“echo annoying”
Lochner and Burger (1958)	speech	50 ms	lead and lag “equally loud”
Schubert & Wernick (1969)	noise		
	a) 20-ms duration	5–6 ms	lead and lag “equally loud”
	b) 50-ms duration	12 ms	
	c) 100-ms duration	22 ms	
Ebata <i>et al.</i> (1968)	clicks	10 ms	fused image at center of the head
Freyman <i>et al.</i> (1991)	clicks	5–9 ms	lag heard on 50% of trials
Yang and Grantham (1997a)	clicks	5–10 ms	lag clearly audible on 75% of trials
Litovsky <i>et al.</i> (1999)	clicks	5–10 ms	lag clearly audible on 75% of trials
<b>DISCRIMINATION CRITICAL THRESHOLDS</b>			
Freyman <i>et al.</i> (1991)	clicks	5–9 ms	$d' = 1$
Yang and Grantham (1997b)	clicks	5–10 ms	discrimination 75% correct
Litovsky <i>et al.</i> (1999)	clicks	5–10 ms	discrimination 75% correct
<b>LOCALIZATION CRITICAL THRESHOLDS</b>			
Litovsky <i>et al.</i> (1997b)	clicks	11.4 ms	lead location chosen on 75% of trials
Litovsky <i>et al.</i> (1997a)	clicks	8 ms	lead location chosen on 75% of trials

facilitate the integration of information about precedence as well as aid in modeling sound processing in reflective environments.

Most of this review is concerned with recent psychophysical research related to the precedence effect; this constitutes the material of Sec. I. In separate subsections, measurements of fusion, localization dominance, and discrimination suppression are discussed individually. Section II considers recent results showing how preceding stimulation affects perception of simulated reflections, and we describe two important findings based on repeating lead–lag stimulus pairs: the buildup and the breakdown of some of the effects of precedence. In Sec. III, material related to human development and animal experiments is considered. In Sec. IV, attention is given to neural mechanisms that might underlie psychophysical phenomena related to precedence; hence, we will attempt to weave together and correlate findings in psychophysics and physiology. Finally, in Sec. V we consider current models of precedence and their limitations.

## I. PSYCHOPHYSICS WITH SIMPLE PAIRED STIMULI

### A. Fusion

Fusion is a striking perceptual effect: at short delays between a sound source and its reflection, while two or more equally loud, spatially separated sounds are physically present, listeners perceive only one fused auditory event. This effect can be quite useful for avoiding multiple sound images that may arise from the source and its reflections. A listener’s ability to negotiate sounds in a reflective space can be enhanced if the auditory system groups a sound source and its reflections in to a single, coherent auditory percept.

In order to measure fusion in a controlled laboratory environment, listeners are usually asked to provide a subjective impression of how many sounds they hear, and measurements are usually repeated at various lead–lag delays. Results can be plotted as psychometric functions that show the percent of trials on which “two sounds” are reported as a

function of delay. At short delays (<5 ms for clicks) most listeners report hearing one sound on the majority of trials, and as the delays are increased the proportion of trials on which “two sounds” are reported increases, usually reaching 100% by 8 to 10 ms. In addition to the audibility of the lag, other perceptual changes occur, including the influence of the lag on aspects of the fused image, such as its loudness, spatial extent, and pitch. These types of perceptual changes also depend on the type of signal, the signal level, the direction of the sources, and whether the sounds are presented in free field or over headphones. It is important to note that the detectability of the lagging source is not of primary interest; listeners are usually able to distinguish between trials in which a lag is present and trials in which the lead is presented alone (Guzman and Yost, 1999). Rather, the focus is on whether the lag is perceptually fused with the lead or whether it is perceived as a separate sound.

A common measure of the temporal boundary between separating perception of “one fused sound” from “two separate sounds” is the *echo threshold* (for review see Zurek, 1987; Blauert, 1997). Quantitative estimates of echo thresholds vary tremendously (2–50 ms), depending on a number of variables. The primary determinant of the echo threshold seems to be the nature of the stimulus, although instructions to subjects are probably important and few laboratories used several types of stimuli within a single study. In addition, spatial separation might be a determinant. Blauert’s (1997) definition assumes that the lead and lag are separated while not all other writings do. In Table I we list examples of echo thresholds obtained with different stimuli, showing that echo thresholds are much shorter for brief stimuli such as clicks than for longer duration stimuli such as noise and ongoing complex stimuli such as running speech. Although the table is not inclusive of all studies conducted on fusion, it represents a good summary of the stimulus variables that might influence echo threshold.

Fusion studies have an inherent subjective nature, hence estimates of echo threshold also depend on the instructions



given to the listener. Thus echo thresholds can vary depending on whether one measures delays at which the “primary auditory event and reflection [are] equally loud” (e.g., Haas, 1951; Meyer and Schroeder, 1952; Lochner and Burger, 1958; Franssen, 1960, 1963; David, 1959), the “reflection [is] annoying” (Haas, 1951; Muncey *et al.*, 1953), “a second sound is heard at the vicinity of the lag speaker” (e.g., Freyman *et al.*, 1991), or “one or more than one potential sources of the perceived sound are detected” (e.g., Yost and Guzman, 1996). A large majority of studies on fusion was conducted prior to 1971 and usually in the free field. These studies are discussed extensively by Blauert (1997). Much less work has been done over headphones, and this work (e.g., Yost and Soderquist, 1984) suggests that echo thresholds are shorter for headphones than for free-field delivered stimuli.

More recently, several additional aspects of fusion have been noted. There seems to be significant intersubject variability in the strength of fusion (e.g., Clifton and Freyman, 1989; Freyman *et al.*, 1991; Clifton *et al.*, 1994). With click stimuli, some listeners no longer experience fusion at relatively short delays (2–4 ms), while others experience a strong effect that lasts beyond 10 ms. In addition, it has been suggested (Litovsky and Colburn, 1998) that spatial separation between lead and lag significantly reduces echo threshold.

In the classic literature on precedence there seems to be an underlying assumption that the precedence effect is a binaural phenomenon, mediated by the binaural circuits in the auditory system, and hence it has been modeled using binaural inputs (e.g., Blauert, 1997). It turns out that several aspects of precedence, including fusion, occur at similar delays under binaural and monaural conditions.

One might ask what natural situation would result in presentation of sounds via a monaural system? In recent years at least two such scenarios have been identified and studied. The first has been to compare listeners’ performance in the azimuthal and median-sagittal planes, where the relative strength of binaural disparity cues and spectral cues differ in the two dimensions (Litovsky *et al.*, 1999; Rakerd *et al.*, 1997). The second approach has been to measure fusion in listeners who have profound monaural deafness (Litovsky *et al.*, 1997a). Monaural listeners are accustomed to functioning in their environment using information from one ear alone, hence they probably represent a “perfect” monaural system much more realistically than would normal-hearing listeners tested under monaural conditions. The results are suggestive of a fusion mechanism that is not dependent on binaural processing. In the azimuthal-median studies, the delay at which listeners perceived “one sound” or “two sounds” for either clicks (Litovsky *et al.*, 1999) or speech (Rakerd *et al.*, 1997) is nearly identical for most listeners. In the study using monaural listeners, fusion was found at similar delays for the normal-hearing and the monaural populations.

To summarize, a basic perceptual consequence of listening in reflective environments is that sources and their reflections become perceptually fused into a single coherent auditory percept. This effect is strongest at short delays (1–8

ms), not unlike those that occur in relatively small enclosed spaces. Sound travels at the speed of approximately 1 meter per 3 ms, hence a reflection might take up to 6 ms to reach the ears from a nearby wall, and up to 10 ms or longer from a far wall. The advantages of fusion might therefore be most noticeable in small rooms, and decrease as the reflective surfaces are placed farther from the listener. Fusion studies have been mostly limited to subjective impressions reported by listeners, with little emphasis on the perceived location of the lead, lag, or fused image. The next section on localization dominance focuses on processing of directional information.

## B. Localization dominance

It is not hard to imagine what a listener might experience if directional information contained in a source and in its reflections were weighted equally by the auditory system. It would be difficult to identify the true location of the source. However, that is not the case. The reader can relate to his/her own experience, whereby a source can usually be correctly localized, and the reflections contribute relatively little directional information. This everyday experience is what we term “localization dominance.” In a controlled laboratory situation this phenomenon is usually studied by simulating one source (lead) and one reflection (lag), and the effect is thought to be strongest at short delays (greater than a millisecond and shorter than echo threshold). This is not to say that directional information from the lag is completely ignored, but that the contribution of the lead to localization of the fused image is much stronger than that of the lag.

### 1. Headphone studies

The bulk of studies on localization dominance have been conducted under headphones using “adjustment” protocols in which listeners match the position of the fused image to that of a reference stimulus (or to midline) by varying a binaural parameter, such as interaural differences in time (ITD) or level (ILD), of either the lead or lag (von Békésy, 1930; Wallach *et al.*, 1949; Haas, 1949; Snow, 1954b; Leakey and Cherry, 1957; Zurek, 1980; Yost and Soderquist, 1984; Shinn-Cunningham *et al.*, 1993, 1995). The dominance of the lead is quite compelling. For example, in their classic study, Wallach *et al.* (1949) reported combinations of lead and lag ITDs (each favoring a different ear) that resulted in a fused auditory image at the center of the head. At a delay of 2 ms an ITD of 100  $\mu$ s in the lead required an ITD of approximately 400  $\mu$ s in the lag. The fact that the lag did succeed in centering the fused image suggests that the lead did not dominate perception completely. In fact, a simple linear weighting function would imply that the directional cues contained in the lead were weighted four times more heavily than those of the lag.

A more precise estimate of localization dominance can be achieved by asking subjects to “point” to the perceived location of the auditory image using an acoustic pointer whose ITD can be manipulated by the listener (e.g., Zurek, 1980; Shinn-Cunningham *et al.*, 1993). The advantage of this protocol is that the perceptual weight of the lead and lag can be directly estimated from the data using very few parameters. Using a simple model, Shinn-Cunningham *et al.*

(1993) succeeded in calculating weighting factors for the lead and lag from several published studies and found that typical values ranged from 80%–90% and 10%–20% for the lead and lag, respectively. Hence, Wallach *et al.*'s (1949) original finding, that while the lead dominates localization the lag contribution is also important, has been confirmed over the years in several labs using various techniques.

## 2. Azimuthal-plane studies in free fields and rooms

While headphone studies were successful in providing a quantitative measure of the perceptual weights of the lead and lag, they certainly did not provide a “realistic” acoustic environment. Free-field studies, on the other hand, do provide more realistic everyday scenarios, although they have their own limitations. Early studies in free field using a two-loudspeaker system showed that the perceived location of the fused image is dominated by the lead speaker (e.g., Wallach *et al.*, 1949; Leakey and Cherry, 1957; Snow, 1954a; Haas, 1951; Leakey, 1959). Recall that these stimuli are further unrealistic in that the lag is not attenuated relative to the lead; this is unlike a true reflection, which can be dramatically attenuated relative to the source (e.g., Fig. 1). A classic example of this effect was demonstrated by Leakey and Cherry (1957) for loudspeakers located at 45° to the right and left, using speech signals. With no delay, the fused image appeared at a central location in front of the listener, and a delay of a few milliseconds in one loudspeaker shifted the entire auditory image to the other, leading speaker. Leakey and Cherry also showed that increasing the level of the delayed signal by several decibels shifts the fused image back to its central location. Similarly, Haas (1951) demonstrated that when the level of the leading source is decreased there appear to be two simultaneous sources in the directions of the two speakers. Hence, there is trade-off in localization dominance between delay and relative signal levels of the lead and lag.

To date, one published study has been conducted in the azimuthal plane with more than two speakers. Figure 4 shows data from Litovsky *et al.* (1997b) with three speakers (nine possible lead–lag combinations) at delays of 0 to 10 ms. In the azimuthal plane (open symbols) at delays of 1–2 ms, the leading source location was chosen on 95% of trials, providing strong evidence that the lead dominated localization. At longer delays (above 5 ms), both lead and lag locations were chosen equally, suggesting that localization dominance was no longer effective (listeners heard two sounds but could not determine which was the leading source). Results of free-field studies are generally consistent with the headphone studies; however, the techniques have been less sophisticated and have yielded little information regarding the relative weights of the lead and lag in localization.

Finally, a related auditory illusion known as the “Franssen effect” occurs with similar stimulus conditions in which the first sounds arriving at a listener dominate spatial perception (Franssen, 1963; see Hartmann and Rakerd, 1989; Yost *et al.*, 1997). This illusion refers to the finding that the location of a long duration tone at one spatial location is identified as arising from the location of a short tone burst that precedes the longer tone. For example, if a 50-ms tone at one

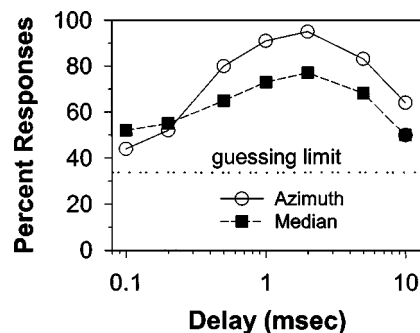


FIG. 4. Localization dominance measures in free field. Lead and lag locations varied in 90° steps along each plane (0, ±90° in azimuth and 0, 90, 180° in median). Listeners reported the location that was nearest to that of the fused event. Mean responses for eight subjects are plotted. For both source planes, the plots show the percentage of trials in which the leading source was nearest to the sound image (reprinted from Litovsky *et al.* 1997b, with permission).

location is presented with a sudden onset and a gradual offset, and at a different location a tone is turned on gradually and remains on for a long time, listeners will perceive a short tone at the onset from the leading speaker, then a room-filling diffuse ongoing sound. When forced to choose a location, subjects resort to the initial tone location. The illusion works for midfrequency tones presented in reverberant rooms. It does not work for noise stimuli, for stimuli presented in anechoic rooms or over headphones, or when the tones are low or very high in frequency (Hartmann and Rakerd, 1989; Yost *et al.*, 1997). The ability to produce the Franssen illusion is correlated with the ability of listeners to localize sounds, in that sounds that are difficult to localize appear to generate the strongest Franssen illusion (Hartmann and Rakerd, 1989; Yost *et al.*, 1997). Thus the Franssen effect is consistent with localization dominance in the precedence effect, in that the first information arriving at the listener controls the reported location of the perceived sound.

## 3. Free-field studies in the median sagittal plane

Since reflections in natural rooms arrive from multiple directions, including the walls, ceiling, and floor, they are likely to contain directional information that includes binaural cues (ITD and ILD) as well as monaural spectral and level cues. Until very recently, studies on localization dominance have focused exclusively on stimuli containing binaural cues, simulating reflections arriving from the walls. Little attention has been paid to scenarios in which reflections might contain few binaural cues and the primary directional cues are monaural spectral, such as when they arrive from the ceiling and floor directly in front of the listener. To the extent that models of localization dominance exist, they have thus focused on binaural mechanisms.

To date, a handful of studies have attempted to measure localization dominance under conditions in which monaural spectral cues are prominent. The first such study (Blauert, 1971) demonstrated that when the lead and lag are presented in the median sagittal plane from 0° (front) or 180° (back) and the lead–lag delay is greater than 0.5 ms, the lead dominated the perceived location of the fused image. However, the longest delay used in that study was 0.88 ms, which is

within the “summing localization” window and not quite within the temporal region normally used in azimuthal/binaural studies on localization dominance (Blauert, 1997). Several recent studies (Rakerd and Hartmann, 1997; Litovsky *et al.*, 1997b; Dizon *et al.*, 1997) measured localization dominance in the median sagittal plane at delays greater than 1 ms. These studies employed a multiple-loudspeaker paradigm with various combinations of lead-lag locations, and listeners were instructed to report which of the loudspeakers was nearest to the perceived location of the fused sound image. The filled symbols in Fig. 4 show the results from Litovsky *et al.* (1997b) in the median-sagittal plane. At delays of 1–2 ms the leading source location was chosen on a majority of trials, as was found in the azimuthal plane. Thus the basic effect of localization dominance occurs at similar delays in the azimuthal or median planes. The effect does seem a bit weaker in the median plane, possibly since localization is poorer in the median plane, and adding a reflection creates coloration caused by a spectral ripple (Yost *et al.*, 1996) which can lead to localization errors (Blauert, 1997).

### C. Lag discrimination suppression

While studies on fusion and localization dominance measure perceptual effects that are somewhat analogous to our experience in reflective environments, there exists a third approach to probing the auditory system’s sensitivity to stimulus parameters of the lead or lag. These experiments on lag discrimination suppression measure the ability of listeners to process spatial information about the lagging stimulus (relative to the leading stimulus). Experiments have been conducted both under headphones and in free field. Headphone experiments measure just-noticeable-difference in the ITD or ILD of the lead and lag (e.g., Zurek, 1980; Gaskell, 1983; Yost and Soderquist, 1984; Saberi and Perrott, 1990; Shinn-Cunningham *et al.*, 1993; Saberi, 1996; Tollin and Henning, 1998), whereas free-field experiments measure discrimination of positional changes (e.g., Perrott *et al.*, 1989; Freyman *et al.*, 1991; Litovsky and Macmillan, 1994; Yang and Grantham, 1997a, 1997b; Litovsky, 1999). In general, when the delay is short (less than 5 ms for clicks), changes in the lag are extremely difficult to discriminate; changes in the lead are much easier to discriminate, although performance is somewhat worse than on single-source discrimination. As Litovsky and colleagues have pointed out (Litovsky and Macmillan, 1994; Litovsky, 1997; Litovsky and Ashmead, 1997), while the lead-discrimination task requires listeners to ignore irrelevant information in the lag, the lag-discrimination task requires that listeners extract information from a sound that is not audible as a separate auditory event. Based on what is known about localization dominance, a prediction can be made that lead discrimination should be somewhat hampered by the presence of the lag, since the latter is not completely suppressed. In addition, lag discrimination should be the worse condition since directional information contained in the lag is the primary target of the suppressive mechanism.

## 1. Studies using headphones

The advantage of headphone studies is that they allow careful stimulus control and independent manipulation of each directional cue. The classic demonstration of the discrimination effect was Zurek’s (1980) report that listeners’ ability to discriminate changes in ITDs or ILDs of the lagging source deteriorates in the presence of the lead. The effect is largest at delays of 2–3 ms, the same delay range at which localization dominance and fusion are also known to be quite strong (see Secs. IA and IB).

Various aspects of this finding have since been replicated, with some caveats regarding the exact delays at which the effect is strongest (e.g., Gaskell, 1983; Tollin and Henning, 1998), limitations of the procedure due to high inter- and intra-subject variability (Yost and Soderquist, 1984), and possible effects of training (Saberi and Perrott, 1990). One of the intriguing outcomes of such studies has been a report that discrimination performance can be directly related to measures of localization dominance. Shinn-Cunningham *et al.* (1993) developed a simple model for localization “adjustment” data, which was able to accurately predict performance on the discrimination task, suggesting a tight relation between these two measurements. Thus when the leading source dominates the perceived location of the fused image, listeners are also unable to extract directional cues from the lag as easily as they can from the lead. This is consistent with Zurek’s (1980) postulation that the precedence effect results from a temporary loss of sensitivity to interaural cues shortly after the onset of the lead.

## 2. Studies in free field

Discrimination studies in free field can be categorized into two complementary experimental paradigms. One approach, which is analogous to ITD or ILD just-noticeable-differences (JNDs), has been to measure the minimum audible angle (MAA; estimating the smallest change in the direction of a sound source that can be reliably detected) for the lead, lag, or a single source sound. In these studies the delay is kept short (2–4 ms), within the range of delays where fusion is known to be strong. Numerous studies have shown that lag-discrimination MAAs are higher than single-source MAAs by a few degrees (Perrott *et al.*, 1987; Perrott and Pacheco, 1989; Perrott *et al.*, 1989; Litovsky and Macmillan, 1994; Litovsky, 1997). In addition, lead-discrimination MAAs are better than lag-discrimination, but still somewhat worse than single-source MAAs (Litovsky and Macmillan, 1994; Litovsky, 1997). These results suggest that, while the presence of the lead renders directional information present in the lag difficult to access, the lag also interferes with lead discrimination. Hence, the precedence effect does not represent complete suppression of the lag, but rather a strong dominance by the lead. A useful method for normalizing performance is the ratio of lead/single or lag/single thresholds (Litovsky, 1997; Tollin and Henning, 1998), which is appropriate for comparing performance across conditions or between groups of various ages or with differing amounts of hearing loss.



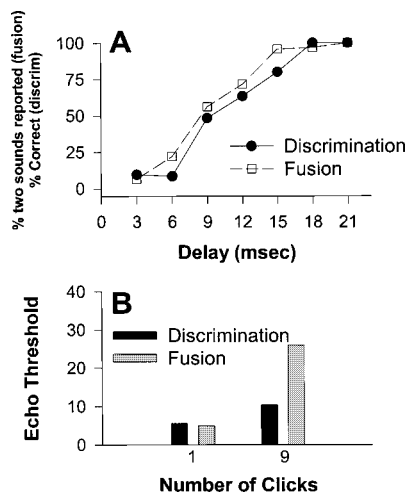


FIG. 5. Discrimination suppression measured in free field is similar to fusion under some conditions, but not others. A: Listeners report how many sources they hear (fusion; squares) or discriminate between two possible lag locations (discrimination suppression; circles). Data were obtained using one pair of 4-ms noise bursts. Results are highly similar for the two precedence tasks (replotted from Freyman *et al.*, 1991, with permission). B: Echo thresholds (delays at which discrimination performance reached 75% correct, or at which listeners reported hearing “two sounds” on 75% of trials) are shown for 1-click and 9-click conditions (Replotted from Yang and Grantham, 1997a, with permission).

A second experimental paradigm in free field has been to measure discrimination of a fixed change in the lag location while varying the delay. This method was originally developed by Freyman *et al.* (1991) in an attempt to link discrimination suppression to fusion. Figure 5(A) shows results from Freyman *et al.* (1991) which suggest that, under some conditions, echo threshold obtained in the fusion task is similar to the thresholds obtained with the discrimination task (defined as the delay for which  $d'$  equals 1). Freyman and colleagues postulated that, if listeners can subjectively “hear out” the lag as a separate sound then they should be able to extract directional information from the lag. Conversely, if the lag is fused with the lead then discrimination performance should be at chance. It turns out that this relation between fusion and discrimination holds true under specific conditions which depend on the locations of the stimuli, the number of stimuli presented within a given trial, and possibly other variations not studied.

There are two very interesting parameters that suggest a decoupling between auditory mechanisms mediating fusion and discrimination. The first parameter is the number of clicks presented within each trial. It appears that while the strength of fusion (measured with echo thresholds) increases substantially as the number of lead-lag pairs increases (further explored in Sec. II), discrimination suppression is less dependent on this stimulus variable [Yang and Grantham, 1997a; see Fig. 5(B)]. This effect relates to the influence of prior stimuli which are discussed in more detail in Sec. II.

A second situation in which there appears to be decoupling between fusion and discrimination is one in which stimuli are presented in the median-sagittal plane. As was discussed earlier, fusion is observed at nearly identical delays in the azimuthal and median planes, where binaural and monaural spectral cues, respectively, dominate localization.

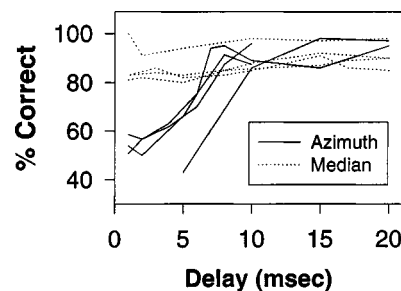


FIG. 6. Discrimination suppression in the azimuthal (solid) and median sagittal (dotted) planes for four listeners. Percent correct is plotted as a function of lead-lag delay (from Litovsky and Colburn, 1998; after Litovsky *et al.*, 1999).

In contrast, discrimination suppression has not been found in the median plane. Figure 6 shows results from a recent study by Litovsky *et al.* (1999), who compared discrimination performance in the two planes. Consistent with previous findings, in the azimuthal plane performance is poor at short delays and improves dramatically by 5–10 ms (e.g., Freyman *et al.*, 1991; Clifton *et al.*, 1994; Yang and Grantham, 1997a). In contrast, in the median-sagittal plane, performance is roughly independent of delay. When asked to describe which auditory cues were used during the task, all subjects reported that in the azimuth the only reliable cue was lateral movement of the sound, but in the median plane perceptual changes in the pitch of the fused auditory event were quite obvious and reliable. The authors postulated that in the median plane listeners were able to rely on cues provided by directional filtering by the pinnae, which are known to be important for identifying the elevation of sources (Fisher and Freedman, 1968; Blauert, 1969; Butler, 1969; Shaw, 1974; Gardner and Gardner, 1973; Middlebrooks and Green, 1991). Given the nature of the objective lag-discrimination task, where feedback is provided and listeners can use whatever cues are available to them, the delay dependence may only be robust when binaural cues are being varied, such as in the azimuthal plane in free field or using ITDs and ILDs under headphones. In fact, a recent report (Freyman *et al.*, 1998) suggests that also in azimuth, listeners are quite sensitive to various aspects of the lag, including its intensity and spectral content. Taken together, these findings are consistent with the notion that although we are usually not actively aware of reflections, we remain sensitive to information carried by these sounds. This sensitivity can be useful for enhancing certain information carried by the source and for a person’s awareness of room acoustics.

#### D. Cross-frequency effects and uncorrelated lead-lag

Most studies on precedence have used idealized short-duration stimuli such as clicks. However, it is clear that in order to understand how reflections are processed by the auditory system, more realistic stimuli should be used. For instance, one could ask whether the reflection must be an identical copy of the source, which can be accomplished by using stimuli with different spectral contents or that are noncorrelated. Several such studies have been conducted, and taken together, they suggest that all aspects of the precedence ef-



fect are strongest when the lead and lag are identical, but to some extent they also operate when the lag is not an exact replica of the lead. In fact, it would be strange if precedence failed completely unless the lead and lag were identical. In most reverberant rooms the acoustics are such that some frequencies contained in the source are reflected while others are absorbed; hence, reflections are rarely identical to the source. However, realistic reflections can only be comprised of frequencies originally contained in the source, thus one might expect noncorrelated lead and lag signals to yield weak precedence.

To our knowledge, only one study has been conducted on cross-frequency effects using the fusion paradigm. Perrott *et al.* (1987) compared performance in free field for broadband noise bursts (50-ms duration) that were correlated to various gradations, and showed that fusion is somewhat weaker when the lead and lag consist of uncorrelated tokens of noise than when the tokens are correlated. In the uncorrelated condition, listeners reported hearing two sounds on most trials, regardless of delay; at delays below 10 ms there was a small proportion of trials on which one sound was reported. In contrast, in the correlated condition listeners experienced the “classic” fusion phenomenon; at short delays they reported hearing one sound on the majority of trials, and at longer delays they reported two sounds on the majority of trials.

The effects of the relative frequency of lead and lag components have been explored in more detail using the localization dominance and discrimination suppression paradigms. Scharf (1974) measured localization for a pair of tones presented from 45° to the right and left in an anechoic room. When the tones were the same frequency, a single fused image was heard from the leading speaker. As the frequency difference increased, the effect weakened, but was still present for tones differing by 1900 Hz. Blauert and Divenyi (1988) used virtual sound stimuli to show that the amount of lag-discrimination suppression increases as the overlap between the lead and lag spectra was increased. Like the fusion study noted above, this result is consistent with the notion that naturally occurring reflections (that contain more spectral overlap with the source) are more likely to be suppressed than unrelated stimuli. Similarly, in a headphone study of ITD discrimination, Divenyi (1992) found that lag-discrimination suppression is stronger if the lead frequency is lower than the lag. Divenyi fixed the lag frequency at 2000 Hz, varied the lead frequency from 500 to 3000 Hz, and found that sensitivity to lag ITD was poorest for the lowest lead frequencies. In fact, when the lead frequency was 2000 Hz or greater, there was little effect on performance, with ITD discrimination thresholds less than 100  $\mu$ s. In contrast, lead frequencies of 500 or 1000 Hz resulted in lag ITD discrimination thresholds as high as 400  $\mu$ s. This finding is somewhat unexpected if one assumes that precedence should be strongest for lead-lag stimuli that are similar in spectral content. However, Divenyi interprets his findings to suggest that discrimination suppression depends on the relative localizability of the lead and lag, rather than their spectral similarity. The lower-frequency lead stimuli are more easily localized than the higher-frequency stimuli, hence they

produce stronger suppression of the directional information contained in the lag.

Similar results have been reported by others. Yang and Grantham (1997b) studied discrimination suppression for lead and lag stimuli that differed in their spectral overlap in both the free field (in an anechoic chamber) and with headphones (by manipulating interaural time differences). They showed that the parameters affecting discrimination suppression appear to be different in the two listening environments. In the free field, the amount of spectral overlap between the lead and lag stimuli was most effective in determining the amount of discrimination suppression as was found by Blauert and Divenyi (1988). However, over headphones, low-frequency information seemed to dominate high-frequency information even when there was little or no spectral overlap between the lead and lag stimuli. This dominance of low frequencies over high frequencies in lag-discrimination suppression is consistent with the fact that thresholds for processing interaural time differences are lower for low frequencies compared with those for high frequencies.

Using the pointer adjustment method with narrow-band noise bursts, Shinn-Cunningham *et al.* (1995) measured the extent to which the ITD in the lag influenced the perceived ITD of the fused image, and found that a high-frequency lag (1250 Hz) received almost no perceptual weight in the presence of a low-frequency (450 Hz) lead, but a low-frequency lag was perceptually weighted equal to a high-frequency lead. To test Divenyi’s hypothesis that these findings arise from differences in the “localization strength” of components with different frequencies, Shinn-Cunningham *et al.* (1995) balanced the low- and high-frequency components by centering the image of simultaneously presented components with adjusted levels. The components were equally effective in centering an image when the level of the high-frequency component was greater than the level of the low-frequency component. When the levels of the lead-lag stimuli were adjusted in this way the asymmetry in discrimination was eliminated and the weight of the lead was the same for both cases.

The emphasis on the lead stimulus seen in localization results can also be seen in results related to the interaural cross correlation (Aoki and Houtgast, 1992). In measurements of the width or the compactness/diffuseness of the sound image, this study demonstrated that the dominance of the lead is the same for localization and diffuseness. These results suggest that the precedence effect, at least under some conditions, is not exclusively one of localization or lateralization *per se*, but rather a more general effect of reduction in the binaural information available in the lag.

These data suggest that suppression of the spatial information contained in the lag is strongest when the lead and lag are spectrally similar. However, there are still many unknowns about how the three measures of precedence depend on the spectral similarity between the lead and lag stimulus. Since real-world reflections are spectrally correlated, but not identical, additional information about precedence and spectral differences is necessary with the originating sound

source in order to fully understand sound processing in reverberant environments.

### E. Summary

We have reviewed psychophysical literature on perceptual phenomena which are based in single pairs of stimuli and which are thought to be related to the precedence effect. While fusion and localization dominance probe listeners' subjective impressions of how many auditory events are heard and what their perceived positions are, lag-discrimination suppression probe the extent to which aspects of the lag or lead are difficult to access due to the presence of the other. The discrimination studies avoid a problem inherent to fusion and localization dominance studies: the subjective nature of the tasks required of the listeners. The fusion and localization dominance tasks measure a perceptual "impression," for which there is no objectively correct answer. Although subjective studies yield valuable information regarding auditory processing, the experimenter must be concerned with changes over time in listeners' criteria. On the other hand, discrimination experiments must be evaluated with full awareness that any of the multiple, changing aspects of the perceptions may be used as a cue. Thus performance may be related to timbre changes, for example, and not to spatial attributes directly.

Over the years, several attempts have been made to explain the time-course of precedence phenomena and to identify their locus in the auditory pathway. McFadden (1973) described a neural network that heavily weights the location of a sound source in the presence of reflections by inhibiting neurons that carry directional information regarding the opposite hemifield. Although this network does not take account of results in the median plane, it emphasizes the possible role of inhibition of directional information carried by the lag. The temporal window of precedence, which lasts for approximately 1–8 ms for clicks, was elegantly described by Harris *et al.* (1963) as a

"...neural gate previous to the place of binaural interaction that closes about 1 ms after the first neural response, permitting no further neural timing signals to be sent to the brain. This gate would reopen two, or slightly more, milliseconds later. A mechanism for such a neural gate could be self-inhibition and inhibition of lateral nerves by the nerves firing on the initial stimulus. This inhibition would have an inherent delay that would permit nerves to be fired in an interval of about 1 ms after the initial neural response. The inhibition would also last for about 2 ms." (p. 677)

This postulated inhibition might effectively reduce sensitivity to directional information, but not to other cues such as timbre, pitch, and loudness.

## II. EFFECTS OF PREVIOUS STIMULATION

While most real-world sound sources and reflections occur only once or a few number of times, recent research on measures of precedence involving repeating the lead and lag stimuli several times have revealed some unexpected results.

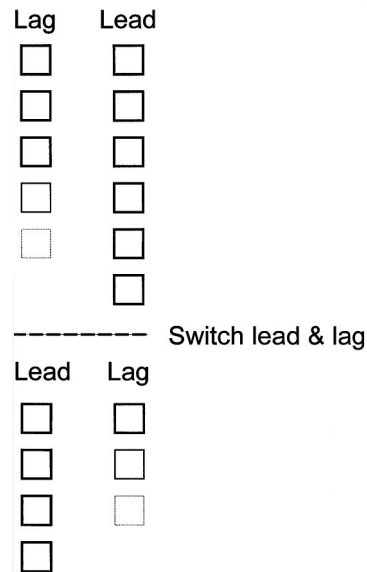


FIG. 7. This figure is a schematic cartoon describing the buildup and the breakdown of fusion. The first presentation of the lead and lag (at top) may not cause fusion leading to the perception of both stimuli. After several presentations fusion occurs. Following the switch of the spatial location of the lead and lag stimuli, fusion breaks down and is reestablished after the lead and lag stimuli are repeated at their new spatial locations.

These results reveal that the measures of precedence are dependent on listeners' immediate prior experience with the lead–lag presentations. If such experience is important for processing sound sources and their reflections then this implies that reflection processing may be more complicated than had previously been thought.

These recent data show that several measures of precedence change after several presentations of the lead–lag stimuli (buildup). For instance, after several repetitions of the lead and lag stimuli, listeners perceive a subjective "fading out" of the lagging stimulus (Clifton and Freyman, 1989). If, after the fade-out some parameters of the lead and lag are suddenly changed, such as their relative locations, then the lag reappears and is perceived again. Fusion is reestablished if the lead and lag stimuli are repeated several more times at their new spatial locations (Clifton, 1987). A schematic diagram of this phenomenon is displayed in Fig. 7. These findings suggest that aspects of the lag become more suppressed with number of presentations (buildup of precedence), and are no longer suppressed after certain parameters are switched (breakdown of precedence). These recent results are reviewed and the implications discussed below.

### A. The buildup effect

Recent experiments have shown that, not only is echo threshold influenced by properties such as stimulus type, duration, and the spectral makeup of lead–lag stimuli, but short-term previous experience with a stimulus also influences echo threshold. In particular, it has been found that if a lead–lag click stimulus is repeated over and over again (referred to as a click train), echo threshold for the last lead–lag

stimulus is raised by several milliseconds (Thurlow and Parks, 1961; Clifton and Freyman, 1989; Freyman *et al.*, 1991), with the most important parameter being the number of click-pairs in the train rather than the train duration or the click rate (Freyman *et al.*, 1991). The slopes of echo threshold as a function of number of click-pairs in the train show the sharpest increase in threshold between 1–5 click-pairs, and an asymptotic value by 12 click-pairs, suggesting that echo threshold becomes stable following the acquisition of a certain amount of information about the lead–lag stimulus (Clifton and Freyman, 1997). This increase in echo threshold suggests that some sort of adaptation occurs over repetitions, such that listeners become less sensitive to reflections. In terms of fusion, lagging clicks are heard at first and then perceptually seem to fade away with stimulus repetitions. In terms of discrimination suppression, the time delay needed to discriminate the lag click increases as multiple repetitions of the stimulus occur. In terms of localization dominance, the image would be expected to move toward the leading source location, although measurements have not been made. This phenomenon has been called buildup of precedence.

Surprisingly, whether the lead stimulus originates from the left or the right side also affects the amount of buildup seen. This was first reported as a finding by Clifton and Freyman (1989), in which it was found that buildup was stronger when the leading click originated from the right side rather than from the left. Grantham (1996) pursued this notion and found that for a single noise burst no difference in discrimination suppression was seen between conditions in which the leading stimulus originated from the left or from the right side. If a judgment followed a train of lead–lag pairs (i.e., after buildup occurred), however, much more discrimination suppression was found on the right side than on the left side. Asymmetry in buildup has been suggested as evidence that more central brain mechanisms are involved in this phenomenon (Clifton and Freyman, 1989; Grantham, 1996).

Another stimulus characteristic which has been found to be important in buildup is the number of lagging clicks presented. Rather than presenting a simple lead–lag stimulus, Yost and Guzman (1996) presented multiple lags at different delays following the lead portion of the stimulus. Yost and Guzman found that, when the stimulus was a simple lead–lag click train (that is, only one lagging click was present), the time delay needed to see a buildup of fusion had to be much shorter than when two lag portions were present. For instance, no buildup of fusion occurred for a 12-ms separation between the lead and lag. However, if two lag clicks were presented, so that the first lag occurred 6 ms after the lead click and the second lag occurred 12 ms after the lead click (the same time separation as for the lead–lag stimulus), buildup was seen. This finding suggests that there is an interaction between the time separation and number of lagging clicks present.

Recent work suggests that the amount of buildup depends upon the type of task a listener is asked to perform (Yang and Grantham, 1997a). Although earlier studies have used both fusion and discrimination suppression tasks virtually interchangeably to measure buildup (Freyman *et al.*,

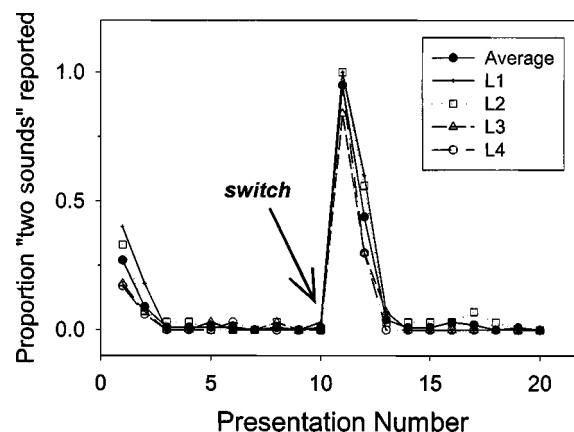


FIG. 8. The percent of times that “more than one stimulus” is reported as a function of the number of presentations of the lead and lag stimulus separated by 6 ms. Between the 10th and 11th presentation the locations of the lead and lag were reversed and remained reversed for the remainder of the presentations. Listeners indicated how many sound sources they perceived for the last lead–lag presentation for each repetition condition. When there are three or fewer presentations, listeners occasionally report perceiving more than one sound source. For additional lead–lag presentations only one sound source is reported (buildup). After the switch, most of the time more than one sound source is reported (breakdown of fusion). Then additional presentations yield reports of one sound source (buildup again). (Adapted from Yost and Guzman, 1996, with permission.)

1991; Clifton *et al.*, 1994), Yang and Grantham suggest that fusion and discrimination suppression may be independent phenomena. They found that for a single lead–lag pair, performance on these two tasks was similar. But, with a train of bursts, fusion increased far more than discrimination suppression. This finding suggests that these two different aspects of precedence, fusion and discrimination suppression, may be governed at least to a certain extent by different mechanisms.

Yang and Grantham (1997a; see Fig. 5B) have shown that when the lead–lag click-pair is preceded by a train of 9 click-pairs there is a strong dissociation between fusion and discrimination. Fusion echo thresholds increase by 15–20 ms. In contrast, discrimination echo thresholds are less affected by the preceding stimulus, and increase by 3–8 ms. It should also be pointed out that a change in the threshold for discriminating an interaural parameter or the spatial location of the lead and lag stimuli in these experiments does not necessarily mean that the lead or lag stimuli were heard as separate signals (i.e., that fusion did not occur). The discrimination can be based on the changes in the fused image such as the diffuseness or spatial extent of the perceived sound image, or lateral position of the fused event (see Yang and Grantham, 1997b).

## B. The breakdown of precedence phenomena

It has been found that buildup of fusion or of discrimination suppression can be interrupted by presenting certain kinds of changes to the lead and lag portions of the stimulus, resulting in a dramatic decrease in echo threshold (cf., Fig. 7). This interruption of buildup is often called a “breakdown” or “release from suppression.” If the locations of the lead and lag suddenly switch, so that the lead now occurs where the lagging stimulus had been and vice versa (referred



to as a reversal switch), fusion is disrupted and listeners hear both clicks. An example of this effect is shown in Fig. 8 with recent data from Yost and Guzman (1996). This release from suppression due to a switch in location was first reported by Clifton (1987) and has thus become known as the “Clifton effect” (Moore, 1996; Blauert, 1997). Other changes to the stimulus have also been shown to disrupt buildup. Clifton *et al.* (1994) found that if the time delay between the lead and lag of the last lead–lag pair was changed, discrimination suppression was reduced and listeners could detect the location of the lagging click more easily. Yost and Guzman (1996) found that only a particular kind of change in location results in a breakdown. If the lead and lag clicks are simply shifted one loudspeaker to the right (referred to as a lateral switch), there is no breakdown in fusion.

Hafter and his colleagues (see reviews in Hafter *et al.*, 1988; Hafter, 1997; Yost and Hafter, 1987) measured the ability of listeners to detect an interaural time difference (and in some experiments, an interaural level difference; see Hafter *et al.*, 1983) of a train of clicks. They consistently found that the early clicks in the click train contribute more than later clicks in the discrimination of a dichotic click train with an interaural difference than from a diotic click train without interaural differences. As one would expect, the more clicks there are in the click train (up to some limit), the lower the interaural difference thresholds become. However, for short temporal separations between the clicks in the train, the later clicks contribute less information than the earlier clicks. They developed a model which assumes the auditory system adapts to the early arriving interaural information, such that the early arriving binaural information dominates. While most of their work has been done using headphone delivered stimuli, Hafter *et al.* (1992) demonstrated that similar effects occur in an anechoic room. These results are like many cited in the precedence literature (see Hafter, 1997) in that spatial processing is dominated by early arriving information (the precedence effect), later samples have less binaural information than earlier samples, and there is a buildup of binaural information over repeated events.

Hafter and Buell (1990; see also Hafter *et al.*, 1988 and Hafter, 1997) studied stimulus conditions involving click trains and interrupted click trains. These experiments share many similarities to those measuring for the breakdown of precedence. For instance, suppose that a train of  $N$  clicks has an interaural time discrimination threshold of approximately  $60 \mu\text{s}$ . If the train of clicks is made half as long, i.e.,  $N/2$  clicks long, the threshold can increase (for instance, to about  $70 \mu\text{s}$ ) since a short click train has less useful binaural information than a longer click train. However, if halfway through the long click train (after the  $N/2$  click) noise is added, or something else (even silence) interrupts the train momentarily, the interaural-time difference threshold for the interrupted, long click train decreases to  $50 \mu\text{s}$ . This indicates that the information in the interrupted click train is more effective than in the continuous train. It is as if the interruption caused the binaural system to restart its adaptation process and the click train is processed as two sets of  $N/2$  clicks, rather than a long train of  $N$  clicks. This is similar to the Clifton effect, in which an interruption in the flow of

repeated clicks causes a restart of the system that governs the way in which interaural information is associated with repeating click events.

### C. Possible consequences of the buildup and the breakdown of precedence

Several ideas regarding processes involved in the precedence effect have emerged based upon findings of buildup and breakdown. The breakdown of suppression, especially the breakdown of fusion, suggests that the auditory system maintains information about the lag (reflection) stimulus, even when fusion occurs and localization is dominated by the lead stimulus. That is, for the early click events in a train of repeating clicks, the lag appears to be completely fused with the lead, and source location is dominated by that of the lead. However, when certain types of “switches” or inconsistencies involving the lead and lag take place, the lag is perceived as separate with its own perceived location. Thus it appears as if information about the lag is not eliminated, but rather that certain information about the lag is suppressed. Certain changes in the acoustic environment then release this suppression.

Phenomena such as left–right asymmetry and the relatively long time course over which buildup and breakdown occur suggest to some researchers that buildup and breakdown are governed by *central* brain processes (Clifton and Freyman, 1997; Grantham, 1996). It has been suggested that breakdown occurs with changes to the stimulus that are incompatible with a listener’s immediate previous experience in an acoustic environment. Blauert and Col (1992) hypothesized that the precedence effect is at least partially controlled by processes in higher centers of the nervous system, and suggest that a model such as one involving a pattern recognition process may be appropriate to account for the listener’s ability to selectively listen to the lagging source under certain conditions. Rakerd and Hartmann (1985) propose what they have termed the “plausibility hypothesis.” According to this theory, interaural parameters are assessed by listeners to determine their plausibility given the information listeners have about the environment (for instance, visual images). Thus interaural variables deemed to be implausible (such as a very large interaural difference of time) are discounted. Clifton and colleagues (e.g., Clifton *et al.*, 1994) suggest a similar hypothesis, in which reflections provide information about a listener’s acoustic environment. Based on this (previous) information, listeners form “expectations” about the sounds that can occur given the acoustic environment. If a sound does not follow their expectations (such as in the case of a reversal switch), listeners reevaluate the acoustic stimulus, causing breakdown of fusion and/or of discrimination suppression. It should be noted that to date there is no evidence that either buildup or breakdown are learned effects or that they can be modified by practice (Clifton and Freyman, 1997).

The fact that buildup and the breakdown of precedence occur for several different conditions has been used to suggest that reflection processing relies on central mechanisms such as those that are involved with cognition (see Blauert, 1997). While this is one possible way to theorize about pre-

cedence mechanisms, data may be too sparse to resolve one theory over another. For instance, if buildup and breakdown of precedence cannot be modified by practice or other conditions that control learning, then many theories of cognition would not be applicable for dealing with these precedence effects. Precedence buildup and breakdown suggest that we do not fully understand the basic processes governing precedence, especially fusion. If fusion changes as a function of immediate prior experience with the lead and lag stimuli, what aspects of the lead and lag (or the environment in which the lead and lag occur) cause fusion? A full understanding of reflection processing in real-world situations requires knowledge of precisely which aspects of the lead and lag alter the effects of precedence.

### III. ANIMAL BEHAVIOR AND HUMAN DEVELOPMENT

Early behavioral-ablation studies showed that unilateral ablation of the auditory cortex severely impairs the ability of cats to accurately choose the hemifield containing the leading source, while the ability to find the hemifield containing a single-source sound was not affected (Cranford *et al.*, 1971; Whitfield *et al.*, 1972). Performance was consistent with elimination of lead dominance in localization on the side of the lesion. The deficit was limited to cases in which the leading source was in the contralateral hemifield (opposite to the lesion site). For instance, if the left auditory cortex was ablated, left–right lead–lag pairs of tone pulses (23-ms delay) were localized at the leading source but right–left lead–lag stimuli were not localized consistently. Later studies (Cranford and Oberholtzer, 1976; Whitfield, 1978) found this effect in some but not all animals. These differences have been attributed to differences in the extent of the lesion, individual differences in behavioral learning strategies amongst cats, and differences in testing procedures.

Behavioral correlates of localization dominance and fusion have been found in several animals at delays that are similar to those reported for humans. Cranford (1982) made psychophysical measurements of localization dominance in normal animals. Click trains were presented with varying lead–lag delays (from 0.2 to 9 ms) either with a speaker on the left leading a speaker on the right, or vice versa. Cats were trained to release a foot pedal on the right if they heard a sound on the right, and another foot pedal on the left if they heard a sound on the left. Cats were only rewarded if they released a foot pedal corresponding to the side of the leading click. (For the case of 0-ms delay, cats were rewarded for releasing either foot. This symmetric case was run in separate sessions.) Mean results from six cats tested in the Cranford study and from studies in rats and humans are shown in Fig. 9. In the Cranford study (circles), all animals responded mostly to the leading side at lead–lag delays of 0.5 to 2 ms. At longer delays some cats still responded only to the lead, while others responded to both lead and lag. These results for cats are very similar to the human localization-dominance results (Litovsky *et al.*, 1997b) plotted in the same figure. It appears that the delays at which cats seem to show localization dominance are about the same as the delays for which human listeners experience localization dominance and fusion.

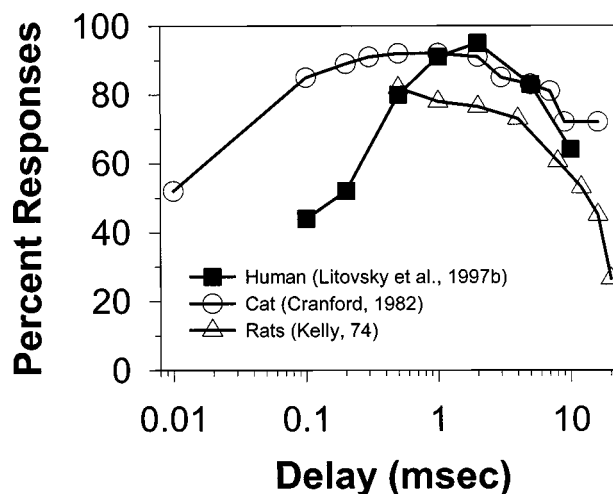


FIG. 9. Behavioral measurements in animals. Data for three species are compared: humans (square; from Litovsky *et al.*, 1997; replotted with permission), cats (circle; from Cranford, 1982; replotted with permission) and rats (triangle; from Kelly, 1974; replotted with permission). In each case performance is plotted as a function of lead–lag delay. Human data are azimuthal data replotted from Fig. 5; percent of trials in which listeners reported that the sound was located at the lead speaker are plotted. Cat data represent the percent of trials in which cats oriented toward the leading speaker. Rat data represent animals’ performance (rated from 0%–100%) score for discrimination of right–left (lead–lag) trials from left–right trials.

Other studies reported similar findings in rats (Kelly, 1974, Fig. 9, triangles) and crickets (Wytenbach and Hoy, 1993). Finally, Keller and Takahashi (1996) measured sound localization in owls for lead–lag stimuli. Since the owls were asleep or not appropriately positioned on 90% of trials, the behavioral data are scarce in seven of the ten animals tested. In three animals, at lead–lag delays of 2 to 10 ms the owls turned their heads consistently toward the leading speaker on the majority of trials. Most recently, Populin and Yin (1998) used a magnetic search coil technique to measure eye movements in cats whose heads were fixed in place. These cats were able to localize a single-source click or noise burst to within a few degrees. Lead–lag trials included two sources placed at 18° to the right and left. At a delay of 0 ms, cats oriented their eyes toward midline, suggesting that they heard a “phantom” source at that location. As the delays were increased, the eye movements shifted gradually toward the leading speaker, approximating single source by 300  $\mu$ s.

Hochster and Kelly (1981), motivated by the earlier ablations studies in cats, measured localization dominance in children 6–16 years of age with temporal-lobe epilepsy. For stimuli consisting of single clicks, these children were able to identify which of two loudspeakers contained the stimulus. On precedence trials containing lead–lag stimulus pairs with short delays (1–4 ms), both epileptic and normal children were able to identify the leading speaker on nearly 100% of trials. However, the effect was substantially weaker in the epileptic children than normal children or adults, suggesting that the temporal lobe might be functionally involved in mediating localization dominance.

A series of precedence studies in human infants and children and in newborn and young dogs were undertaken, partly motivated by the animal behavior studies (for extensive reviews see Clifton, 1985; Litovsky and Ashmead,

1997). Localization dominance, measured in much the same way as it was in the cats, does not seem to be present in newborn human infants (Clifton *et al.*, 1981), although single-source discrimination occurs within hours after birth (see Clifton, 1985). Localization dominance is first measured in humans at 4 to 5 months of age, but at that age fusion echo thresholds are quite high (25 to 45 ms) compared with adult thresholds measured under the same conditions (8 to 15 ms). Echo thresholds in human infants were measured using a conditioned head-turning task. Lead-lag sounds were presented from loudspeakers located at 90° to the right and left and their relative onsets were delayed by 5–50 ms. At the onset of each trial there was a 7-ms delay, which had previously been shown to produce effective fusion in infants. The delay was abruptly lengthened; when infants heard the lagging sound at its location they responded by turning their heads toward this “novel” sound. Echo thresholds in 5-month-old infants are approximately 26 ms for click stimuli, which is significantly higher than adults’ thresholds of 9 ms. By 5 years of age, adult thresholds are reached for click stimuli, however, for long duration more complex stimuli echo thresholds are significantly higher in children than in adults (Morrongiello *et al.*, 1984). Behavioral studies on localization dominance have also been conducted in young dogs, who do not show any evidence of precedence through the fifth month of life (Ashmead *et al.*, 1986).

Discrimination suppression also undergoes significant developmental changes during early childhood. Litovsky (1997) measured MAAs in free field for the leading source (in the presence of the lag at midline), for the lagging source (in the presence of the lead at midline), and for a single source, at ages 18 months, 5 years, and adult. The stimuli were 25-ms noise bursts and the lead-lag delay was 5 ms. Since this lead-lag delay is below the echo threshold for these stimuli, the image is fused. Single-source MAAs are adultlike (1 to 2°) at 5 years of age and fairly low (5°) by 18 months of age. Lead MAAs are quite low in adults (1.7°), somewhat elevated at 5 years (4.4°), and substantially higher at 18 months (23°). Lag MAAs are still low in adults (1.7°), but substantially higher in 5-year olds (27.5°) and 18-month olds (65°). Lead MAAs reflect listeners’ ability to focus on the first-arriving wave front and to discriminate between leading source locations in the presence of the lagging source, and this ability improves dramatically with age. Lag MAAs reflect listeners’ ability to extract directional information from a sound that is not heard as a separate auditory event; this ability improves somewhat with age but is still quite underdeveloped at 5 years relative to adults. These findings are consistent with the fusion echo threshold data which were obtained using a task that measures children’s ability to localize the lag as a separate sound. In the MAA study, normalizing the lead and lag results by the single-source results maintained the developmental differences observed, suggesting that lead and lag MAAs are not merely the “by product” of a “noisy” single-source discrimination ability. As Litovsky (1997) points out, while the developmental work may point to maturational changes in the central auditory pathway, attentional and learning processes cannot be ruled out.

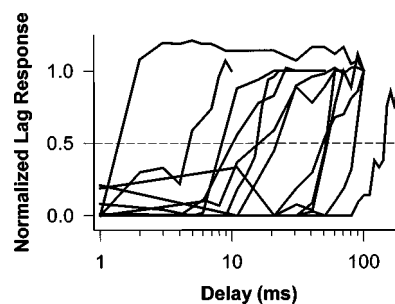


FIG. 10. Examples of recovery functions of neural half-maximal ISDs are plotted. For all neurons, both leading and lagging stimuli were at locations or ITDs which elicited robust responses with single stimuli. The recovery functions represent the normalized lagging responses as a function of ISD. (Replotted from Litovsky and Yin, 1998a, with permission.)

#### IV. PHYSIOLOGICAL CORRELATES OF PRECEDENCE

In recent years there has been a surge of activity exploring physiological substrates of precedence. Two decades ago (Altman, 1975), and more recently (Carney and Yin, 1989), there were suggestions made that such a substrate might be found in responses of single neurons in the central nucleus of the inferior colliculus (IC). Using stimuli that consisted of a single click to each ear with unusually long ITDs (tens of milliseconds), Carney and Yin (1989) found that the response to the lagging click was suppressed in most cells, even when the leading click did not elicit a response. In addition, at very short delays (less than 1 ms), the lagging click often produced a backward suppression of the leading click. Carney and Yin (1989) hypothesized that while the long-lasting forward suppression might reflect a neural correlate of echo suppression, the backward suppression may be important for summing localization.

Several years later, Yin (1994) tested that hypothesis by presenting stimuli from two locations (or with different ITDs) during the time period associated with summing localization (less than 1 ms), when listeners hear a fused auditory event at a phantom location near the leading speaker. Several IC neurons gave responses consistent with summing localization: as the lead-lag delay increased from zero, the responses to the lead-lag combination followed the response of the neuron to ITDs in between the lead and lag delays, progressing with increasing delay to the response expected from the lead alone. Yin (1994) suggested that the discharge rates of these neurons might be related to the location at which the animal would perceive the phantom sound source. The same study also explored neural responses to lead-lag stimuli at longer delays. Note that IC neurons are usually sensitive to specific ITDs and azimuthal locations, hence measurements were consistently made for lead and lag locations at each neuron’s “best” azimuth or ITD. At short delays most neurons responded only to the leading source, and as the delays were increased the lagging response recovered from suppression, resulting in recovery curves much like the fusion psychometric functions seen in human listeners. For the 65 cells studied, the half-maximal delays (at which the lagging response recovered to half of the nonsuppressed response) ranged from 1 to 100 ms with a median of 20 ms.



Similar results were reported by Litovsky and Yin (1998a) for a population of 94 neurons studied in free field [e.g., Fig. 3(A)] and with dichotic stimuli [e.g., Fig. 3(B)]; half-maximal delays ranged from 1.5 to 154 ms with a median of 27 ms. Figure 10 shows normalized lagging responses (lag response at each delay divided by lag response at maximal delay) from a sample of 12 neurons in the IC. The data suggest that all neurons show strong suppression at the shorter delays tested and no suppression at long delays. In addition, there is substantial variability in delays at which IC neurons exhibit suppression of the lagging response. (In fact, this variability is also apparent in Fig. 12, where IC data are compared with data collected at other levels in the auditory pathway.)

Seeking further physiological correlates of precedence, Litovsky and Yin (1998a, 1998b) measured responses of IC neurons using stimulus parameters that are known to influence echo thresholds in human listeners. They reported a tendency for suppression to last longer at lower overall stimulus levels, which is consistent with psychophysical results for localization dominance (e.g., Shinn-Cunningham *et al.*, 1993). Suppression was also stronger for noise bursts than for clicks, for long-duration noise than for short-duration noise, and when the leading level was increased, all of which are consistent with psychophysics (see Blauert, 1997, and Table I). The most striking correlate was obtained for stimuli presented either in the azimuthal plane or in the median-sagittal plane (see also Litovsky *et al.*, 1997b). Comparisons were made by positioning the lagging source at 0° azimuth–0° elevation (a location common to both planes), and the leading source at locations along the azimuth and median planes that produced similar discharge rates. For the 39 neurons studied there was a high correlation in half-maximal delays ( $r=0.8$ ) for the two planes. These results are consistent with the psychophysical measurements of fusion (Litovsky and Colburn, 1998) and localization dominance (Litovsky *et al.*, 1997b; see Fig. 4) in the azimuthal and median planes. The combined physiological and psychophysical findings suggest that fusion is mediated by the same neural mechanisms regardless of whether binaural disparity cues (azimuth) or spectral cues (elevation) are prominent. As was discussed above (Sec. IB 3), Litovsky *et al.* (1997b) have argued that models of fusion which assume that interaural delays are an integral aspect of the precedence effect (Lindemann, 1986a; Franssen, 1963; Shinn-Cunningham *et al.*, 1993) do not address the similarity of fusion for stimuli in vertical and horizontal planes. Models which integrate spectral cues are yet to be developed.

The studies discussed thus far were conducted in barbiturate-anesthetized animals. The potential problems in data interpretation stems from the fact that barbiturates are thought to enhance physiological inhibition in the central auditory pathway (Kuwada *et al.*, 1989). Responses of IC neurons to lead–lag stimulus pairs have also been measured in awake rabbits (Fitzpatrick *et al.*, 1995) and owls (Keller and Takahashi, 1996). In the awake rabbit, half-maximal delays in 55 neurons (using dichotic click stimuli with both lead and lag at the neuron's "best" ITD) ranged from 1 to 64 ms with a median of 6.3 ms, which is somewhat lower

than the values obtained in anesthetized cats. In the awake barn owl (*Tyto alba*), Keller and Takahashi (1996) measured responses of 51 neurons in the external nucleus of the inferior colliculus (ICx) to pairs of noise bursts (3 or 100 ms) with lead–lag delays of 0.5 to 5 ms. These neurons, which are excited maximally by stimuli presented from specific locations in space, responded only to the leading stimulus at these short delays. While some neurons' lagging response recovered by 5 ms, others did not. Behaviorally, at these same delays the owls only turned their heads toward the leading source (see Sec. III above).

As several authors point out (e.g., Fitzpatrick *et al.*, 1995; Litovsky and Yin, 1998a), it is difficult to determine whether the differences stem from species differences or the effect of barbiturates. One of the challenges in trying to relate neural evidence of echo suppression to behavioral echo thresholds, is that for brief stimuli a large proportion of neurons recover from suppression at delays that extend far beyond behavioral echo thresholds. Behavioral echo thresholds for brief stimuli are usually reached by 5 or 10 ms in humans and other animals. However, the proportions of neurons with half-maximal delays less than 10 ms ranged from approximately 15% (Yin, 1994) to 22% (Litovsky and Yin, 1998a) in anesthetized cats, and reached over 30% in awake rabbits (Fitzpatrick *et al.*, 1995). It has been argued (Yin, 1994) that perceptual echo thresholds are most likely generated by those neurons with the lowest half-maximum delays in much the same way that behavioral thresholds of pure tones have been related to physiological thresholds of auditory nerve fibers (Lieberman, 1978). It has also been noted that several aspects of precedence, such as accurate localization of the lag at its respective position and equal-loudness perception of the lead and lag, are not released from the influence of the lead until the delays are in the tens of milliseconds, which may account for the neurons with long half-maximal delays (Fitzpatrick *et al.*, 1995; Litovsky and Yin, 1998b).

The extent of suppression in the IC also depends on the location or ITD of the leading stimulus, but the relationship between the lead and lag locations seems to affect neurons in a variety of ways. In a headphone study using dichotic clicks, Fitzpatrick *et al.* (1995) reported that while half of the neurons show longer suppression when the ITD is near the "best" ITD, the other half show stronger suppression with the lead near the "worst" ITD. The "worse" neurons, which never fire in response to the lead, provide the strongest evidence that the suppression observed in ICC neurons is not merely a function of refractoriness, whereby once a neuron has already fired in response to the leading source the probability that it will fire in response to the lagging source is diminished. Litovsky and Yin (1998b) made measurements both dichotically and in free field. In the latter case the lag was held constant at each neuron's "best" location and the lead location was varied along the azimuth in 15° increments. When considering only the neurons whose lead response is modulated with azimuth, they found that the great majority (85%) showed maximal suppression when the lead was at the neuron's "best" location, and a minority showed maximal suppression when the lead was at the neuron's

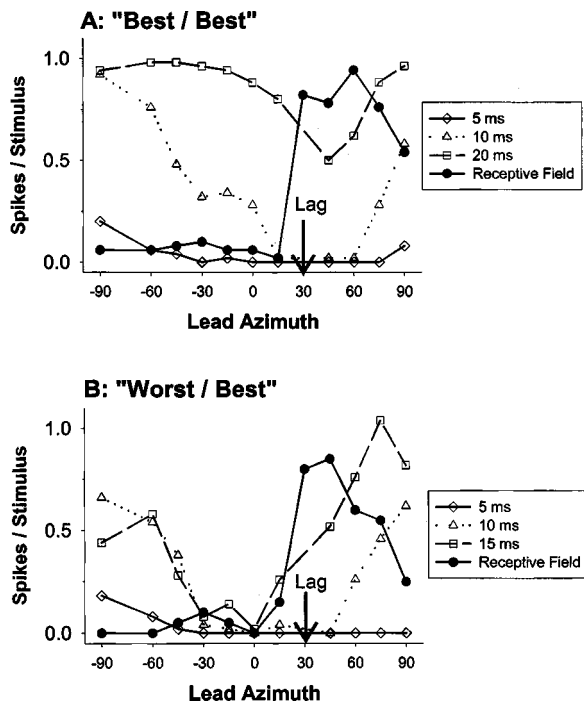


FIG. 11. Examples of neurons with “best” and “worse” lead location suppression. A: Example of a “best/best” neuron where suppression is strongest when the lead is in the neuron’s “best” azimuthal locations (CF=3 kHz). Responses to single clicks are shown in dark circles, representing the neuron’s receptive field. The arrow points to the location of the lagging source (+30°); if no suppression occurs then the lagging response should equal the response of the neuron to a single click at 30°. Responses to the lagging clicks at ISDs of 5, 10, and 20 ms are shown. At 20 ms, suppression is relatively weak and occurs only when the lead is at the best response area. At 10 ms, the suppression spreads out to the neuron’s non-responsive area, and at 5 ms suppression is almost complete, regardless of the lead location. B: Example of a “worse/best” neuron where suppression is strongest when the lead is in the neuron’s “worse” azimuthal locations (CF=1.6 kHz). Suppression for this neuron also increases as the delays are decreased, with almost complete suppression at 5 ms, regardless of the lead location. (Reprinted from Litovsky and Yin, 1998b, with permission.)

“worse” location. Examples of these two neuron types are shown in Fig. 11.

Of course, the full story is only told when all of the neurons in the population are considered, including those that do not show a directionally sensitive response to the lead. In a recent study using “virtual space” stimuli, Litovsky *et al.* (1999) quantified both the amount of response modulation (directional sensitivity) related to the excitation produced by the lead, and the modulation related to the suppression of the lag. In their population of IC neurons, 80% show modulation of both lead excitation and lag suppression, and 20% show modulation of either the lead excitation or the lag suppression. In their study, the directional properties of the lead stimulus were then digitally manipulated such that, as the azimuth of the lead was varied the ILD was held constant while the ITD and spectrum varied naturally. This manipulation resulted in loss of either lead excitation or lag suppression in the majority of neurons (83%). The authors have thus argued that the inputs mediating directional sensitivity to the source (excitation produced by lead) and the reflection (suppression of the lag) may be decoupled, and arrive from different populations of neurons.

Although the physiological recordings discussed above were made in the IC, which is a major site of binaural convergence and interaction in the auditory pathway, the generation of inhibitory effects may not necessarily be in the IC itself. Inputs to the IC arrive in the form of direct monaural input from the cochlear nuclei, indirect binaural input via the superior olive, and multisynaptic inputs via the lateral lemniscus. Auditory studies in the monaural pathway have shown that in response to click pairs with varying lead-lag delays suppression of the lagging response occurs in the auditory nerve (Parham *et al.*, 1996) and ventral cochlear nucleus (Wickesberg, 1996; Parham *et al.*, 1998). In these studies, the half-maximal suppression occurs at around 2–4 ms with full recovery seen for all fibers by 8 ms. In a slice preparation, Wickesberg and Oertel (1990) found putative inhibitory projections from one subdivision of the cochlear nucleus to another, which is most effective at interstimulus intervals of 2 ms; they have thus suggested that intrinsic inhibitory circuits in the cochlear nucleus might mediate a monaural correlate of echo suppression. Several other studies have shown that a neural correlate of forward-masking in response to tones can be seen in the cochlear nucleus (Boettcher *et al.*, 1990; Kaltenbach *et al.*, 1993; Shore, 1995). In the superior olivary complex (SOC), Fitzpatrick *et al.* (1995) found that most neurons recovered by 4 to 8 ms (median 1.9 ms) with monaural stimulation. However, the short time courses of monaural suppression compared with the much longer time course of suppression observed in the IC suggests that much of the suppression is generated at levels above the monaural pathway. In fact, reports of measurements made in the auditory cortex (Reale *et al.*, 1995; Fitzpatrick *et al.*, 1995) suggest that suppression can last for hundreds of milliseconds.

In an attempt to compare the amount of suppression that can be measured at various levels in the auditory pathway, we have replotted data from a number of studies conducted in the following areas: auditory nerve (derived from Parham *et al.*, 1996), antero-ventral cochlear nucleus (derived from Parham *et al.*, 1998), inferior colliculus (replotted from Litovsky and Yin, 1998a) and auditory cortex (from Reale *et al.*, 1995). Figure 12 shows population histograms of half-maximal delays at each of these stages in the auditory pathway. These panels point to the marked increase in suppression as one ascends the auditory pathway, going from monaural circuits to binaural ones. While suppression in the cochlear nucleus is no longer present by 10 ms, in the IC suppression is quite strong at those delays for most neurons, and cortical neurons do not begin to recover before 50 ms. Finally, given the extensive physiological measurements conducted to date, in Table II we summarize half-maximal delays observed at various levels in the auditory pathway, in a number of different species, and with either anesthetized or awake preparations.

Yin (1994) makes a convincing argument that long-lasting suppression observed in the IC is not due to long refractory periods caused by intrinsic neural mechanisms, and is therefore probably due to synaptic inhibition. The IC contains abundant inhibitory synapses, primarily from the lateral superior olive (LSO, e.g., Saintmarie, 1989) and the



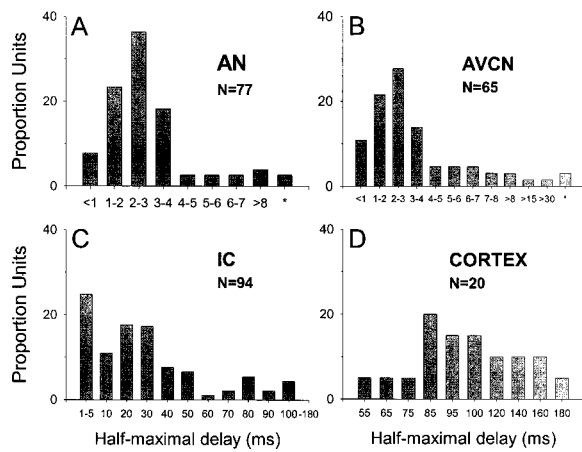


FIG. 12. Population histograms of half-maximal delays for click stimuli measured in cats at various levels in the auditory pathway. A and B: Recordings made in the auditory nerve (AN) and anteroventral cochlear nucleus (AVCN), respectively. In both populations stimulation was monaural and click stimulus level was 85 dB SPL peak *re* 20  $\mu$ Pa. AN data were derived from 77 fibers (with permission from Parham *et al.*, 1996; for two fibers values were not obtainable [\*]). AVCN data were derived from 65 units (with permission from Parham *et al.*, 1998; for two units values were not obtainable [\*]). C: Recordings made in 94 units in the inferior colliculus (IC) by Litovsky and Yin (1998a). Recordings were made using binaural stimulation with ITDs or azimuthal locations of both lead and lag placed in each neuron's maximal response area. Levels ranged from 50–80 dB. Data are reprinted with permission. D: Recordings made in 20 units of the auditory cortex by Reale *et al.* (1995). As in the IC, ITDs of both lead and lag elicited maximal responses. Data provided to the authors by Dr. John Brugge via personal communication, 1999.

dorsal nucleus of the lateral lemniscus (DNLL) (e.g., Adams and Mugnaini, 1984; Moore and Moore, 1987), as well as some that originate within the IC itself (Oliver *et al.*, 1994). Fitzpatrick *et al.* (1995) have postulated that the DNLL projections may play a prominent role in the results seen in the IC. Most neurons in the DNLL are sensitive to ITDs, hence the “best” ITD and “worse” ITD neurons could be invoked by ipsilateral and contralateral activation, respectively. Based on similar assumptions, Cai *et al.* (1998a) have suc-

cessfully modeled both “best” ITD and “worse” ITD neurons, as well as other effects observed by Litovsky and Yin (1998a, 1998b), assuming ITD-sensitive inhibition from the DNLL.

To date, little else is known about the actual physiological mechanisms that mediate precedence phenomena. It is unlikely that perception can be explained solely on the basis of single-neuron responses, hence more sophisticated analyses such as determination of population codes may be necessary. Further, although initial stages of echo suppression may occur in the brainstem, single-neuron results cannot account for all perceptual phenomena related to precedence. For example, Litovsky and Yin (1998a) conducted an analysis which suggests that single neurons do not show signs of buildup. Perhaps the most convincing evidence that the IC may not account for behavioral measurements of precedence comes from a recent study by Litovsky (1998), who showed that neurons in the ICC of newborn cats exhibit the same type of suppressive effects as the adult neurons. This occurs as early as 8–9 days of age, before the full maturation of the structure or function of the auditory system and before the time that a young cat functions behaviorally in its auditory environment. Recall that human infants and young dogs do not exhibit localization dominance or fusion early in life, and that the phenomenon is only observed at 4–5 months of age. Thus, Litovsky (1998) concludes that behavioral manifestation of precedence is mediated at higher levels in the auditory pathway than the IC, which is consistent with the cat lesion studies and the human development studies, all of which suggest that the auditory cortex is essential for precedence to occur behaviorally.

## V. MODELS AND THEORIES RELATED TO PRECEDENCE EFFECTS

There have been a few published models addressed to the precedence effect; however, there is no model currently published that is able to accommodate available data satis-

TABLE II. Physiological studies in mammals.

Recording site	Animal	State	Intra/extra	Stim presentation (Clicks)	50% recovery range (median)	Study
Auditory nerve	cat	decerebrate	xtra-cell	headphones	>1 to 20 ms (2.4)	Parham <i>et al.</i> (1996)
Cochlear nucleus (DCN)	cat	decerebrate	xtra-cell	headphones	>1 to 32 ms (2.7)	Parham <i>et al.</i> (1996)
Cochlear nucleus (VCN)	chinchilla	anesthetized	xtra-cell	headphones	1–2 ms	Wickesberg (1996)
Cochlear nucleus (PVCN)	mouse	slice	intra-cell	shock	2 ms	Wickesberg & Oertel (1990)
Superior olivary complex	rabbit	awake	xtra-cell	headphones	4–8 ms (1.9)	Fitzpatrick <i>et al.</i> (1995)
Inferior colliculus	cat	anesthetized	xtra-cell	headphones	1.5–100 ms (20)	Yin (1994)
Inferior colliculus	cat	anesthetized	xtra-cell	free field	1–90 ms (20)	Yin (1994)
Inferior colliculus	cat	anesthetized	xtra-cell	free-field azimuth	2–183 ms (28)	Litovsky & Yin (1988a, 1998b)
Inferior colliculus	cat	anesthetized	xtra-cell	free-field elevation	1–76 (16.5)	Litovsky & Yin (1998a, 1998b)
Inferior colliculus	cat	anesthetized	xtra-cell	headphones	1.5–110 ms (28)	Litovsky & Yin (1998a, 1998b)
Inferior colliculus	cat	anesthetized	xtra-cell	virtual azimuth	2–80 ms (19)	Litovsky <i>et al.</i> (1998)
Inferior colliculus	rabbit	awake	xtra-cell	headphones	2–60 ms (6.3)	Fitzpatrick <i>et al.</i> (1995)
Inferior colliculus	owl	awake	xtra-cell	free field	1–5 ms	Keller & Takahashi (1996)
Auditory cortex	cat	anesthetized	xtra-cell	virtual 2D free field	48–175 ms (103)	Reale <i>et al.</i> (1995; Abstract)
Auditory cortex	rabbit	awake	xtra-cell	headphones	1–100 ms (20)	Fitzpatrick <i>et al.</i> (1998; Abstract)

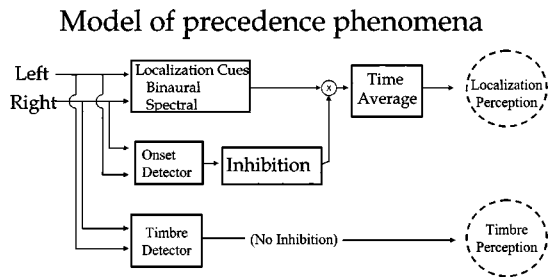


FIG. 13. This model represents a modified version of a model proposed by Zurek (1987) to account for precedence under binaural conditions, in azimuth. In its original form the model contained the localization inputs and an onset detector, which accounted for inhibition of localization information contained in echoes. In its extended form the model also contains a “timbre detector,” which lacks inhibition and allows listeners to attend to spectral information contained in echoes. (Upper path reproduced from Zurek, 1987, with permission.)

factorily. In addition, none of the models can account for phenomena such as the buildup or breakdown of the precedence effect, which are thought to be more cognitive. The discussion here focuses on four models of psychophysical performance (Zurek, 1980, 1987; Lindemann, 1986b; Shinn-Cunningham *et al.*, 1993; Tollin and Henning, 1999), presented in order of their appearance in the literature. As we shall see, these models illustrate fundamentally different approaches and have distinct advantages and disadvantages. In addition, they take a “bottom-up” approach with a focus on relatively peripheral effects. At the moment there is no satisfactory model that encompasses basic aspects of the precedence effect. In addition to these psychophysically oriented models, some models of physiological responses (e.g., Wickesberg and Oertel, 1990; Cai *et al.*, 1998a, 1998b) describe neural responses and mechanisms that may be related to some aspects of the precedence effect and they are briefly noted. Finally, we discuss briefly the general question of how much of the precedence phenomena can be understood from a relative peripheral mechanism point of view and how much a more cognitive model is required.

The first model, proposed by Zurek (1980, 1987), is a phenomenological model that is not designed for quantitative predictions but provides an intuitive representation of the basic observations. This model is illustrated in Fig. 13. The upper path of the model was originally proposed by Zurek (1987), who suggested that stimulus onsets (or onsetlike transitions) initiate an inhibition or suppression process that blocks the generation of location information from the ongoing stimulus for a brief period. This upper path thus represents a key aspect of the precedence effect: that the early part of a stimulus leads to suppressive effects on the later part of the stimulus. However, given additional data collected since 1987, we have modified the model by adding the lower path, which emphasized the fact that only localization/lateralization information is suppressed and not other subjective attributes of the stimulus such as loudness, pitch, or timbre. Although this model provides a useful representation of the basic observations, it does not allow quantitative predictions, nor an internal mechanism for the generation of these effects from the stimulus wave forms. A computational model that incorporated these ideas was analyzed by Martin

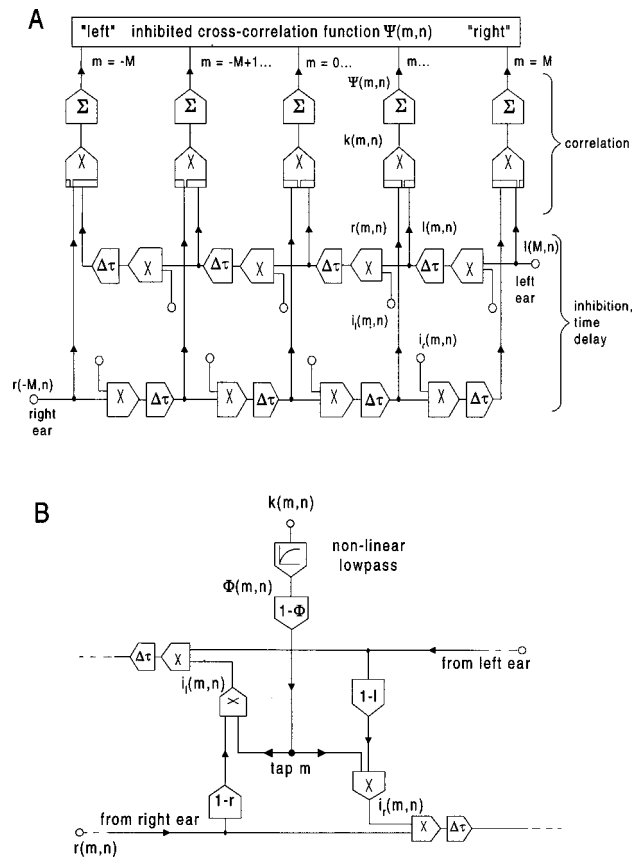


FIG. 14. Inhibited interaural cross-correlation model (Lindemann, 1986b). A: Input signals are represented  $[r(-M,n)$  and  $l(M,n)]$  which are functions of discrete time variables,  $n$ . The delayed and attenuated input signals are represented at location  $m$  as  $r(m,n)$  and  $l(m,n)$ . Input signals are attenuated as they pass along their delay lines, indicated by the symbol  $\chi$ . Attenuation is derived such that it increased with level of the contralateral signal. The output at a given location is roughly the correlation between the left and right signals at a given location in the delay line. Hence, wave forms from each side can cancel each other as they pass along the delay line, and it is possible for strong output to inhibit outputs at other locations. (After Lindemann, 1986b; reproduced from Colburn, 1996, with permission.)

(1997); the model reproduced the basic trends of the data. The general idea of a transient, central inhibitory process that suppressed location information is consistent with the suggestion of Harris *et al.* (1963) as quoted above (end of Sec. I).

The second model, proposed by Lindemann (1986a, 1986b), is an extension of the classic mechanism for sound localization suggested by Jeffress (1948). The Jeffress mechanism, described primarily as a network of coincidence detectors, can be represented as a two-dimensional array of neurons with each neuron's characteristic frequency along one axis and its internal delay along the other axis. Each neuron responds maximally when its immediate inputs are coincident; thus a given neuron responds most vigorously when the stimulus delay is compensated by the neuron's internal delay. For correlated inputs, therefore, interaural delay is translated to a place of maximal stimulation for each frequency. Alternatively, the neural response versus internal delay can be thought of as a cross-correlation function of the bandpass-filtered input wave forms for each characteristic frequency. [For a general discussion of models of binaural

processing and the role of the Jeffress-type network in these models, see the review papers of Stern and Trahiotis (1995) and Colburn (1996).] Lindemann (1986a) extended the correlation mechanism to include lateral inhibition along the delay line (see Fig. 14). This inhibition, which includes a static and a dynamic component, suppresses the propagating wave form across the delay lines so that outputs away from the current site of coincidence output are reduced. For stationary stimuli (tones and noise), the model was shown to be consistent with time-intensity trading and with perceptions of image width for decorrelated noises. For nonstationary signals, such as bandpass-filtered transients (Lindemann 1986b), the inhibitory factors lead to a reinforcement of the current locations of activity and a consequent emphasis on the lead sound when there is a lead-lag click pair. In the same paper, this model was also shown to be consistent with several sets of measurements of echo threshold, including the general time scales and the influence of the leading stimulus on the location of the echo. Specifically, the stimulus overlap resulting from the narrow-band filtering of the input leads to the echo location being pulled toward the lead position, even after the echo is perceived as a distinct object. Although this is a powerful model that makes important contributions to our theoretical understanding, the model has not been applied extensively to precedence effect situations, presumably because of its complexity and lack of available implementations. This model is consistent with suggestions of McFadden (1973) in that initial sounds from one direction suppress later sounds from different directions more than sounds from the initial direction. Extensions of this model have been developed by Gaik (1993) to improve the model's response to naturally occurring interaural level differences and by Bodden (1993) to address aspects of the cocktail-party effect.

The third model, proposed by Shinn-Cunningham *et al.* (1993), is a descriptive model that allows the quantitative prediction of lead-lag discrimination data from localization dominance data. The model has been applied to experiments in which the stimulus contains a lead stimulus with ITD  $\tau_1$  and a lag stimulus with an ITD  $\tau_2$ , separated by the lead-lag delay. According to this model, an internal variable is generated which corresponds to the ITD  $\alpha$  of the pointer in a localization matching experiment when the lead ITD is  $\tau_1$  and the lag ITD is  $\tau_2$ . The model postulates that

$$\alpha = c * \tau_1 + (1 - c) * \tau_2,$$

where  $c$  is a non-negative weight that represents the emphasis on the lead relative to the lag stimulus. For example, if the value of  $c=0.5$ , then lead and lag are equally weighted, there is no precedence effect, and the matching ITD is equal to the average of the lead and lag ITDs. If the value of  $c$  is unity ( $c=1.0$ ), then the lead completely dominates the generation of the location and the matching ITD is equal to the lead ITD. (If the value of  $c$  were zero, this equation would describe lateral position that is determined by the lag ITD, an extreme form of "antiprecedence.") The linear relation assumed here appears to be adequate for almost all position matching data. The dependence of the parameter  $c$  on stimulus attributes, such as lead-lag delay and stimulus frequency, allows a simple, quantitative description of localization-

dominance results. In addition, by assuming that discrimination of ITD is made on the basis of the combined variable  $\alpha$ , and that the separate ITDs are not available, one can predict results of lead-lag discrimination experiments. When these predictions were evaluated (Shinn-Cunningham *et al.*, 1993), the model was shown to explain the quantitative relationship between the results of these two different types of experiments. Specifically, Shinn-Cunningham *et al.* (1993) measured these two tasks in the same listeners. Their model parameters, which were estimated from data obtained with the "pointer adjustment" method (see Sec. IB 1), were able to predict results of the ITD discrimination task quite accurately. The disadvantage of this model is that no mechanism is described, so that the model does not *a priori* predict the size of  $c$  for different experiments, nor is there a natural extension to other types of experiments or stimuli. Also, as described so far, the model is limited to ITD experiments and to two stimulus components with well-defined lead and lag delays with a single fused image, but the model could be naturally extended to multiple bursts, to angles instead of delays, and to multiple images. This model has been extended to free-field experiments by Litovsky and Macmillan (1994). More recently, Stellmack *et al.* (1998) confirmed the usefulness of this model for interpreting discrimination experiments and obtained similar values for the weighting parameter  $c$ .

The fourth model (Tollin, 1998) provides a mechanism for the combination of interaural phase and level information with the time relations of the stimulus components explicitly included in the model. This model has been applied to pairs of click stimuli with a focus on the discrimination of ITD in lead or lag stimulus. In general, estimates of ITD and ILD are combined to generate a decision variable corresponding to lateral position. A notable feature of this model is that, for the cases to which it has been applied, namely the discrimination of wideband clicks, there is no inhibition contained within the model, and yet the model is able to predict that discrimination of the lead click is significantly better than discrimination of the lag click. The basic elements of this model include: a bandpass, rectified cross-correlation function calculation to generate an ITD-based estimate of location; a short-term energy ratio calculation for an ILD-based estimate of location; and a linear weighting of these estimates to generate a final lateral position. A central feature of the model is that the bandpass filter that is used for the estimate of position of a wideband transient stimulus is the filter that is close to 750 Hz (the dominance region) and that is near a peak of the stimulus magnitude spectrum for one ear. This model is also able to predict several "anomalous localization" results seen in reports from Wallach *et al.* (1949), Gaskell (1983), and Tollin and Henning (1998). In these cases, the localized direction is opposite to the direction expected from the overall ITD and ILD of the stimulus. The focus on the 750-Hz region and the combination of time and intensity differences for the generation of the lateral position variable and the discrimination decision variable are consistent with the ideas in Gaskell (1983). The model has yet to be applied to a wide class of stimuli and to a wide range of parameters.



To the extent that precedence effects are consistent with the physiological observations discussed in the previous section, models of the physiology should lead to models of precedence. Thus for example, the mechanism suggested by Wickesberg and Oertel (1990) may be involved in monaural processing that influences precedence. Similarly, the models of Cai *et al.* (1998a, 1998b), which show that simple mechanisms of inhibition can successfully describe the “best-ITD-inhibited” and “worst-ITD-inhibited” neurons described above, may capture aspects of processing that are critical for precedence effects. These physiological models have not been applied to predict the psychophysical performance explicitly and they are not considered further in this review.

It may be noted that the models described in this section are focused on relatively peripheral levels and, as such, are limited to relatively simple, noncognitive mechanisms. However, some aspects of the precedence effects are much more naturally described in cognitive terms, including the buildup and breakdown effects seen with sequences of stimuli. These aspects of precedence illustrate that the peripheral mechanisms included in these models may tell only part of the story. It is appropriate to pursue both types of models, and to keep in mind that a general understanding of precedence effects may require more elaborate interpretations.

## VI. CONCLUSION

We have attempted to provide an updated overview of recent psychoacoustical and physiological work on the effects of precedence. By introducing the framework of fusion, localization dominance, and lag-discrimination suppression along with the buildup and breakdown of suppression, we wanted to provide a framework within which to describe the effects of precedence and to help in the integration of data from both psychophysical and physiological experiments. While the motivation for many of the experiments is partly rooted in our desire to understand how the auditory system processes multiple arrays of directional cues in reverberant spaces, many of the studies do not directly address “real-world” issues. In fact, the precedence effect as an auditory phenomenon has little to do with realistic acoustic environments. However, we believe that the work in this area can be instrumental in motivating future work in more realistic scenarios, and that it offers a unique window into the auditory system, by viewing auditory processes through human perception, animal behavior and physiology, as well as development. It is probably only through the combined efforts of these various approaches that a full theory of precedence will evolve and useful models will be developed. Since this topic of research is dynamic and ongoing, our paper represents a mere snapshot that will clearly need updating as the field progresses.

## ACKNOWLEDGMENTS

The work of Ruth Litovsky is supported by NIDCD Grant (Nos. DC02696 and DC00100) and by ONR-managed MURI Grant (No. Z883401). The work of H. Steven Colburn is supported by NIDCD Grant (No. DC00100). The work of William A. Yost and Sandra Guzman is supported by an

NIDCD Program Project Grant (No. DC00293) and by the Air Force Office of Scientific Research. We are grateful to Drs. Rachel Clifton, Pierre Divenyi, Daniel Tollin, and an anonymous reviewer, for their insightful comments on an earlier version of the manuscript. A special thanks are due to Bob Dizon, for many insightful comments and stimulating discussions during the past few years. We would also like to thank Drs. John Brugge and Duck Kim for contributing physiological data collected in their laboratories.

- Adams, J., and Mugnaini, E. (1984). “Dorsal nucleus of the lateral lemniscus: A nucleus of GABAergic projection neurons,” *Brain Res. Bull.* **13**, 585–590.
- Altman, J. A. (1975). “Neurophysiological mechanisms in auditory localization,” in *Brain information services, Soviet research reports* (BRI Publication Office, Univ. of California, Los Angeles), Vol. 1, pp. 27–32.
- Aoki, S., and Houtgast, T. (1992). “A precedence effect in the perception of interaural cross correlation,” *Hearing Res.* **59**, 25–30.
- Ashmead, D. H., Clifton, R. K., and Reese, E. P. (1986). “Development of auditory localization in dogs: Single source and precedence effect sounds,” *Dev. Psychobiol.* **19**, 91–103.
- von Békésy, G. (1930). “Zur Theorie des Hörens: Über das Richtungshören bei einer Zeitdifferenz oder Lautstärkeungleichheit der beidseitigen Schalleinwirkungen [On the theory of hearing: On directional hearing in connection with a time difference or inequality of loudness of the effective sound between the two sides],” *Phys. Z.* **31**, 824–838; **31**, 857–868.
- Blauert, J. (1969). “Sound localization in the median plane,” *Acustica* **22**, 205–213.
- Blauert, J. (1971). “Localization and the law of the first wavefront in the median plane,” *J. Acoust. Soc. Am.* **50**, 466–470.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition* (The MIT Press, Cambridge, MA).
- Blauert, J., and Col, J. P. (1992). “Irregularities in the precedence effect,” in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demaney, and K. Horner (Pergamon, Oxford), pp. 531–538.
- Blauert, J., and Divenyi, P. L. (1988). “Spectral selectivity in binaural contralateral inhibition,” *Acustica* **66**, 267–274.
- Bodden, M. (1993). “Modeling human sound source localization and the cocktail-party-effect,” *Acta Acust.* **1**, 43–55.
- de Boer, K. (1940). “Three-dimensional sound reproduction,” *Philips Tech. Rev.* **5**, 107–115.
- Boettcher, F. A., Salvi, R. J., and Saunders, S. S. (1990). “Recovery from short-term adaptation in single neurons in the cochlear nucleus,” *Hearing Res.* **48**, 125–144.
- Butler, R. A. (1969). “Monaural and binaural localization of noise bursts vertically in the median sagittal plane,” *J. Auditory Res.* **3**, 230–235.
- Cai, H., Carney, L. H., and Colburn, H. S. (1998a). “A model for binaural response properties of inferior colliculus neurons: I. A model with ITD-sensitive excitatory and inhibitory inputs,” *J. Acoust. Soc. Am.* **103**, 475–493.
- Cai, H., Carney, L. H., and Colburn, H. S. (1998b). “A model for binaural response properties of inferior colliculus neurons: II. A model with ITD-sensitive excitatory and inhibitory inputs and an adaptation mechanism,” *J. Acoust. Soc. Am.* **103**, 494–506.
- Carney, L. H., and Yin, T. C. T. (1989). “Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: Excitatory and inhibitory components,” *J. Neurophysiol.* **62**, 144–161.
- Clifton, R. K. (1985). “The precedence effect: Its implications for developmental questions,” in *Auditory Development in Infancy*, edited by S. E. Trehub and B. Schneider (Plenum, New York), pp. 85–99.
- Clifton, R. K. (1987). “Breakdown of echo suppression in the precedence effect,” *J. Acoust. Soc. Am.* **82**, 1834–1835.
- Clifton, R. K., and Freyman, R. L. (1989). “Effect of click rate and delay on breakdown of the precedence effect,” *Percept. Psychophys.* **46**(2), 139–145.
- Clifton, R. K., and Freyman, R. L. (1997). “The precedence effect: Beyond echo suppression,” in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, Mahwah, NJ), pp. 233–255.
- Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (1994). “Listeners’ expectations about echoes can raise or lower echo threshold,” *J. Acoust. Soc. Am.* **95**(3), 1525–1533.

- Clifton, R. K., Morriongiello, B. A., Kulig, J. W., and Dowd, J. M. (1981). "Newborns' orientation toward sound: Possible implications for cortical development," *Child Dev.* **52**, 833–838.
- Colburn, H. S. (1996). "Computational models of binaural processing," in *Springer Handbook of Auditory Research: Auditory Computation*, edited by R. R. Fay and A. N. Popper (Springer, New York), pp. 332–400.
- Cranford, J. L. (1982). "Localization of paired sound sources in cats: Effects of variable arrival times," *J. Acoust. Soc. Am.* **72**(4), 1309–1311.
- Cranford, J. L., and Oberholtzer, M. (1976). "Role of neocortex in binaural hearing in the cat, II: The precedence effect in sound localization," *Brain Res.* **111**, 225–239.
- Cranford, J. L., Ravizza, R., Diamond, I. T., and Whitfield, I. C. (1971). "Unilateral ablation of the auditory cortex in the cat impairs complex sound localization," *Science* **172**, 286–288.
- David, E. E. (1959). "Comment on the precedence effect," in *Proceedings of the 3rd International Congress On Acoustics*, Stuttgart, Vol. 1, pp. 144–146.
- Divenyi, P. L. (1992). "Binaural suppression of nonechoes," *J. Acoust. Soc. Am.* **91**(2), 1078–1084.
- Dizon, R., Litovsky, R. Y., and Colburn, H. S. (1997). "Positional dependence on localization dominance in the median-sagittal plane," *J. Acoust. Soc. Am.* **101**, 3106(A).
- Ebata, M., Sone, T., and Nimura, T. (1968). "On the perception of direction of echo," *J. Acoust. Soc. Am.* **44**, 542–547.
- Fisher, H. and Freedman, S. J. (1968). "The role of the pinna in auditory localization," *J. Auditor. Res.* **8**, 15–26.
- Fitzpatrick, D. C., Kuwada, S., Batra, R., and Trahiotis, C. (1995). "Neural responses to simple, simulated echoes in the auditory brainstem of the unanesthetized rabbit," *J. Neurophysiol.* **74**, 2469–2486.
- Fitzpatrick, D. C., Kuwada, S., and Kim, D. O. (1998). "Responses of neurons to click-pair stimuli: Auditory nerve to auditory cortex," *Assoc. Res. Otolaryngology* (Abstract; p. 53).
- Franssen, N. V. (1960). "Some considerations on the mechanism of directional hearing," Ph.D. thesis, Technische Hogeschool, Delft, The Netherlands.
- Franssen, N. V. (1963). *Stereophony* (Phillips Technical Bibliography, Eindhoven).
- Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (1991). "Dynamic processes in the precedence effect," *J. Acoust. Soc. Am.* **90**, 874–884.
- Freyman, R. L., McCall, D. M., and Clifton, R. K. (1998). "Intensity discrimination for precedence effect stimuli," *J. Acoust. Soc. Am.* **103**, 2031–2041.
- Gaik, W. (1993). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Am.* **94**, 98–110.
- Gardner, M. B., and Gardner, R. S. (1973). "Problem of localization in the median plane: Effect of pinnae cavity occlusion," *J. Acoust. Soc. Am.* **53**, 400–408.
- Gaskell, H. (1983). "The precedence effect," *Hearing Res.* **12**, 277–303.
- Gilkey, R. H., and Anderson, T. R. (Eds.) (1997). *Binaural and Spatial Hearing in Real and Virtual Environments* (Lawrence Erlbaum, Mahwah, NJ).
- Grantham, D. W. (1996). "Left-right asymmetry in the buildup of echo suppression in normal-hearing adults," *J. Acoust. Soc. Am.* **99**(2), 1118–1123.
- Guzman, S. J., and Yost, W. A. (1999). "Forward masking in the precedence effect," *J. Acoust. Soc. Am.* (submitted).
- Haas, H. (1949). "The influence of a single echo on the audibility of speech," *J. Audiol. Eng. Soc.* **20**, 145–159, English translation (1972).
- Haas, H. (1951). "On the influence of a single echo on the intelligibility of speech," *Acustica* **1**, 48–58.
- Hafter, E. R. (1997). "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 Buell, T. N. (1990). "Restarting the adapted binaural system," *J. Acoust. Soc. Am.* **88**(2), 806–812.
- Hafter, E. R., Buell, T. N., and Richards, V. (1988). "Onset-coding in lateralization: Its form, site, and function," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 647–676.
- Hafter, E. R., Dye, R. H., and Wenzel, E. (1983). "Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number," *J. Acoust. Soc. Am.* **73**, 1708–1713.
- Hafter, E. R., Saberi, K., Jensen, E. R., and Briolle, F. (1992). "Localisation in an echoic environment," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demaney, and K. Horner (Pergamon, Oxford), pp. 555–561.
- Harris, G. G., Flanagan, J. L., and Watson, B. J. (1963). "Binaural interaction of a click with a click pair," *J. Acoust. Soc. Am.* **35**, 672–678.
- Hartmann, W. M., and Rakerd, B. (1989). "Localization of sound in rooms IV: The Franssen effect," *J. Acoust. Soc. Am.* **86**(4), 1366–1373.
- Hochster, M. E., and Kelly, J. B. (1981). "The precedence effect and sound localization by children with temporal lobe epilepsy," *Neuropsychologia* **19**, 49–55.
- Jeffress, L. A. (1948). "A place theory of sound localization," *J. Comp. Physiol. Psychol.* **41**(1), 35–39.
- Kaltenbach, J. A., Melega, R. J., Falzarano, P. R., Myers, S. F., and Simpson, T. H. (1993). "Forward masking properties of neurons in the dorsal cochlear nucleus: Possible role in the process of echo suppression," *Hearing Res.* **67**, 35–44.
- Keller, C. H., and Takahashi, T. T. (1996). "Responses to simulated echoes by neurons in the barn owl's auditory space map," *J. Comp. Physiol.* **178**(4), 499–512.
- Kelly, J. B. (1974). "Localization of paired sound sources in the rat: Small time differences," *J. Acoust. Soc. Am.* **55**, 1277–1284.
- Kuwada, S., Batra, R., and Stanford, T. R. (1989). "Monaural and binaural response properties of neurons in the inferior colliculus of the rabbit: Effects of sodium pentobarbital," *J. Neurophysiol.* **61**, 269–282.
- Leakey, D. M. (1959). "Some measurements of the effects of interchannel intensity and time differences in two channel sound systems," *J. Acoust. Soc. Am.* **31**, 977–986.
- Leakey, D. M., and Cherry, E. C. (1957). "Influence of noise upon the equivalence of intensity differences and small time delays in two-loudspeaker systems," *J. Acoust. Soc. Am.* **29**, 284–286.
- Lieberman, M. C. (1978). "Auditory nerve responses from cats raised in a low-noise chamber," *J. Acoust. Soc. Am.* **63**, 442–455.
- Lindemann, W. (1986a). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.* **80**, 1608–1622.
- Lindemann, W. (1986b). "Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front," *J. Acoust. Soc. Am.* **80**, 1623–1630.
- Litovsky, R. Y. (1997). "Developmental changes in the precedence effect: Estimates of Minimal Audible Angle," *J. Acoust. Soc. Am.* **102**, 1739–1745.
- Litovsky, R. Y., Cransten, B. R., and Delgutte, B. (1998). "Neural correlates of the precedence effect in the inferior colliculus: Effect of localization cues," *Assoc. Res. Otolaryngology* (Abstract, p. 40).
- Litovsky, R. Y., Hawley, M. L., and Colburn, H. S. (1997a). "Measurement of precedence in monaural listeners," Meeting of the American Speech and Hearing Association, Boston, MA.
- Litovsky, R. Y., Rakerd, B., Yin, T. C. T., and Hartmann, W. M. (1997b). "Psychophysical and physiological evidence for a precedence effect in the median sagittal plane," *J. Neurophysiol.* **77**, 2223–2226.
- Litovsky, R. Y. (1998). "Physiological studies on the precedence effect in the inferior colliculus of the kitten," *J. Acoust. Soc. Am.* **103**, 3139–3152.
- Litovsky, R. Y. and Colburn, H. S. (1998). "Precedence effects in the azimuthal and sagittal planes," *Assoc. Res. Otolaryngology Abstracts* **21**, p. 53.
- Litovsky, R. Y., and Yin, T. C. T. (1998a). "Physiological studies of the precedence effect in the inferior colliculus of the cat: I. Correlates of psychophysics," *J. Neurophysiol.* **80**, 1285–1301.
- Litovsky, R. Y., and Yin, T. C. T. (1999b). "Physiological studies of the precedence effect in the inferior colliculus of the cat: II. Neural mechanisms," *J. Neurophysiol.* **80**, 1302–1316.
- Litovsky, R. Y., Dizon, R. M., and Colburn, H. S. (1999). "Studies of the precedence effect in the median-sagittal and azimuthal planes in a virtual acoustic space," *J. Acoust. Soc. Am.* (submitted).
- Litovsky, R. Y., and Ashmead, D. H. (1997). "Developmental aspects of binaural and spatial hearing," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, Mahwah, NJ), pp. 571–592.
- Litovsky, R. Y., and Macmillan, N. A. (1994). "Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli," *J. Acoust. Soc. Am.* **96**(2) Pt. 1, 752–758.
- Lochner, J. P. A., and Burger, J. F. (1958). "The subjective masking of

- short time delayed echoes, their primary sounds, and their contribution to the intelligibility of speech," *Acustica* **8**, 1–10.
- McFadden, D. (1973). "Precedence effect and auditory cells with long characteristic delays," *J. Acoust. Soc. Am.* **54**, 528–530.
- Martin, K. D. (1997). "Echo suppression in a computational model of the precedence effect," presented at the 1997 IEEE Mohonk Workshop on Applications of Signal Processing to Acoustics and Audio, New Paltz, NY.
- Meyer, E., and Schroeder, G. R. (1952). "Über den Einfluss von Schallrückwürfen auf Richtungslokalisation und Lautstärke bei Sprache [On the influence of reflected sound on directional localization and loudness of speech]," *Nachr. Akad. Wiss. Goett. II, Math.-Phys. Kl.* **6**, 31–42.
- Middlebrooks, J. C., and Green, D. M. (1991). "Sound localization by human listeners," *Annu. Rev. Psychol.* **42**, 135–159.
- Moore, B. C. J. (1996). *An Introduction to the Psychology of Hearing* (Academic, New York).
- Moore, J. K., and Moore, R. Y. (1987). "Glutamic-acid decarboxylase-like immunoreactivity in brain-stem auditory nuclei of the rat," *J. Comp. Neurol.* **260**(2), 157–174.
- Morronegello, B. A., Kulig, J. W., and Clifton, R. K. (1984). "Developmental changes in auditory temporal perception," *Child Dev.* **55**, 461–471.
- Muncey, R. W., Nickson, A. F. B., and Dubout, P. (1953). "The acceptability of speech and music with a single artificial echo," *Acustica* **3**, 168–173.
- Oliver, D. L., Beckius, G. E., and Ostapoff, E. M. (1994). "Connectivity of neurons in identified auditory circuits studied with transport of dextran and microspheres plus intracellular injection of Lucifer Yellow," *J. Neurosci. Methods* **53**, 23–27.
- Parham, K., Zhao, H. B., and Kim, D. O. (1996). "Responses of auditory nerve fibers to unanesthetized decerebrate cat to click pairs as simulated echoes," *J. Neurophysiol.* **76**, 17–28.
- Parham, K., Zhao, H. B., Ye, Y., and Kim, D. O. (1998). "Responses of anteroventral cochlear nucleus neurons of the unanesthetized decerebrate cat to click pairs as simulated echoes," *Hearing Res.* **125**, 131–146.
- Perrott, D. R., and Pacheco, S. (1989). "Minimal auditory angle thresholds for broadband noise as a function of delay between the onset of the lead and lag signals," *J. Acoust. Soc. Am.* **85**, 2669–2672.
- Perrott, D. R., Marlborough, K., Merrill, P., and Strybel, T. Z. (1989). "Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate," *J. Acoust. Soc. Am.* **85**, 282–288.
- Perrott, D. R., Strybel, T. Z., and Manligas, C. L. (1987). "Conditions under which the Haas precedence effect may or may not occur," *J. Audit. Res.* **27**, 59–72.
- Populin, I. C., and Yin, T. C. T. (1998). "Behavioral studies of sound localization in the cat," *J. Neurosci.* **18**, 2147–2160.
- Reale, R. A., Brugge, J. F., and Hind, J. E. (1995). "Encoding of stimulus direction by AI neurons in the cat: Effects of sounds arriving at different times and from different directions," *Neurosci. Abs.* **21**, 667.
- Rakerd, B., and Hartmann, W. M. (1985). "Localization of sound in rooms, II: The effects of a single reflecting surface," *J. Acoust. Soc. Am.* **78**(2), 524–533.
- Rakerd, B., Hsu, J., and Hartmann, W. M. (1997). "The Haas effect with and without binaural differences," *J. Acoust. Soc. Am.* **101**(5), 3083.
- Saberi, K., and Perrott, D. R. (1990). "Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate," *J. Acoust. Soc. Am.* **87**(4), 1732–1737.
- Saberi, K. (1996). "Observer weighting of interaural delays in filtered impulses," *Percept. Psychophys.* **58**(7), 1037–1046.
- Saintmarie, R. L., Ostapoff, E. M., Morest, D. K., and Wenthold, R. J. (1989). "Glycine-immunoreactive projection of the cat lateral superior olive: Possible role in midbrain ear dominance," *J. Comp. Neurol.* **279**, 382–396.
- Scharf, B. (1974). "Localization of unlike tones from two loudspeakers," in *Sensation and measurement: papers in honor of S.S. Stevens*, edited by H. R. Moskowitz, B. Scharf, and J. C. Stevens (Reidel, Dordrecht), pp. 309–314.
- Schubert, E. D., and Wernick, J. (1969). "Envelope versus microstructure in the fusion of dichotic signals," *J. Acoust. Soc. Am.* **45**, 1525–1531.
- Shaw, A. G. (1974). "The external ear," in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer-Verlag, New York), Vol. V/1, pp. 455–490.
- Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (1995). "Cross-frequency interactions in the precedence effect," *J. Acoust. Soc. Am.* **98**(1), 164–171.
- Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (1993). "Adjustment and discrimination measurements of the precedence effect," *J. Acoust. Soc. Am.* **93**(5), 2923–2932.
- Shore, S. E. (1995). "Recovery of forward-masked responses in ventral cochlear nucleus neurons," *Hearing Res.* **82**, 31–43.
- Snow, W. (1954a). "Basic principles of stereophonic sound," *J. Soc. Motion Picture Television Eng.* **61**, 567–589.
- Snow, W. B. (1954b). "Effect of arrival time on stereophonic localization," *J. Acoust. Soc. Am.* **26**, 1071–1074.
- Stellmack, M. A., Dye, R. H., and Guzman, S. J. (1999). "Observer weighting of binaural information in source and echo clicks," *J. Acoust. Soc. Am.* **105**, 377–387.
- Stern, R. M., and Trahiotis, C. (1995). "Models of binaural interaction" in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, New York), Vol. 6, pp. 347–386.
- Thurlow, W. R., and Parks, T. E. (1961). "Precedence-suppression effects for two click sources," *Percept. Mot. Skills* **13**, 7–12.
- Tollin, D. J. (1998). "Computational model of the lateralisation of clicks and their echoes," in *Proceedings of the NATO Advanced Study Institute on Computational Hearing*, I-Lucca, Italy, 1–12 July 1998, edited by S. Greenberg and M. Slaney, pp. 77–82.
- Tollin, D. J., and Henning, G. B. (1998). "Some aspects of the lateralization of echoed sound in man. I: Classical interaural delay-based precedence," *J. Acoust. Soc. Am.* **104**, 3030–3038.
- Tollin, D. J., and Henning, G. B. (1999). "Some aspects of the lateralization of echoed sound in man. II: The role of stimulus spectrum," *J. Acoust. Soc. Am.* **105**, 838–849.
- Wallach, H., Newman, E. B., and Rosenzweig, M. R. (1949). "The precedence effect in sound localization," *Am. J. Psychol.* **LXII**(3), 315–336.
- Warncke, H. (1941). "The fundamentals of room-related stereophonic reproduction in sound films," *Akust. Zh.* **6**, 174–188.
- Whitfield, I. C. (1978). "Auditory cortical lesions and the precedence effect in a four-choice situation," *J. Physiol. (London)* **289**, 81.
- Whitfield, I. C., Cranford, J., Ravizza, R., and Diamond, I. T. (1972). "Effects of unilateral ablation of auditory cortex in cat on complex sound localization," *J. Neurophysiol.* **35**, 718–731.
- Wickesberg, R. E. (1996). "Rapid inhibition in the cochlear nuclear complex of the chinchilla," *J. Acoust. Soc. Am.* **100**(3), 1691–1702.
- Wickesberg, R. E., and Oertel, D. (1990). "Delayed, frequency-specific inhibition in the cochlear nuclei of mice: A mechanism for monaural echo suppression," *J. Neurosci.* **10**(6), 1762–1768.
- Wyttenbach, R. A., and Hoy, R. R. (1993). "Demonstration of the precedence effect in an insect," *J. Acoust. Soc. Am.* **94**, 777–784.
- Yang, X., and Grantham, D. W. (1997a). "Echo suppression and discrimination suppression aspects of the precedence effect," *Percept. Psychophys.* **59**, 1108–1117.
- Yang, X., and Grantham, D. W. (1997b). "Cross-spectral and temporal factors in the precedence effect: Discrimination suppression of the lag sound in free-field," *J. Acoust. Soc. Am.* **102**, 2973–2983.
- Yin, T. C. T. (1994). "Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat," *J. Neurosci.* **14**, 5170–5186.
- Yost, W. A., Patterson, R., and Sheft, S. (1996). "A time domain description of iterated rippled noise," *J. Acoust. Soc. Am.* **99**(2), 1066–1078.
- Yost, W. A., and Guzman, S. J. (1996). "Auditory processing of sound sources: Is there an echo in here?" *Curr. Directions Psychol. Sci.* **5**(4), 125–131.
- Yost, W. A., and Hafter, E. R. (1987). "Lateralization," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 49–84.
- Yost, W. A., and Soderquist, D. R. (1984). "The precedence effect: Revisited," *J. Acoust. Soc. Am.* **76**(5), 1377–1383.
- Yost, W. A., Mapes-Riordan, D., and Guzman, S. J. (1997). "The relationship between localization and the Franssen effect," *J. Acoust. Soc. Am.* **101**(5), 2994–2997.
- Zurek, P. M. (1980). "The precedence effect and its possible role in the avoidance of interaural ambiguities," *J. Acoust. Soc. Am.* **67**(3), 952–964.
- Zurek, P. M. (1987). "The precedence effect," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 85–105.



# Errors due to sensor and position mismatch in planar acoustic holography

Kyoung-Uk Nam and Yang-Hann Kim<sup>a)</sup>

Center for Noise and Vibration Control (NOVIC), Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Science town, Taejon-shi, 305-701, Korea

(Received 19 October 1998; revised 16 June 1999; accepted 17 June 1999)

There are many sources of measurement errors when one attempts to construct a hologram of near-field acoustic holography. These errors are amplified when predicting pressure at a point which is located near an acoustic source. Errors due to a sensor mismatch and an inaccurate measurement location, or position mismatch, are the main interests of this study. We theoretically measure the errors in terms of their probabilities, bias, and random error. Analysis in a wave number domain shows that the bias error can be regarded as negligible, but the random error is significant in a backward prediction. The random error is suggested to be measured in terms of random error energy. Then, the amplification ratio of the random error energy is related to measurement spacing on a hologram plane, prediction spacing on a prediction plane, and the distance between the hologram and prediction plane. © 1999 Acoustical Society of America. [S0001-4966(99)01610-0]

PACS numbers: 43.20.-f, 43.35.Sx, 43.60.Cg, 43.60.Sx [AN]

## INTRODUCTION

A hologram is a plane that has all the necessary information to predict any unmeasured acoustic value. The predicted values are acoustic pressure, velocity, intensity, potential energy, and kinetic energy. The quality of the prediction, therefore, depends on the quality of the hologram. If one has good hologram data, then the predicted sound fields will be very reliable. However, this conclusion does not have any practical value, unless one defines the measure of the hologram quality.

The quality of a hologram can be determined by the quantity and quality of the hologram data. The number of data per unit area and the area of the hologram define the quantity of the hologram data. This simply means that more information produces better result. This can be achieved only by using as many microphones as possible; however, this is not within the context of our paper. The quality of the hologram data is directly related to the accuracy of measurements, and this is what we can study.

This paper deals with the errors of a hologram, and their effects on the predicted sound field. Our attempt is to theoretically formulate a relation between the errors of a hologram and the predicted sound field variable, i.e., acoustic pressure on a selected plane.

There will be many sources of errors on a measured hologram. One of them is a sensor mismatch among microphones. This introduces the errors in magnitude, as well as phase, thus degrading the quality of the hologram data. Similar errors could be significant due to inaccurate information on the location of the sensors (position mismatch).

## I. THEORETICAL FORMULATION

Figure 1 illustrates a discrete planar acoustic holograph, where  $p_H(m,n)$  represents the pressure at a selected fre-

quency on the hologram plane. Pressure on a prediction plane at arbitrary  $z$ ,  $p_Z(m,n)$ , is estimated by the pressure on the hologram plane,  $p_H(m,n)$ . The estimated pressure contains errors due to aliasing, finite aperture size, and measurement errors. The problems associated with the aliasing and finite aperture size are well addressed in Refs. 1 and 2. This paper studies the effect of the measurement errors on the estimated pressure. The measurement errors are introduced by the sensor or position mismatch.

### A. Expression of measurement value

The error due to the sensor mismatch will be significant, especially when many sensors are used. The transfer functions of the sensors are hardly identical. There will be magnitude and phase differences among the sensors. Thus one can write the measured pressure ( $\hat{p}_H$ ), which has the magnitude and phase error with respect to true pressure ( $p_H$ ), as

$$\hat{p}_H(m,n) = (1 + \varepsilon_a(m,n))e^{j\varepsilon_\phi(m,n)}p_H(m,n), \quad (1)$$

where  $m,n = -N/2+1, \dots, 0, \dots, N/2$ , and  $N$  represents the number of data along one axis.  $\varepsilon_a$  and  $\varepsilon_\phi$  with real values represent the magnitude and phase mismatch of the transfer function of the sensor. These will vanish if the sensors are perfectly identical—no sensor mismatch. The magnitude of  $\hat{p}_H$  is different than the magnitude of  $p_H$  multiplied by  $\varepsilon_a$ , and the phase of  $\hat{p}_H$  is different than the phase of  $p_H$  by  $\varepsilon_\phi$ . One can assume that  $\varepsilon_a$  and  $\varepsilon_\phi$  are reasonably small. This assumption enables Eq. (1) to be expressed in terms of Taylor series. That is,

$$\begin{aligned} \hat{p}_H(m,n) \approx & (1 + \varepsilon_a(m,n) + j\varepsilon_\phi(m,n) \\ & + \frac{1}{2}j^2\varepsilon_a(m,n)\varepsilon_\phi(m,n) \\ & - \frac{1}{2}\varepsilon_\phi^2(m,n))p_H(m,n). \end{aligned} \quad (2)$$

The position mismatch is another source of the measurement errors. It happens when an actual measurement location is different than a desired one. If  $\varepsilon_x$  and  $\varepsilon_y$  are the position

<sup>a)</sup>Electronic mail: yhkim@sorak.kaist.ac.kr

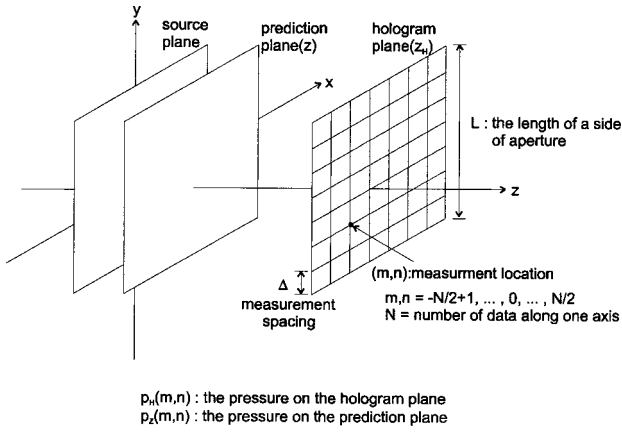


FIG. 1. Illustration of discrete planar acoustic holography.

errors with respect to  $x$  and  $y$  coordinates, respectively, then the pressure at  $(x + \varepsilon_x, y + \varepsilon_y)$  on the hologram plane can be written as, to the second order approximation,

$$\begin{aligned}
 p_H(x + \varepsilon_x, y + \varepsilon_y) \approx & p_H(x, y) + \frac{\partial p_H}{\partial x} \varepsilon_x + \frac{\partial p_H}{\partial y} \varepsilon_y \\
 & + \frac{1}{2} \frac{\partial^2 p_H}{\partial x^2} \varepsilon_x^2 + \frac{1}{2} \frac{\partial^2 p_H}{\partial x \partial y} \varepsilon_x \varepsilon_y \\
 & + \frac{1}{2} \frac{\partial^2 p_H}{\partial y^2} \varepsilon_y^2.
 \end{aligned} \quad (3)$$

Thus the measured pressure at the position  $(m, n)$  with the errors can be written as

$$\begin{aligned}
 \hat{p}_H(m, n) \approx & p_H(m, n) + v_x(m, n) \varepsilon_x(m, n) \\
 & + v_y(m, n) \varepsilon_y(m, n) + c_{xx}(m, n) \varepsilon_x^2(m, n) \\
 & + c_{xy}(m, n) \varepsilon_x(m, n) \varepsilon_y(m, n) \\
 & + c_{yy}(m, n) \varepsilon_y^2(m, n),
 \end{aligned} \quad (4)$$

where

$$\begin{aligned}
 v_x = \frac{\partial p_H}{\partial x}, \quad v_y = \frac{\partial p_H}{\partial y}, \quad c_{xx} = \frac{1}{2} \frac{\partial^2 p_H}{\partial x^2}, \\
 c_{xy} = \frac{1}{2} \frac{\partial^2 p_H}{\partial x \partial y}, \quad c_{yy} = \frac{1}{2} \frac{\partial^2 p_H}{\partial y^2}.
 \end{aligned}$$

$v_{x,y}$  represent the slopes of the pressure, and  $c_{xx,xy,yy}$  represent the curvatures of the pressure.  $\hat{p}_H$ ,  $p_H$ ,  $v_{x,y}$ , and  $c_{xx,xy,yy}$  are complex, but  $\varepsilon_x$  and  $\varepsilon_y$  are real.

We expressed the measured pressure on the hologram plane in terms of error sources:  $\varepsilon_{a,\phi,x,y}$ . The next section discusses the error sources.

## B. Assumption of error sources

First, let us consider the magnitude mismatch  $\varepsilon_a$ . The value of  $\varepsilon_a$  can have any value less than unity, depending on the sensor. This will be true for all sensors. Therefore, we can assume that  $\varepsilon_a$  is distributed irregularly on the hologram plane. In other words,  $\varepsilon_a$  is regarded as drawn arbitrarily from a population  $S_a$ . By the same token, one can assume that other error sources  $\varepsilon_{\phi,x,y}$  are drawn arbitrarily from

populations  $S_{\phi,x,y}$ . This simply implies that the hologram measurement can be regarded as a sample of all possible measurement events or ensemble.

It is noteworthy that if one denotes  $\theta$ , a quantity estimated by the true pressure on the hologram plane, then  $\hat{\theta}$  estimated by the measured pressure in a measurement event would be different than  $\theta$ . The error of the estimated value  $\hat{\theta}$  cannot be estimated, unless all the values of the error sources are known at all the measurement points. However, if the statistical properties of the populations are known, then it is possible to analyze the probability of the error in a measurement event. Bias and random error can express the probability of the error.<sup>3</sup>

Let us consider that the mean and variance of the population  $S_{a,\phi,x,y}$  are 0 and  $\sigma_{a,\phi,x,y}^2$ . Because the error sources are drawn arbitrarily from the populations, the ensemble average of  $\varepsilon_{a,\phi,x,y}$  and  $\varepsilon_{a,\phi,x,y}^2$  at a point are equal to 0 and  $\sigma_{a,\phi,x,y}^2$ . By the same reason,  $\varepsilon_{a,\phi,x,y}$  at a point are uncorrelated with those at other points. These relations can be written as

$$\begin{aligned}
 E[\varepsilon_{a,\phi,x,y}(m, n)] &= 0, \\
 E[\varepsilon_{a,\phi,x,y}(m, n) \varepsilon_{a,\phi,x,y}(m', n')] & \\
 &= \sigma_{a,\phi,x,y}^2 \delta(m' - m) \delta(n' - n),
 \end{aligned} \quad (5)$$

where  $E$  represents ensemble average. In addition, the error sources can be assumed to be statistically independent of one another. That is,

$$\begin{aligned}
 E[f(\varepsilon_a)g(\varepsilon_\phi)] &= E[f(\varepsilon_a)]E[g(\varepsilon_\phi)], \\
 E[f(\varepsilon_x)g(\varepsilon_y)] &= E[f(\varepsilon_x)]E[g(\varepsilon_y)],
 \end{aligned} \quad (6)$$

where  $f$  and  $g$  are arbitrary functions.

## II. BIAS AND RANDOM ERROR ON HOLOGRAM AND PREDICTION PLANE

Using Eqs. (2), (4), (5), and (6), we can derive the bias and random errors on the hologram plane as well as those on the prediction plane. The detailed derivations are shown in Appendix A. This section analyzes the results.

### A. Spatial distribution of bias and random error on hologram plane

Equations (A10) and (A11) express the bias and random error on the hologram plane due to the sensor mismatch as

$$b(\hat{p}_H(m, n)) = -\frac{1}{2} \sigma_\phi^2 p_H(m, n), \quad (7)$$

and

$$\sigma^2(\hat{p}_H(m, n)) = (\sigma_a^2 + \sigma_\phi^2) |p_H(m, n)|^2. \quad (8)$$

Equations (7) and (8) show that the bias and random error are proportional to the pressure on the hologram plane and the magnitude square of the pressure, respectively. The equations also state that the bias error is influenced only by the phase mismatch, but the random error is influenced by the magnitude as well as phase mismatch. To understand these new findings, it is convenient to consider a distribution of measured pressure in a complex plane (Fig. 2). Figure 2(a)



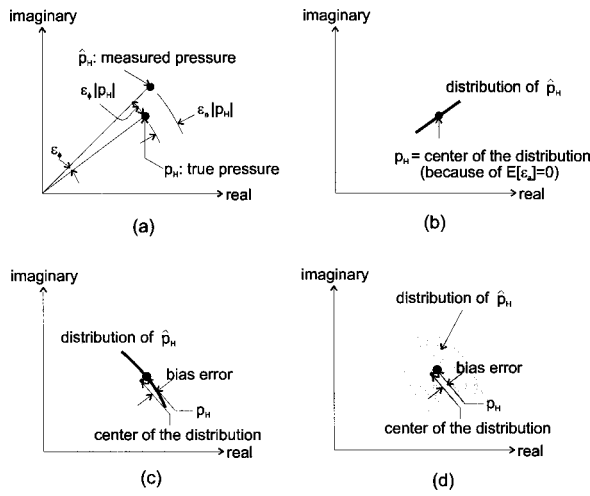


FIG. 2. Distribution of measured pressure in a complex plane due to sensor mismatch; (a) the measured pressure in a measurement event; (b) the distribution of the measured pressure when  $\epsilon_\phi=0$ ; (c) the distribution of the measured pressure when  $\epsilon_m=0$ ; (d) the distribution of the measured pressure when  $\epsilon_m \neq 0$  and  $\epsilon_\phi \neq 0$ .

shows the measured pressure at a point in a measurement event. The measured pressure is distorted in a radial direction by  $\epsilon_a|p_H|$ , and in a circumferential direction by  $\epsilon_\phi|p_H|$ . When  $\epsilon_\phi=0$ , the measured pressure in all the measurement events is distributed in the radial direction [Fig. 2(b)]. The center of the distribution is equal to the true value, because the shape of the distribution is a straight line and  $E[\epsilon_a]=0$ . When  $\epsilon_a=0$ , the measured pressure is distributed in the circumferential direction [Fig. 2(c)]. However, the center of the distribution is not equal to the true pressure in spite of  $E[\epsilon_\phi]=0$ , because the shape of the distribution is not a straight line. Therefore, there is a bias error. When  $\epsilon_a \neq 0$  and  $\epsilon_\phi \neq 0$ , the center of the distribution is different than the true value [Fig. 2(d)]. The phase mismatch mainly determines the difference. The effect of the magnitude mismatch is much less than that of phase mismatch. Therefore, only  $\sigma_\phi^2$  influences the bias error. However,  $\sigma_a^2$  as well as  $\sigma_\phi^2$  influence the random error, because they determine how wide the distribution of the measured pressure is.

Equations (A12) and (A13) express the bias and random error due to the position mismatch as

$$b(\hat{p}_H(m,n)) = \sigma_x^2 c_{xx}(m,n) + \sigma_y^2 c_{yy}(m,n), \quad (9)$$

and

$$\sigma^2(\hat{p}_H(m,n)) = \sigma_x^2 |v_x(m,n)|^2 + \sigma_y^2 |v_y(m,n)|^2. \quad (10)$$

Equations (9) and (10) show that the curvatures of the pressure influence the bias error, and the slopes influence the random error. To understand these findings, let us consider the case of  $\epsilon_y=0$  in a fixed  $y$  (Fig. 3). Figure 3(a) shows the measured pressure distorted by  $\epsilon_x$  in a measurement event. Figure 3(b) shows the distribution of  $\epsilon_x$  and the measured pressure. If the true pressure has a small variation within the distribution of  $\epsilon_x$ , the distribution length of the measured pressure is approximately the distribution length of  $\epsilon_x$  multiplied by the slope of the true pressure. Because the random error represents how long the distribution of the measured pressures is, the slope influences the random error. If the true

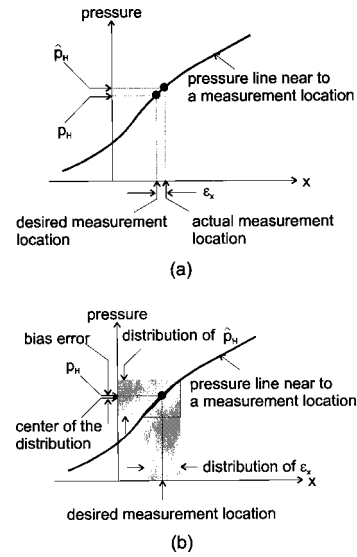


FIG. 3. Distribution of measured pressure due to position mismatch in a fixed  $y$  coordinate when  $\epsilon_y=0$ : (a) the measured pressure in a measurement event; (b) the distribution of  $\epsilon_x$  and the measured pressure.

pressure were a straight line within the distribution of  $\epsilon_x$ , the center of the distribution of the measured pressure would be equal to the true pressure. However, if there was curvature, the center of the distribution would be different than the true pressure. The difference is the bias error. Therefore, the curvature influences the bias error.

We have studied the errors on the hologram plane. The next step is to discuss the errors on the prediction plane.

## B. Spatial distribution of bias and random error on prediction plane

If one calculates the pressure on the prediction plane in wave number domain, the pressure on the prediction plane in space domain can be expressed as a circular convolution<sup>4</sup> between the pressure on the hologram plane and a propagator, which propagates the pressure from the hologram to the prediction plane.<sup>1</sup> That is,

$$p_Z(m,n) = \Delta^2 \sum_{m'} \sum_{n'} p_H(m',n') h_{ZH}(m-m'+pN, n-n'+qN), \quad (11)$$

where  $h_{ZH}$  is the propagator, and  $\Delta$  represent measurement spacing.  $p$  and  $q$  are integers so that  $-N/2 < m-m'+pN \leq N/2$  and  $-N/2 < n-n'+qN \leq N/2$ . Using this equation, the bias and random error on the prediction plane [Eqs. (A15) and (A19)] are expressed as

$$b(\hat{p}_Z(m,n)) = \Delta^2 \sum_{m'} \sum_{n'} b(\hat{p}_H(m',n')) h_{ZH}(m-m'+pN, n-n'+qN), \quad (12)$$

and

$$\sigma^2(\hat{p}_Z(m,n)) = \Delta^4 \sum_{m'} \sum_{n'} \sigma^2(\hat{p}_H(m',n')) |h_{ZH}(m-m'+pN, n-n'+qN)|^2. \quad (13)$$

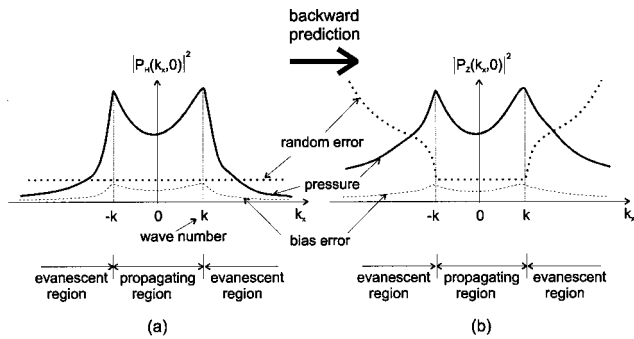


FIG. 4. Bias and random errors at  $k_y=0$  line in a wave number domain due to sensor mismatch; (a) on the hologram plane; (b) on the prediction plane.

Equation (12) shows that the bias error on the prediction plane is obtained by filtering the bias error on the hologram plane with the propagator  $h_{ZH}$ . Equation (13) shows that the random error on the prediction plane is obtained by filtering the random error on the hologram plane with  $|h_{ZH}|^2 \Delta^2$ . However, it is difficult to understand the errors on the prediction plane from Eqs. (12) and (13). The analysis of the errors in the wave number domain would provide an easier way to understand the physical meaning.

### C. Bias and random error in wave number domain

We define discrete Fourier transform from space domain to wave number domain as

$$P(r,s) = \Delta^2 \sum_m \sum_n p(m,n) e^{-j2\pi rm/N - j2\pi sn/N}, \quad (14a)$$

and define the inverse as

$$p(m,n) = \frac{1}{(N\Delta)^2} \sum_r \sum_s P(r,s) e^{j2\pi rm/N + j2\pi sn/N}, \quad (14b)$$

where  $r, s = -N/2 + 1, \dots, 0, \dots, N/2$ , which correspond to  $k_x$  and  $k_y$  in a continuous form, wave numbers in the  $x$  and  $y$  direction. The measurement spacing  $\Delta$  determines the highest wave number,  $k_{\max} = \pi/\Delta$ .<sup>1</sup>

Equations (A26) and (A27) express the bias and random error on the hologram plane in wave number domain due to the sensor mismatch as

$$b(\hat{P}_H(r,s)) = -\frac{1}{2} \sigma_\phi^2 P_H(r,s), \quad (15)$$

and

$$\sigma^2(\hat{P}_H(r,s)) = (\sigma_m^2 + \sigma_\phi^2) \frac{1}{N^2} \sum_{r'} \sum_{s'} |P_H(r',s')|^2. \quad (16)$$

Equations (A33) and (A34) express the bias and random error on the prediction plane in wave number domain as

$$b(\hat{P}_Z(r,s)) = H_{ZH}(r,s) b(\hat{P}_H(r,s)), \quad (17)$$

$$\sigma^2(\hat{P}_Z(r,s)) = |H_{ZH}(r,s)|^2 \sigma^2(\hat{P}_H(r,s)). \quad (18)$$

Thus the relation between the bias errors on the hologram and prediction plane is equal to that between pressures on the hologram and prediction plane in wave number domain. And, the relation between the random errors on the hologram

and prediction plane is equal to that between the magnitude squares of pressures on the hologram and prediction plane.

Equation (15) states that the bias error has the same shape as the pressure. However, the bias error would be much less than the pressure in the entire wave number domain because of small  $\sigma_\phi^2$  [Fig. 4(a)]. The bias error on the prediction plane would also be much less than the pressure on the prediction plane in the entire wave number domain, because the bias error is amplified as much as the pressure [Fig. 4(b)]. Therefore, the bias error is negligible. Equation (16) shows that the random error is constant in the entire wave number domain. This means that evanescent wave components would be distorted relatively largely, even though the random error energy is much less than the pressure energy [Fig. 4(a)]. Therefore, the backward prediction, which increases the evanescent wave component exponentially, can produce large errors [Fig. 4(b)].

Equations (A28) and (A29) express the bias and random error due to the position mismatch as

$$b(\hat{P}_H(r,s)) = \sigma_x^2 C_{xx}(r,s) + \sigma_y^2 C_{yy}(r,s), \quad (19)$$

and

$$\begin{aligned} \sigma^2(\hat{P}_H(r,s)) &= \sigma_x^2 \frac{1}{N^2} \sum_{r'} \sum_{s'} |V_x(r',s')|^2 \\ &+ \sigma_y^2 \frac{1}{N^2} \sum_{r'} \sum_{s'} |V_y(r',s')|^2, \end{aligned} \quad (20)$$

where  $V_{x,y}$  and  $C_{xx,yy}$  are the Fourier transforms of  $v_{x,y}$  and  $c_{xx,yy}$ . The expression of the bias and random error on the prediction plane in wave number domain are equal to Eqs. (17) and (18).

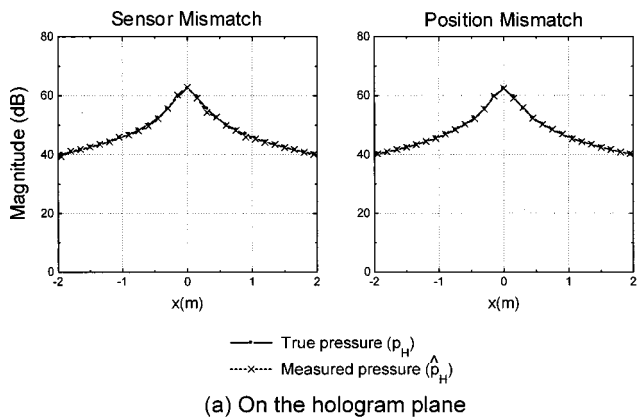
The curvatures in Eq. (19) are the second order derivatives of the pressure with respect to  $x$  and  $y$  in space domain. Therefore, Eq. (19) can be rewritten as

$$b(\hat{P}_H(r,s)) = -(\sigma_x^2 k_x^2(r,s) + \sigma_y^2 k_y^2(r,s)) P_H(r,s), \quad (21)$$

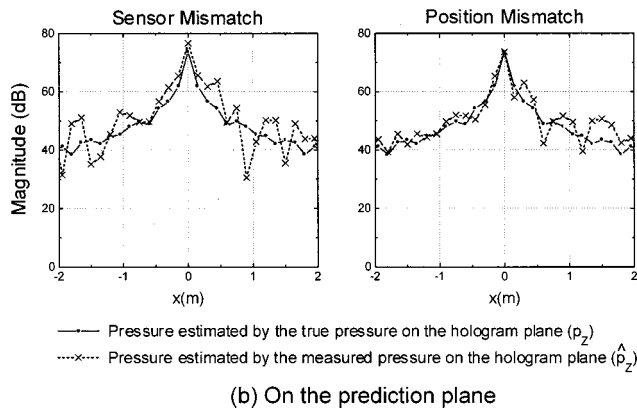
where  $k_x(r,s)$  and  $k_y(r,s)$  represent the wave numbers in continuous form, corresponding to  $r$  and  $s$  in discrete form. If the evanescent wave component decays more rapidly than the increase of  $k_x^2$  and  $k_y^2$  in the  $k_x$  and  $k_y$  direction, one can assume that the energy of the curvature would be concentrated inside the radiation circle. For example, the evanescent wave component of the monopole source decays exponentially in the  $k_x$  and  $k_y$  direction.<sup>5</sup> In such a case, the bias error would be much less than the pressure in the entire wave number domain because of a small  $\sigma_x^2$  and  $\sigma_y^2$ . The bias error on the prediction plane is also small in the entire wave number domain. Therefore, the bias error is negligible. Equation (20) shows that the random error is constant like sensor mismatch. Therefore, the backward prediction can produce large errors.

### D. Numerical simulation

To examine the theoretical results, we performed a numerical simulation. An acoustic field was generated by a monopole source located at the origin of Fig. 1. In Fig. 1,  $z_H$  is 0.15 m and  $z$  is 0.01 m. The frequency was 300 Hz and a



(a) On the hologram plane



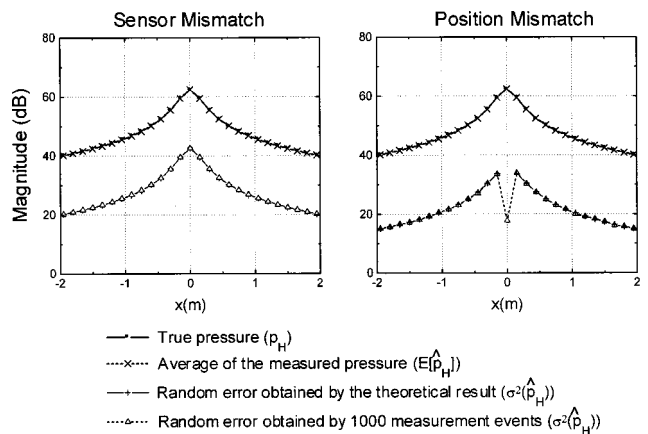
(b) On the prediction plane

FIG. 5. Magnitude of pressure on the hologram and prediction plane in a measurement event ( $\Delta = 0.15$  m,  $y = 0$ ).

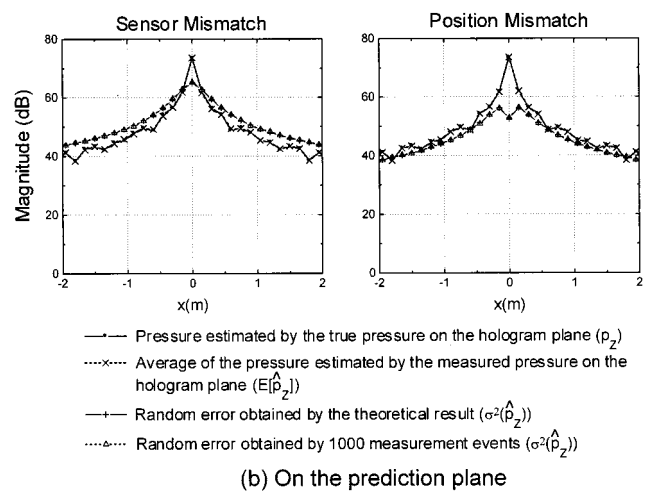
rectangular window was used. The measurement spacing  $\Delta$  was 0.15 m, and the number of the measurement points was  $32 \times 32$ .

For the sensor mismatch,  $\varepsilon_m$  was drawn arbitrarily from the population with the normal distribution, of which the mean was 0 and the standard deviation was 0.05.  $\varepsilon_\phi$  was drawn arbitrarily from the population with normal distribution, of which the mean was 0 and the standard deviation was  $5^\circ = 0.0872$  rad. The measurement values were obtained by substituting the drawn  $\varepsilon_m$  and  $\varepsilon_\phi$  for Eq. (1). For the position mismatch,  $\varepsilon_x$  and  $\varepsilon_y$  were drawn arbitrarily from the populations with the normal distribution, of which the means were 0 and the standard deviations were 0.01 m. The measurement values were the pressures at the  $(x + \varepsilon_x, y + \varepsilon_y)$  on the hologram plane.

Figure 5 shows the magnitude of the pressure at  $y = 0$  line in a measurement event. Figure 5(a) shows the results on the hologram plane. It compares the measured pressure ( $\hat{p}_H$ ) with the true pressure ( $p_H$ ). Figure 5(b) shows the results on the prediction plane. It compares the pressure ( $\hat{p}_Z$ ) estimated by the measured pressure with the pressure ( $p_Z$ ) estimated by the true pressure. If  $\hat{p}_Z$  is compared with the true pressure on the prediction plane, it is difficult to investigate only the effect of the measurement error because of the effect of aliasing and finite aperture size. Therefore, we plotted the pressure estimated by the true pressure in Fig. 5(b). Figure 5 shows small errors on the hologram plane, but large errors on the prediction plane. Figure 6 shows the results obtained by averaging 1000 measurement events. There are little differ-



(a) On the hologram plane



(b) On the prediction plane

FIG. 6. Statistical properties of pressure on the hologram and prediction plane obtained by 1000 measurement events ( $\Delta = 0.15$  m,  $y = 0$ ).  $10 \log_{10} \sigma^2$  is plotted for the random error.

ences between the true pressure and the average of the measured pressure on the hologram plane. It is also similar on the prediction plane. Therefore, the bias error can be regarded to be negligible. In Fig. 6, the random errors ( $10 \log_{10} \sigma^2$ ) are also plotted. The random errors obtained by the theoretical results highly agree with the ones by average. The random errors are highly amplified from the hologram to the prediction plane. The random errors due to the sensor and position mismatch have a different shape from each other. However, the global behaviors of the amplification seem to be similar. This motivates us to study the random errors in terms of their energies.

### III. AMPLIFICATION RATIO OF RANDOM ERROR IN TERMS OF ENERGY

#### A. Amplification ratio of random error energy

The detail derivations of the equations in this section are described in Appendix B.

We define the random error energy on the hologram and prediction plane,  $g_H$  and  $g_Z$ , as

$$g_H = \Delta^2 \sum_m \sum_n \sigma^2(\hat{p}_H(m,n)),$$

$$g_z = \Delta^2 \sum_m \sum_n \sigma^2(\hat{p}_z(m,n)).$$
(22)

It is noteworthy that Eq. (22) is the sum on  $m$  and  $n$  from  $-N/2+1$  to  $N/2$ . Although the hologram data are padded zeroes, the random error energy on the prediction plane does not eliminate the energy of the zero padded area. Therefore, the results in this chapter do not consider the zero padding effect.

For the sensor mismatch, the random error energy  $g_H$  can be approximated as

$$g_H = \Delta^2 \sum_m \sum_n (\sigma_m^2 + \sigma_\phi^2) |p_H(m,n)|^2$$

$$\approx (\sigma_m^2 + \sigma_\phi^2) \int_{-L/2}^{L/2} \int_{-L/2}^{L/2} |p_H(x,y)|^2 dx dy,$$
(23a)

and for the position mismatch,  $g_H$  is approximated as

$$g_H = \Delta^2 \sum_m \sum_n \{ \sigma_x^2 |v_x(m,n)|^2 + \sigma_y^2 |v_y(m,n)|^2 \}$$

$$\approx \sigma_x^2 \int_{-L/2}^{L/2} \int_{-L/2}^{L/2} |v_x(x,y)|^2 dx dy$$

$$+ \sigma_y^2 \int_{-L/2}^{L/2} \int_{-L/2}^{L/2} |v_y(x,y)|^2 dx dy,$$
(23b)

where  $L$  is the length of the aperture. The equations show that  $g_{H,s}$  are fixed when the variances of the error sources are fixed.

Equation (B2) expresses the relation between the random error energies on the hologram and prediction plane,  $g_H$  and  $g_z$ , as

$$g_z = g_H \times R_\sigma, \text{ where } R_\sigma = \Delta^4 \sum_m \sum_n |h_{ZH}(m,n)|^2.$$
(24)

We call  $R_\sigma$  the amplification ratio of random error. By Parseval's theorem, it can be rewritten as

$$R_\sigma = \frac{1}{N^2} \sum_r \sum_s |H_{ZH}(r,s)|^2,$$
(25)

where  $H_{ZH}$  is the Fourier transform of  $h_{ZH}$ . In continuous form,<sup>1</sup> it is expressed as

$$H_{ZH}(k_x, k_y) = e^{j\sqrt{k^2 - k_x^2 - k_y^2}(z - z_H)}.$$
(26)

Therefore,  $R_\sigma$  is greater than 1 in the backward prediction, because  $|H_{ZH}(r,s)| \geq 1$ .

The amplification ratio  $R_\sigma$  is the function of the following parameters: Measurement spacing  $\Delta$ , distance between the hologram and prediction plane  $d$ , and wave number  $k$ ; making the parameters nondimensional; then  $R_\sigma$  depends on  $d/\Delta$  and  $kd$  [Eq. (B4)].

Figure 7 is the contour plot of  $R_\sigma$  with respect to  $d/\Delta$  and  $kd$ . As  $d/\Delta$  increases,  $R_\sigma$  increases. It is because the reduction of  $\Delta$  increases  $k_{\max}$ , which is the maximum wave

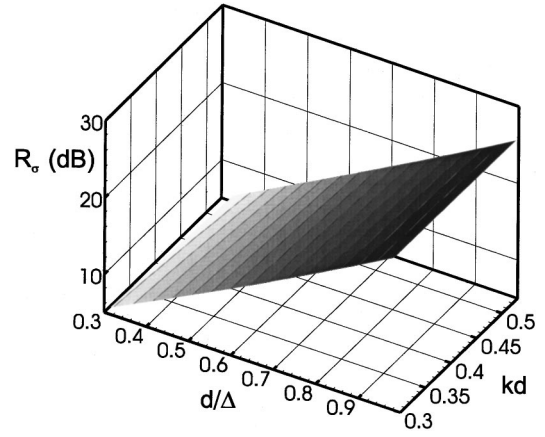


FIG. 7. Contour plot of  $R_\sigma$  with respect to  $d/\Delta$  and  $kd$ .

number. The increase of  $k_{\max}$  produces the increase of the maximum of  $|H_{ZH}(r,s)|^2$ . However, the effect of  $kd$  is negligible. Equation (B6) approximates  $R_\sigma$  (Fig. 7) as

$$10 \log_{10} R_\sigma \approx 24.9(d/\Delta) - 5.92(kd) + 1.55(d/\Delta)^2$$

$$+ 5.07(d/\Delta)(kd) - 0.25(kd)^2 \text{ dB.}$$
(27)

Equation (27) shows that the first order term of  $d/\Delta$  is dominant. In Fig. 6, the amplification ratio is about 23.5 dB as Eq. (27) predicts.

## B. Modified amplification ratio of random error

According to Sec. II C, the higher the wave number component, the relatively larger the distortion of the information. Therefore, it is natural to expect that better results would be obtained by eliminating the high wave number components, i.e., wave number filtering. However, the filtering reduces the resolution on the prediction plane. It means that the resolution on the prediction plane is different from the resolution on the hologram plane. To understand the effect of the difference clearly, let us consider an example of various filtering methods. This example uses the information only within  $-k'_{\max} \leq k_x \leq k'_{\max}$  and  $-k'_{\max} \leq k_y \leq k'_{\max}$  in the wave number domain, where  $k'_{\max} \leq k_{\max}$ . The prediction spacing is  $\Delta_Z = \pi/k'_{\max}$  and the measurement spacing is  $\Delta_H = \pi/k_{\max}$ . Then, the definitions of the random error energies in Eq. (22) will be modified as

$$g_H = \Delta_H^2 \sum_m \sum_n \sigma^2(\hat{p}_H(m,n)),$$
(28)

$$g_z = \Delta_Z^2 \sum_m \sum_n \sigma^2(\hat{p}_z(m,n)).$$

In the wave number domain, the random error energy per unit area is  $g_H/k_{\max}^2$ . Therefore, the random error energy within the area  $4k_{\max}'^2$  is  $g_H \times k_{\max}'^2/k_{\max}^2 = g_H \times \Delta_H^2/\Delta_Z^2$ . The random error energy  $g_H \times \Delta_H^2/\Delta_Z^2$  is amplified by  $R_\sigma$ . Therefore, Eq. (24) can be written as

$$g_z = g_H \times R'_\sigma \text{ where } R'_\sigma = R_\sigma \times \left( \frac{\Delta_H}{\Delta_Z} \right)^2,$$
(29)



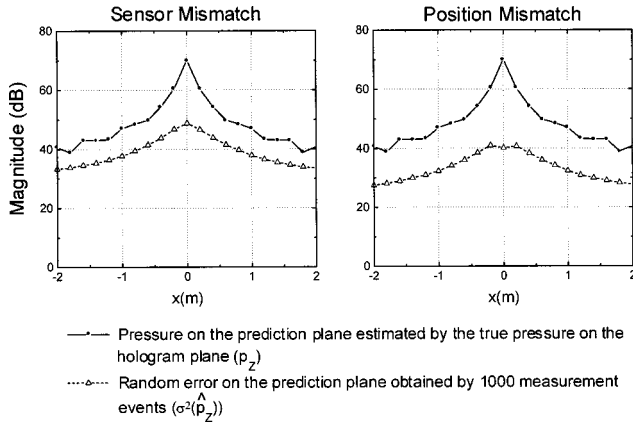


FIG. 8. The random error on the prediction plane obtained by 1000 measurement events ( $\Delta_H=0.1$  m,  $\Delta_Z=0.2$  m,  $y=0$ ).  $10 \log_{10} \sigma^2$  is plotted for the random error.

where  $R_\sigma$  is the function of  $d/\Delta_Z$  instead of  $d/\Delta$ , because  $k'_{\max}$  is the highest wave number component used for the prediction. We call  $R'_\sigma$  the modified amplification ratio of random error. In Eq. (29),  $R'_\sigma$  goes to zero for fixed  $\Delta_Z$ , as  $\Delta_H$  goes to zero. The area  $4k'^2_{\max}$  increases with the reduction of  $\Delta_H$ . The random error at each wave number component decreases so that the random error is constant in the entire wave number domain and total random error energy is the same. Thus the random error energy within the area  $4k'^2_{\max}$  decreases. Therefore, this filtering process can eliminate more random error energy by the reduction of  $\Delta_H$ . Equations (27) and (29) approximate  $R'_\sigma$  as

$$10 \log_{10} R'_\sigma \approx 24.9(d/\Delta_Z) - 5.92(kd) + 1.55(d/\Delta_Z)^2 + 5.07(d/\Delta_Z)(kd) - 0.25(kd)^2 + 20 \log_{10}(\Delta_H/\Delta_Z) \text{ dB.} \quad (30)$$

Figure 8 shows the random errors after the wave number filtering, where  $\Delta_H=0.1$  m and  $\Delta_Z=0.2$  m, the number of measurement points was  $48 \times 48$ , and the number of prediction points was  $24 \times 24$  for the same aperture size as the simulation in Sec. II D. Other conditions were the same as the conditions of the simulation in Sec. II D. The random errors on the prediction plane in 1000 measurement events are much less than those in Fig. 6.  $R'_\sigma$  is 10.2 dB as Eq. (30) predicts. Figure 9 shows that the error in a measurement event is significantly reduced by the reduction of the random error.

However, the increase of the prediction spacing must compromise with the reduction of the resolution on the prediction plane. The reduction of the measurement spacing must compromise the number of measurement points.

#### IV. CONCLUSIONS

We analyzed the effect of measurement errors on predicted pressure. Specifically, the sensor and position mismatch were dealt with. We analyzed the errors in terms of their probability, bias, and random error.

These errors were derived on the hologram and prediction plane. The bias error on the prediction plane was obtained by filtering the bias error on the hologram plane with

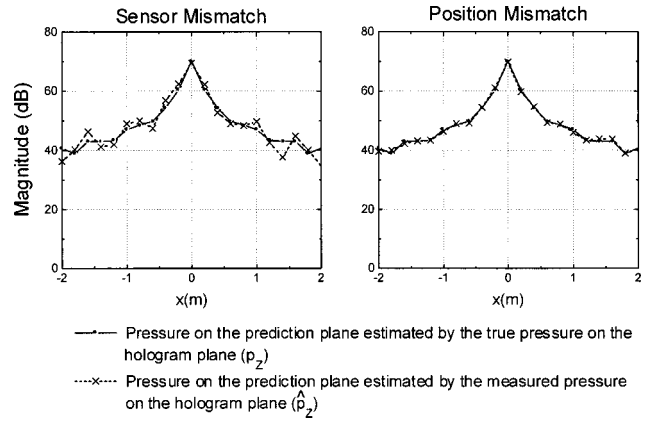


FIG. 9. Magnitude of pressure on the prediction plane in a measurement event ( $\Delta_H=0.1$  m,  $\Delta_Z=0.2$  m,  $y=0$ ).

the propagator. The random error on the prediction plane was obtained by filtering the random error on the hologram plane with the magnitude square of the propagator. The theoretical results were analyzed in the wave number domain as well as using simulations. The analysis drew a very important and useful conclusion. That is, the bias error is negligible, but the random error can be amplified significantly in the backward prediction.

The characteristics of the amplification of the random error are analyzed in terms of random error energy. That was defined as the summation of the random error. The amplification ratio of the random error energy was also analyzed. The amplification ratio was found to be related to the measurement parameters, the distance between the hologram and prediction plane ( $d$ ), the measurement spacing on the hologram plane ( $\Delta$ ), and the wave number ( $k$ ). The effect of the wave number was negligible. Thus the amplification ratio was approximately  $24.9d/\Delta$  on the dB scale. We also considered a wave number filtering process, which made the measurement spacing on the hologram plane ( $\Delta_H$ ) different than the prediction spacing on the prediction plane ( $\Delta_Z$ ). The modified amplification ratio was approximately  $24.9(d/\Delta_Z) + 20 \log_{10}(\Delta_H/\Delta_Z)$  on the dB scale.

#### APPENDIX A: THE DERIVATION OF BIAS AND RANDOM ERRORS

The measured pressure due to the sensor and position mismatch [Eqs. (2) and (4)] can be rewritten as

TABLE AI. Variables in sensor and position mismatch, corresponding to variables in Eqs. (A1)–(A2).

Equations (A1) and (A2)	Sensor mismatch	Position mismatch
$\varepsilon_1$	$\varepsilon_a$ (magnitude error)	$\varepsilon_x$ (position error in $x$ direction)
$\varepsilon_2$	$\varepsilon_\phi$ (phase error)	$\varepsilon_y$ (position error in $y$ direction)
$a_1$	$p_H$ (pressure)	$v_x$ (slope in $x$ direction)
$a_2$	$j p_H$	$v_y$ (slope in $y$ direction)
$b_{11}$	0	$c_{xx}$ (curvature)
$b_{12}$	$\frac{1}{2} j p_H$	$c_{xy}$ (curvature)
$b_{22}$	$-\frac{1}{2} p_H$	$c_{yy}$ (curvature)
$\sigma_1$	$\sigma_m$ (variance of $\varepsilon_a$ )	$\sigma_x$ (variance of $\varepsilon_x$ )
$\sigma_2$	$\sigma_\phi$ (variance of $\varepsilon_\phi$ )	$\sigma_y$ (variance of $\varepsilon_y$ )

$$\begin{aligned}\hat{p}_H(m,n) &\approx p_H(m,n) + a_1(m,n)\varepsilon_1(m,n) \\ &+ a_2(m,n)\varepsilon_2(m,n) + b_{11}(m,n)\varepsilon_1^2(m,n) \\ &+ b_{12}(m,n)\varepsilon_1(m,n)\varepsilon_2(m,n) \\ &+ b_{22}(m,n)\varepsilon_2^2(m,n),\end{aligned}\quad (\text{A1})$$

where  $\varepsilon_{1,2}$  are errors, and  $a_{1,2}$  and  $b_{11,12,22}$  are the weighting factors or contributions on the first and second order of the errors. It is noteworthy that  $\varepsilon_{1,2}$  are obviously real value numbers, however,  $a_{1,2}$  and  $b_{11,12,22}$  are complex to express magnitude and phase. In the sensor and position mismatch, the variables corresponding to  $\varepsilon_{1,2}$ ,  $a_{1,2}$ , and  $b_{11,12,22}$  are listed in Table AI. Equations (5) and (6) can be rewritten as

$$\begin{aligned}E[\varepsilon_{1,2}(m,n)] &= 0, \\ E[\varepsilon_{1,2}(m,n)\varepsilon_{1,2}(m',n')] &= \sigma_{1,2}^2\delta(m'-m)\delta(n'-n),\end{aligned}\quad (\text{A2})$$

$$E[f(\varepsilon_1)g(\varepsilon_2)] = E[f(\varepsilon_1)]E[g(\varepsilon_2)],\quad (\text{A3})$$

where  $\sigma_1^2$  and  $\sigma_2^2$  are the variances of the populations of  $\varepsilon_1$  and  $\varepsilon_2$ .  $f$  and  $g$  are arbitrary functions. These relations are necessary and sufficient to express that the errors are independent of each other. Using these equations, we derive the bias and random error on the hologram and prediction plane. If  $\theta$  and  $\hat{\theta}$  are true and estimated value, the bias and random error of  $\hat{\theta}$  are defined as

$$b(\hat{\theta}) = E[\hat{\theta}] - \theta,\quad (\text{A4})$$

$$\sigma^2(\hat{\theta}) = E[|\hat{\theta}|^2] - |E[\hat{\theta}]|^2.\quad (\text{A5})$$

Using Eqs. (A1)–(A3),  $E[\hat{p}_H(m,n)]$  and  $E[|\hat{p}_H(m,n)|^2]$  can be derived as

$$\begin{aligned}E[\hat{p}_H(m,n)] &\approx E[p_H(m,n) + a_1(m,n)\varepsilon_1(m,n) + a_2(m,n)\varepsilon_2(m,n) + b_{11}(m,n)\varepsilon_1^2(m,n) + b_{12}(m,n)\varepsilon_1(m,n)\varepsilon_2(m,n) \\ &+ b_{22}(m,n)\varepsilon_2^2(m,n)] = p_H(m,n) + \sigma_1^2 b_{11}(m,n) + \sigma_2^2 b_{22}(m,n),\end{aligned}\quad (\text{A6})$$

$$\begin{aligned}E[|\hat{p}_H(m,n)|^2] &\approx E[\{p_H(m,n) + a_1(m,n)\varepsilon_1(m,n) + a_2(m,n)\varepsilon_2(m,n) + b_{11}(m,n)\varepsilon_1^2(m,n) + b_{12}(m,n)\varepsilon_1(m,n)\varepsilon_2(m,n) \\ &+ b_{22}(m,n)\varepsilon_2^2(m,n)\}\{p_H^*(m,n) + a_1^*(m,n)\varepsilon_1(m,n) + a_2^*(m,n)\varepsilon_2(m,n) + b_{11}^*(m,n)\varepsilon_1^2(m,n) \\ &+ b_{12}^*(m,n)\varepsilon_1(m,n)\varepsilon_2(m,n) + b_{22}^*(m,n)\varepsilon_2^2(m,n)\}] \\ &= |p_H(m,n)|^2 + \sigma_1^2 |a_1(m,n)|^2 + \sigma_2^2 |a_2(m,n)|^2 + \{\sigma_1^2 b_{11}^*(m,n) + \sigma_2^2 b_{22}^*(m,n)\} p_H(m,n) \\ &+ \{\sigma_1^2 b_{11}(m,n) + \sigma_2^2 b_{22}(m,n)\} p_H^*(m,n),\end{aligned}\quad (\text{A7})$$

where the superscript “\*” represents complex conjugate. The higher order terms than the second order of  $\varepsilon$  were neglected in the derivation. Using Eqs. (A6) and (A7), the bias and random error on the hologram plane can be written as

$$\begin{aligned}b(\hat{p}_H(m,n)) &= E[\hat{p}_H(m,n)] - p_H(m,n) \\ &= \sigma_1^2 b_{11}(m,n) + \sigma_2^2 b_{22}(m,n),\end{aligned}\quad (\text{A8})$$

$$\begin{aligned}\sigma^2(\hat{p}_H(m,n)) &= E[|\hat{p}_H(m,n)|^2] - |E[\hat{p}_H(m,n)]|^2 \\ &= \sigma_1^2 |a_1(m,n)|^2 + \sigma_2^2 |a_2(m,n)|^2.\end{aligned}\quad (\text{A9})$$

Substituting the variables of Table AI for the above equations, the bias and random error on the hologram plane due to the sensor mismatch can be obtained as

$$b(\hat{p}_H(m,n)) = -\frac{1}{2}\sigma_\phi^2 p_H(m,n),\quad (\text{A10})$$

$$\sigma^2(\hat{p}_H(m,n)) = (\sigma_m^2 + \sigma_\phi^2) |p_H(m,n)|^2.\quad (\text{A11})$$

Similarly, the errors due to the position mismatch can be expressed as

$$b(\hat{p}_H(m,n)) = \sigma_x^2 c_{xx}(m,n) + \sigma_y^2 c_{yy}(m,n),\quad (\text{A12})$$

$$\sigma^2(\hat{p}_H(m,n)) = \sigma_x^2 |v_x(m,n)|^2 + \sigma_y^2 |v_y(m,n)|^2.\quad (\text{A13})$$

The next step is to derive the bias and random error on the prediction plane. Using Eqs. (11) and (A6),  $E[\hat{p}_z(m,n)]$  can be derived as

$$\begin{aligned}E[\hat{p}_z(m,n)] &= \Delta^2 \sum_{m'} \sum_{n'} E[\hat{p}_H(m',n')] h_{ZH}(m-m'+pN, \\ &n-n'+qN) \\ &= p_Z(m,n) + \Delta^2 \sum_{m'} \sum_{n'} \{\sigma_1^2 b_{11}(m',n') \\ &+ \sigma_2^2 b_{22}(m',n')\} h_{ZH}(m-m'+pN, \\ &n-n'+qN).\end{aligned}\quad (\text{A14})$$

Using Eqs. (A8) and (A14), the bias error on the prediction plane are derived as

$$\begin{aligned}b(\hat{p}_z(m,n)) &= E[\hat{p}_z(m,n)] - p_Z(m,n) \\ &= \Delta^2 \sum_{m'} \sum_{n'} \{\sigma_1^2 b_{11}(m',n') \\ &+ \sigma_2^2 b_{22}(m',n')\} h_{ZH}(m-m'+pN, \\ &n-n'+qN) \\ &= \Delta^2 \sum_{m'} \sum_{n'} b(\hat{p}_H(m',n')) h_{ZH}(m-m'+pN, \\ &n-n'+qN).\end{aligned}\quad (\text{A15})$$



Using Eq. (11),  $E[|\hat{p}_z(m,n)|^2]$  can be readily written as

$$\begin{aligned}
 E[|\hat{p}_z(m,n)|^2] &= E \left[ \left( \Delta^2 \sum_{m'} \sum_{n'} p_H(m',n') h_{ZH}(m-m'+pN,n-n'+qN) \right) \right. \\
 &\quad \left. \times \left( \Delta^2 \sum_{m''} \sum_{n''} p_H^*(m'',n'') h_{ZH}^*(m-m''+p'N,n-n''+q'N) \right) \right] \\
 &= \Delta^4 \sum_{m'} \sum_{n'} \sum_{m''} \sum_{n''} E[\hat{p}_H(m',n') \hat{p}_H^*(m'',n'')] \\
 &\quad \times h_{ZH}(m-m'+pN,n-n'+qN) h_{ZH}^*(m-m''+p'N,n-n''+q'N). \tag{A16}
 \end{aligned}$$

The correlation of the measured pressure  $E[\hat{p}_H(m',n') \hat{p}_H(m'',n'')]$  can be also readily derived as

$$\begin{aligned}
 E[\hat{p}_H(m',n') \hat{p}_H^*(m'',n'')] &\approx E[\{p_H(m',n') + a_1(m',n') \varepsilon_1(m',n') \\
 &\quad + a_2(m',n') \varepsilon_2(m',n') + b_{11}(m',n') \varepsilon_1^2(m',n') + b_{12}(m',n') \varepsilon_1(m',n') \varepsilon_2(m',n') + b_{22}(m',n') \varepsilon_2^2(m',n')\} \\
 &\quad \times \{p_H^*(m'',n'') + a_1^*(m'',n'') \varepsilon_1(m'',n'') + a_2^*(m'',n'') \varepsilon_2(m'',n'') + b_{11}^*(m'',n'') \varepsilon_1^2(m'',n'') \\
 &\quad + b_{12}^*(m'',n'') \varepsilon_1(m'',n'') \varepsilon_2(m'',n'') + b_{22}^*(m'',n'') \varepsilon_2^2(m'',n'')\}] \\
 &\approx p_H(m',n') p_H^*(m'',n'') + \{\sigma_1^2 a_1(m',n') a_1^*(m'',n'') + \sigma_2^2 a_2(m',n') a_2^*(m'',n'')\} \delta(m''-m') \delta(n''-n') \\
 &\quad + \{\sigma_1^2 b_{11}^*(m'',n'') + \sigma_2^2 b_{22}^*(m'',n'')\} p_H(m',n') + \{\sigma_1^2 b_{11}(m',n') + \sigma_2^2 b_{22}(m',n')\} p_H^*(m'',n''). \tag{A17}
 \end{aligned}$$

Substituting Eq. (A17) for Eq. (A16),  $E[|\hat{p}_z(m,n)|^2]$  can be obtained as

$$\begin{aligned}
 E[|\hat{p}_z(m,n)|^2] &\approx |p_z(m,n)|^2 + \Delta^4 \sum_{m'} \sum_{n'} \{\sigma_1^2 |a_1(m',n')|^2 + \sigma_2^2 |a_2(m',n')|^2\} |h_{ZH}(m-m'+pN,n-n'+qN)|^2 \\
 &\quad + p_z(m,n) \Delta^2 \sum_{m'} \sum_{n'} \{\sigma_1^2 b_{11}^*(m',n') + \sigma_2^2 b_{22}^*(m',n')\} h_{ZH}^*(m-m'+pN,n-n'+q'N) \\
 &\quad + p_z^*(m,n) \Delta^2 \sum_{m'} \sum_{n'} \{\sigma_1^2 b_{11}(m',n') + \sigma_2^2 b_{22}(m',n')\} h_{ZH}(m-m'+pN,n-n'+q'N). \tag{A18}
 \end{aligned}$$

Using Eqs. (A9), (A14), and (A18), the random error on the prediction plane will be expressed as

$$\begin{aligned}
 \sigma^2(\hat{p}_z(m,n)) &= E[|\hat{p}_z(m,n)|^2] - |E[\hat{p}_z(m,n)]|^2 \\
 &= \Delta^4 \sum_{m'} \sum_{n'} \{\sigma_1^2 |a_1(m',n')|^2 + \sigma_2^2 |a_2(m',n')|^2\} |h_{ZH}(m-m'+pN,n-n'+qN)|^2 \\
 &= \Delta^4 \sum_{m'} \sum_{n'} \sigma^2(\hat{p}_H(m',n')) |h_{ZH}(m-m'+pN,n-n'+qN)|^2. \tag{A19}
 \end{aligned}$$

Until now, the bias and random error on the hologram and prediction plane were derived in space domain. However, it is often easier to interpret the results in wave number domain. The next paragraph derives the bias and random error in wave number domain.

Using the definition of the Fourier transform in Eq. (14),  $E[\hat{P}_H(r,s)]$  in the wave number domain is derived as

$$\begin{aligned}
 E[\hat{P}_H(r,s)] &= \Delta^2 \sum_m \sum_{n'} E[\hat{p}_H(m,n)] e^{-j2\pi r m/N - j2\pi s n/N} \\
 &= P_H(r,s) + \sigma_1^2 B_{11}(r,s) + \sigma_2^2 B_{22}(r,s), \tag{A20}
 \end{aligned}$$

where  $B_{11,22}$  are the Fourier transforms of  $b_{11,22}$  in Eq. (A1). Using Eq. (A20), the bias error in wave number domain is derived as

$$\begin{aligned}
 b(\hat{P}_H(r,s)) &= E[\hat{P}_H(r,s)] - P_H(r,s) \\
 &= \sigma_1^2 B_{11}(r,s) + \sigma_2^2 B_{22}(r,s). \tag{A21}
 \end{aligned}$$

Using Eq. (14),  $E[|\hat{P}_H(r,s)|^2]$  is expressed as

$$\begin{aligned}
& E[|\hat{P}_H(r,s)|^2] \\
&= E\left[\left(\Delta^2 \sum_m \sum_n \hat{p}_H(m,n) e^{-j2\pi r m/N - j2\pi s n/N}\right)\right. \\
&\quad \left.\times \left(\Delta^2 \sum_{m'} \sum_{n'} \hat{p}_H^*(m',n') e^{j2\pi r m'/N + j2\pi s n'/N}\right)\right] \\
&= \Delta^4 \sum_m \sum_n \sum_{m'} \sum_{n'} E[\hat{p}_H(m,n) \hat{p}_H^*(m',n')] \\
&\quad \times e^{j2\pi r(m'-m)/N + j2\pi s(n'-n)/N}. \tag{A22}
\end{aligned}$$

Substituting Eq. (A17) for above equation,  $E[|\hat{P}_H(r,s)|^2]$  is rewritten as

$$\begin{aligned}
E[|\hat{P}_H(r,s)|^2] &= |P_H(r,s)|^2 + \Delta^4 \sum_m \sum_n \{ \sigma_1^2 |a_1(m,n)|^2 \\
&\quad + \sigma_2^2 |a_2(m,n)|^2 \} + \{ \sigma_1^2 B_{11}^*(r,s) \\
&\quad + \sigma_2^2 B_{22}^*(r,s) \} P_H(r,s) + \{ \sigma_1^2 B_{11}(r,s) \\
&\quad + \sigma_2^2 B_{22}(r,s) \} P_H^*(r,s). \tag{A23}
\end{aligned}$$

Using Parseval's theorem,  $E[|\hat{P}_H(r,s)|^2]$  is finally derived as

$$\begin{aligned}
E[|\hat{P}_H(r,s)|^2] &= |P_H(r,s)|^2 \\
&\quad + \frac{1}{N^2} \sum_{r'} \sum_{s'} \{ \sigma_1^2 |A_1(r',s')|^2 \\
&\quad + \sigma_2^2 |A_2(r',s')|^2 \} + \{ \sigma_1^2 B_{11}^*(r,s) \\
&\quad + \sigma_2^2 B_{22}^*(r,s) \} P_H(r,s) + \{ \sigma_1^2 B_{11}(r,s) \\
&\quad + \sigma_2^2 B_{22}(r,s) \} P_H^*(r,s), \tag{A24}
\end{aligned}$$

where  $A_{1,2}$  are the Fourier transforms of  $a_{1,2}$ . Using Eqs. (A20) and (A24), the random error in wave number domain is derived as

$$\begin{aligned}
\sigma^2(\hat{P}_H(r,s)) &= E[|\hat{P}_H(r,s)|^2] - |E[\hat{P}_H(r,s)]|^2 \\
&= \frac{1}{N^2} \sum_r \sum_s \{ \sigma_1^2 |A_1(r',s')|^2 \\
&\quad + \sigma_2^2 |A_2(r',s')|^2 \}. \tag{A25}
\end{aligned}$$

Substituting the variables in Table AI for Eqs. (A21) and (A25), the bias and random errors in the wave number domain due to the sensor mismatch are derived as

$$b(\hat{P}_H(r,s)) = -\frac{1}{2} \sigma_\phi^2 P_H(r,s), \tag{A26}$$

$$\sigma^2(\hat{P}_H(r,s)) = (\sigma_m^2 + \sigma_\phi^2) \frac{1}{N^2} \sum_{r'} \sum_{s'} |P_H(r',s')|^2. \tag{A27}$$

Similarly, the errors due to the position mismatch in the wave number domain are derived as

$$b(\hat{P}_H(r,s)) = \sigma_x^2 C_{xx}(r,s) + \sigma_y^2 C_{yy}(r,s), \tag{A28}$$

$$\begin{aligned}
\sigma^2(\hat{P}_H(r,s)) &= \sigma_x^2 \frac{1}{N^2} \sum_{r'} \sum_{s'} |V_x(r',s')|^2 \\
&\quad + \sigma_y^2 \frac{1}{N^2} \sum_{r'} \sum_{s'} |V_y(r',s')|^2, \tag{A29}
\end{aligned}$$

where  $V_{x,y}$  and  $C_{xx,yy}$  are the Fourier transforms of  $v_{x,y}$  and  $c_{xx,yy}$ .

Pressure on the prediction plane  $P_Z(r,s)$  is expressed as

$$P_Z(r,s) = H_{ZH}(r,s) P_H(r,s), \tag{A30}$$

where  $H_{ZH}(r,s)$  is the Fourier transform of the propagator  $h_{ZH}$ . Thus  $E[P_Z(r,s)]$  and  $E[|P_Z(r,s)|^2]$  are expressed is

$$E[P_Z(r,s)] = H_{ZH}(r,s) E[P_H(r,s)], \tag{A31}$$

$$E[|P_Z(r,s)|^2] = |H_{ZH}(r,s)|^2 E[|P_H(r,s)|^2]. \tag{A32}$$

Using these equations, the bias and random error on the prediction plane in the wave number domain are derived as

$$b(\hat{P}_Z(r,s)) = H_{ZH}(r,s) b(\hat{P}_H(r,s)), \tag{A33}$$

$$\sigma^2(\hat{P}_Z(r,s)) = |H_{ZH}(r,s)|^2 \sigma^2(\hat{P}_H(r,s)). \tag{A34}$$

## APPENDIX B: THE DERIVATION OF THE AMPLIFICATION RATIO OF RANDOM ERROR

Let us perform summation with respect to  $m$  and  $n$  in Eq. (13), and multiply by  $\Delta^2$ . Then

$$\begin{aligned}
& \Delta^2 \sum_m \sum_n \sigma^2(\hat{p}_Z(m,n)) \\
&= \Delta^6 \sum_m \sum_n \sum_{m'} \sum_{n'} \sigma^2(\hat{p}_H(m',n')) \\
&\quad \times |h_{ZH}(m-m'+pN, n-n'+qN)|^2 \\
&= \Delta^6 \sum_{m'} \sum_{n'} \left\{ \sigma^2(\hat{p}_H(m',n')) \sum_m \sum_n |h_{ZH}(m-m' \right. \\
&\quad \left. + pN, n-n'+qN)|^2 \right\}, \tag{B1}
\end{aligned}$$

where  $\sum_m \sum_n |h_{ZH}(m-m'+pN, n-n'+qN)|^2 = \sum_m \sum_n |h_{ZH} \times (m,n)|^2$ , because  $h_{ZH}(m-m'+pN, n-n'+qN)$  is the expression of the circular convolution of  $h_{ZH}(m,n)$ . Therefore, Eq. (B1) is rewritten as

$$\begin{aligned}
\Delta^2 \sum_m \sum_n \sigma^2(\hat{p}_Z(m,n)) &= \left\{ \Delta^2 \sum_m \sum_n \sigma^2(\hat{p}_H(m,n)) \right\} \\
&\quad \times \left\{ \Delta^4 \sum_m \sum_n |h_{ZH}(m,n)|^2 \right\}. \tag{B2}
\end{aligned}$$

In Eq. (B2),  $\Delta^2 \sum_m \sum_n \sigma^2(\hat{p}_H(m,n))$  and  $\Delta^2 \sum_m \sum_n \sigma^2(\hat{p}_Z(m,n))$  represent the random error energy on the hologram and prediction plane, respectively. We call  $\Delta^4 \sum_m \sum_n |h_{ZH}(m,n)|^2$  the amplification ratio of random error and express it as  $R_\sigma$ . By Parseval's theorem,  $R_\sigma$  is rewritten as

$$R_\sigma = \frac{1}{N^2} \sum_r \sum_s |H_{ZH}(r,s)|^2$$

$$\approx \frac{1}{4k_{\max}^2} \int_{-k_{\max}}^{k_{\max}} \int_{-k_{\max}}^{k_{\max}} |H_{ZH}(k_x, k_y)|^2 dk_x dk_y, \quad (\text{B3})$$

where  $H_{ZH}(r,s)$  is the Fourier transform of  $h_{ZH}(m,n)$ , and  $H_{ZH}(k_x, k_y)$  is the continuous form of the  $H_{ZH}(r,s)$ .  $k_{\max}$  is the maximum wave number obtained by measurement. Because  $k_{\max} = \pi/\Delta$ , and  $H_{ZH}(k_x, k_y) = e^{j\sqrt{k^2 - k_x^2 - k_y^2}(z - z_H)}$ ,  $R_\sigma$  is expressed as

$$R_\sigma \approx \frac{\Delta^2}{(2\pi)^2} \int_{-\pi/\Delta}^{\pi/\Delta} \int_{-\pi/\Delta}^{\pi/\Delta} |e^{j\sqrt{k^2 - k_x^2 - k_y^2}(z - z_H)}|^2 dk_x dk_y.$$

If  $d = z_H - z$ ,  $\alpha = k_x d$ ,  $\beta = k_y d$ , the above equation can be rewritten as

$$R_\sigma \approx \frac{1}{(2\pi)^2} \left(\frac{\Delta}{d}\right)^2$$

$$\times \int_{-\pi d/\Delta}^{\pi d/\Delta} \int_{-\pi d/\Delta}^{\pi d/\Delta} |e^{-j\sqrt{(kd)^2 - \alpha^2 - \beta^2}}|^2 d\alpha d\beta$$

$$= f(d/\Delta, kd). \quad (\text{B4})$$

That is,  $R_\sigma$  is the function of  $d/\Delta$  and  $kd$ .  $R_\sigma$  can be obtained by numerical integration. Figure 7 plots the result. When  $d = 0$ ,  $R_\sigma$  is 0 dB. Therefore,  $R_\sigma$  can be approximately expressed as

$$10 \log_{10} R_\sigma \approx c_1(d/\Delta) + c_2(kd) + c_{11}(d/\Delta)^2$$

$$+ c_{12}(d/\Delta)(kd) + c_{22}(kd)^2. \quad (\text{B5})$$

In order to determine the coefficients, we used the following five points in Fig. 7:  $(d/\Delta, kd, R_\sigma) = (1, 0.524, 26.0 \text{ dB})$ ,  $(0.333, 0.524, 6.20 \text{ dB})$ ,  $(1, 0.286, 26.2 \text{ dB})$ ,  $(0.333, 0.286, 7.26 \text{ dB})$ ,  $(0.5, 0.370, 11.6 \text{ dB})$ . Substituting above data for Eq. (B5), we determined the coefficients, and approximated  $R_\sigma$  as

$$10 \log_{10} R_\sigma \approx 24.9(d/\Delta) - 5.92(kd) + 1.55(d/\Delta)^2$$

$$+ 5.07(d/\Delta)(kd) - 0.25(kd)^2 \quad (\text{dB}). \quad (\text{B6})$$

The difference between Eq. (B6) and the data of Fig. 7 are less than 0.5 dB.

<sup>1</sup>J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).

<sup>2</sup>H.-S. Kwon and Y.-H. Kim, "Minimization of bias error due to windows in planar acoustic holography using a minimum error window," J. Acoust. Soc. Am. **98**, 2104–2111 (1995).

<sup>3</sup>J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedure*, 2nd ed. (Wiley, New York, 1986), Chap. 8, pp. 252–290.

<sup>4</sup>D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1984), pp. 60–74.

<sup>5</sup>H.-S. Kwon, "Sound visualization by using enhanced planar acoustic holographic reconstruction," Ph.D. Dissertation, Dept. of Mechanical Engineering, KAIST (1997).

# A new approach to reflection-transmission between viscoelastic half-spaces

G. Caviglia

*Dipartimento di Matematica, Università, Via Dodecaneso 35, 16146 Genova, Italy*

A. Morro

*Università, DIBE, Via Opera Pia 11a, 16145 Genova, Italy*

(Received 27 May 1997; revised 15 April 1998; accepted 10 June 1999)

Reflection and transmission of time-harmonic waves at the welded boundary between linear, viscoelastic, isotropic half-spaces is examined. A basis consisting of inhomogeneous waves is considered, for wave fields obeying Snell's law; the basis elements are parameterized by the component  $\mathbf{k}_{\parallel}$  of the wave vector parallel to the interface; incident and reflected-transmitted waves are expressed as superpositions of incoming and outgoing inhomogeneous waves. The characterization of incoming and outgoing inhomogeneous waves through the phase, the amplitude, and the mean energy intensity is shown to provide nonequivalent mathematical models for reflection and transmission. On the basis of the general properties of the energy flux, the energy criterion is regarded as the appropriate one. The resulting approach is developed in connection with the reflection-transmission process. In particular, the incoming or outgoing character of any inhomogeneous wave is shown to depend on  $\mathbf{k}_{\parallel}$ . In correspondence with such a dependence, jump discontinuities of the reflection and transmission coefficients are shown to occur. © 1999 Acoustical Society of America. [S0001-4966(99)03209-9]

PACS numbers: 43.20.Fn, 43.20.Bi [AN]

## INTRODUCTION

Reflection-transmission of time-harmonic waves at a solid-solid plane interface has been investigated both theoretically and numerically.<sup>1-10</sup> Dissipative solids have also been considered<sup>11-14</sup> by letting the stiffness matrix be complex valued. The particular attention paid to the behavior of plane waves is also due to the circumstance that, in linear materials, the results carry over to any time dependence through Fourier analysis.

Any investigation of reflection and transmission of time-harmonic waves shows some basic features. First, the reflected and transmitted waves may involve complex-valued wave vectors even though the incident wave is homogeneous in that the wave vector  $\mathbf{k}$  is real valued. This indicates that a self-consistent approach requires a systematic use of inhomogeneous waves where the real and imaginary parts,  $\Re\mathbf{k}$  and  $\Im\mathbf{k}$ , of  $\mathbf{k}$  need not be parallel. In addition, inhomogeneous waves are necessarily involved when the stiffness matrix is complex valued. Second, reflected and transmitted waves are regarded as superpositions of independent inhomogeneous (plane) waves with a common factor  $\exp[i(\mathbf{k}_{\parallel} \cdot \mathbf{x} - \omega t)]$ ;  $\mathbf{x}$  is the position vector,  $\omega$  is the angular frequency, and  $\mathbf{k}_{\parallel}$  is the wave vector parallel to the interface. By appealing to causality, the reflected and transmitted waves are taken to be generated by the basis elements that propagate away from the interface. Third, the reflection-transmission problem is solved through the requirement that displacement and traction be continuous across the interface; the resulting linear system provides the reflection and transmission coefficients, that is, the coefficients entering the superpositions of outgoing inhomogeneous waves.

The characterization of the direction of a wave can be

given different statements which prove to be nonequivalent. In viscoelasticity, an inhomogeneous wave is often regarded as incoming or outgoing according to the direction of the phase vector<sup>11</sup>  $\Re\mathbf{k}$ . In elasticity, an inhomogeneous wave is considered as outgoing if the amplitude decreases with distance from the interface (cf. Refs. 2, 7-10) while, for homogeneous waves, the sign of the time-averaged energy intensity at the interface is the property characterizing outgoing or incoming waves (cf. Refs. 6-10). It is an outstanding result of wave propagation within elastic anisotropic solids that consistency of reflection-transmission problems is guaranteed if  $\mathbf{k}_{\parallel}$  is real, the energy criterion is adopted for homogeneous waves, and the amplitude criterion is considered for inhomogeneous waves, in that the mean energy flux of inhomogeneous waves is parallel to the interface.<sup>6,8</sup>

In the present investigation of reflection and transmission between isotropic viscoelastic half-spaces we consider a basis of six inhomogeneous waves associated with a given  $\mathbf{k}_{\parallel}$ ; in general  $\Re\mathbf{k}$  is not collinear with  $\Im\mathbf{k}$ . By means of a numerical example we show that an inhomogeneous wave may prove to be outgoing or incoming, depending on the criterion adopted. The selection of outgoing waves is performed on the basis of the energy flux. For any value of  $\mathbf{k}_{\parallel}$ , exactly three waves prove to be outgoing in each half-space; hence the mathematical problem associated with the reflection-transmission process results in an algebraic system for the reflection and transmission coefficients whose solution exists and is unique, provided surface or interface waves are not allowed. Despite the pertinent peculiarities, the three criteria select the same waves when the value of the incidence angle is small or  $\mathbf{k}_{\parallel}$  is real.

With reference to a specific material, the direction of the energy flux associated with an inhomogeneous wave is

shown to depend on  $\mathbf{k}_{\parallel}$ . As a consequence, the basis for incident, reflected, and transmitted waves may change as  $\mathbf{k}_{\parallel}$  varies. Hence, the description of the reflection-transmission process, namely the selection of outgoing waves, is parameterized by  $\mathbf{k}_{\parallel}$ . Anyway, the partition into horizontally and vertically polarized waves still applies and is preserved at the interface.

There may be values of  $\mathbf{k}_{\parallel}$  such that the energy flux of two inhomogeneous waves is parallel to the interface. This is the analogue of the phenomenon of critical incidence. Around such values of  $\mathbf{k}_{\parallel}$  the reflected amplitude is shown to suffer a jump discontinuity.

## I. BASIC FRAMEWORK

We consider two viscoelastic, homogeneous, isotropic solids, occupying two half-spaces with a common plane boundary. We denote by  $\mathbf{x}$  the position vector of a point in the equilibrium unstressed configuration and by  $\mathbf{u}(\mathbf{x}, t)$  the displacement of point  $\mathbf{x}$  at time  $t$  from the reference position. Cartesian coordinates  $x, y, z$  with unit vectors  $\mathbf{e}_x, \mathbf{e}_y$ , and  $\mathbf{n} = \mathbf{e}_z$  are considered, such that the interface coincides with the  $(x, y)$ -plane. The incident wave is taken to propagate in the upper half-space  $z > 0$ .

We consider time-harmonic inhomogeneous waves and express the displacement field in the form

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{A} \exp[i(\mathbf{k} \cdot \mathbf{x} - \omega t)], \quad (1)$$

where  $\mathbf{A}$  and  $\mathbf{k}$  denote the vector amplitude and the wave vector; the angular frequency  $\omega$  is real while  $\mathbf{A}$  and  $\mathbf{k}$  are complex. Substitution for the viscoelastic stress tensor  $\mathbf{T}$  and the displacement  $\mathbf{u}$  from Eq. (1) into the equation of motion yields the propagation condition

$$(\lambda + \mu)(\mathbf{k} \cdot \mathbf{A}) \mathbf{k} + (\mu \mathbf{k} \cdot \mathbf{k} - \rho \omega^2) \mathbf{A} = 0, \quad (2)$$

where  $\rho$  is the density while  $\mu$  and  $\lambda$  are the complex, frequency-dependent, moduli in the upper half-space ( $z > 0$ ); they are defined by<sup>15</sup>

$$\mu = \mu_0 + \int_0^{\infty} \mu'(s) \exp(i\omega s) ds,$$

$$\lambda = \lambda_0 + \int_0^{\infty} \lambda'(s) \exp(i\omega s) ds.$$

The complex valuedness of  $\mu$  and  $\lambda$  makes it evident that, in general, the wave vector  $\mathbf{k}$  and the amplitude  $\mathbf{A}$  are necessarily complex valued.

At the interface, the waves are required to obey Snell's law. This means that the wave vectors of the pertinent waves have a common value of the horizontal part  $\mathbf{k}_{\parallel} = k_x \mathbf{e}_x + k_y \mathbf{e}_y$ , which is then determined by the incident wave. As a consequence, the dependence on  $x, y$ , and  $t$  is through the common exponential factor

$$\exp[i(\mathbf{k}_{\parallel} \cdot \mathbf{x} - \omega t)]. \quad (3)$$

Since  $k_x$  and  $k_y$  are complex we cannot choose any system of coordinates that makes  $k_x$  or  $k_y$  vanish. That is why the reflection-transmission problem is genuinely three-dimensional.

Consider any complex valued vector  $\mathbf{k}_{\parallel}$ . Substitution into Eq. (2) leads to the determination of six independent inhomogeneous waves.<sup>16</sup> These waves are involved in the formulation of the welded contact conditions at the interface, as continuity of the displacement  $\mathbf{u}$  and the traction  $\mathbf{t} = \mathbf{T} \mathbf{n}$ . Following Stroh<sup>8</sup> it is convenient to combine the displacement  $\mathbf{u}$  and the traction  $\mathbf{t}$  pertaining to each wave into a six-dimensional column vector  $\mathbf{p}$ . In block matrix form we have  $\mathbf{p} = [\mathbf{u}, \mathbf{t}]^T$ , where  $T$  means transpose. Let  $k_{\parallel}^2 = \mathbf{k}_{\parallel} \cdot \mathbf{k}_{\parallel}$ . The six vectors  $\mathbf{p}$ , associated with the independent inhomogeneous waves, are proportional to<sup>16</sup>

$$\mathbf{q}_L^{\pm} = \begin{bmatrix} \mathbf{k}_{\parallel} \pm \sigma_L \mathbf{n} \\ \pm a \mathbf{k}_{\parallel} + b \mathbf{n} \end{bmatrix} \exp[i(\mathbf{k}_L^{\pm} \cdot \mathbf{x} - \omega t)],$$

$$\mathbf{q}_T^{\pm} = \begin{bmatrix} \mp \sigma_T \mathbf{k}_{\parallel} + k_{\parallel}^2 \mathbf{n} \\ -b \mathbf{k}_{\parallel} \pm c \mathbf{n} \end{bmatrix} \exp[i(\mathbf{k}_T^{\pm} \cdot \mathbf{x} - \omega t)],$$

$$\mathbf{q}_{TH}^{\pm} = \begin{bmatrix} \mathbf{k}_{\parallel} \times \mathbf{n} \\ \pm i \mu \sigma_T \mathbf{k}_{\parallel} \times \mathbf{n} \end{bmatrix} \exp[i(\mathbf{k}_{TH}^{\pm} \cdot \mathbf{x} - \omega t)];$$

the symbol  $\mathbf{q}$ , rather than  $\mathbf{p}$ , is a reminder that the physical dimensions are not those of  $\mathbf{p}$ . The coefficients  $a, b, c$  are defined by

$$a = 2i\mu\sigma_L, \quad b = i\mu(\sigma_T^2 - k_{\parallel}^2) = i(\rho\omega^2 - 2\mu k_{\parallel}^2), \\ c = 2i\mu\sigma_T k_{\parallel}^2.$$

Let

$$h_{L,T} = \begin{cases} \rho\omega^2/(\lambda + 2\mu) - k_{\parallel}^2, & \text{for } L\text{-waves} \\ \rho\omega^2/\mu - k_{\parallel}^2, & \text{for } T, TH\text{-waves.} \end{cases} \quad (4)$$

It is understood that  $h_{L,T} \neq 0$ . The complex wave vectors  $\mathbf{k}$  are represented as

$$\mathbf{k}_{L,T}^{\pm} = \mathbf{k}_{\parallel} \pm \sigma_{L,T} \mathbf{n},$$

where

$$\sigma_{L,T} = \sqrt{h_{L,T}}. \quad (5)$$

For definiteness,  $\sigma_{L,T}$  is taken to have a positive real part or, if the real part of  $\sigma$  is zero, to have a positive imaginary part. The real and imaginary parts of  $\mathbf{k}$  are perpendicular to planes of constant phase and constant amplitude. The  $+, -$  signs refer to the sign of the  $z$ -component of the phase. The suffixes  $L, T, TH$  specify the longitudinal, transverse, horizontal polarization of the wave. For, the displacement is parallel to  $\mathbf{k}_L^{\pm} = \mathbf{k}_{\parallel} \pm \sigma_L \mathbf{n}$  in  $\mathbf{q}_L^{\pm}$  while it is perpendicular to  $\mathbf{k}_T^{\pm} = \mathbf{k}_{\parallel} \pm \sigma_T \mathbf{n}$  in  $\mathbf{q}_T^{\pm}$  and to both  $\mathbf{k}_{\parallel}$  and  $\mathbf{n}$  in  $\mathbf{q}_{TH}^{\pm}$ .

The inhomogeneous waves  $\mathbf{q}$  form a basis, in the sense that the traction displacement vector associated with any wave field proportional to Eq. (3) may always be expressed as a linear combination of the form

$$\mathbf{p} = \begin{bmatrix} \mathbf{u} \\ \mathbf{t} \end{bmatrix} = \alpha_L^+ \mathbf{q}_L^+ + \dots + \alpha_{TH}^- \mathbf{q}_{TH}^-,$$

where  $\alpha_L^+, \dots, \alpha_{TH}^-$  are complex valued coefficients with appropriate dimensions.

Let  $\mathcal{F}$  be the time-averaged energy flux vector; for time-harmonic waves<sup>16,17</sup>  $\mathcal{F} = -i\omega(\mathbf{T}\mathbf{u}^* - \mathbf{T}^*\mathbf{u})/4$  where  $*$  means



complex conjugate. We denote by  $\mathcal{F}$  the time-averaged intensity in the direction of the  $z$ -axis, at  $z=0$ , that is  $\mathcal{F}=\mathcal{F}\cdot\mathbf{n}$  or

$$\mathcal{F}(\mathbf{p})=-\frac{1}{4}i\omega(\mathbf{t}\cdot\mathbf{u}^*-\mathbf{t}^*\cdot\mathbf{u}).$$

If  $\mathbf{p}$  is proportional to any of the inhomogeneous waves  $\mathbf{q}_L^\pm$ ,  $\mathbf{q}_T^\pm$ ,  $\mathbf{q}_{TH}^\pm$ , we have

$$\begin{aligned}\mathcal{F}(\alpha\mathbf{q}_L^+)&=-\mathcal{F}(\alpha\mathbf{q}_L^-) \\ &=|\alpha|^2\omega[|\mathbf{k}_\parallel|^2\Re(\mu\sigma_L)+\frac{1}{2}\rho\omega^2\Re\sigma_L \\ &\quad -\Re(\sigma_L^*\mu k_\parallel^2)]\exp[-2\Im(\mathbf{k}_\parallel)\cdot\mathbf{x}],\end{aligned}\quad (6)$$

$$\begin{aligned}\mathcal{F}(\alpha\mathbf{q}_T^+)&=-\mathcal{F}(\alpha\mathbf{q}_T^-) \\ &=|\alpha|^2\omega\{k_\parallel^2(k_\parallel^2)^*\Re(\mu\sigma_T)+|\mathbf{k}_\parallel|^2[\frac{1}{2}\rho\omega^2\Re\sigma_T \\ &\quad -\Re(\sigma_T^*\mu k_\parallel^2)]\}\exp[-2\Im(\mathbf{k}_\parallel)\cdot\mathbf{x}],\end{aligned}\quad (7)$$

$$\begin{aligned}\mathcal{F}(\alpha\mathbf{q}_{TH}^+)&=-\mathcal{F}(\alpha\mathbf{q}_{TH}^-)=\frac{1}{2}\omega|\alpha|^2|\mathbf{k}_\parallel|^2\Re(\mu\sigma_T) \\ &\quad \times\exp[-2\Im(\mathbf{k}_\parallel)\cdot\mathbf{x}].\end{aligned}\quad (8)$$

## II. GENERATORS OF REFLECTED AND TRANSMITTED WAVES

Since the solution is sought in terms of (inhomogeneous) plane waves, in any reflection-transmission problem a basis of outgoing plane waves has to be selected for each half-space and hence the solution is determined as a superposition of the selected waves. In many cases a non-trivial question arises as to the criterion for the selection of the outgoing elementary waves. The question is also of practical importance in that the result of the reflection-transmission process, namely the reflected and transmitted waves, is affected by the choice of the outgoing waves.<sup>18,19</sup>

Any outgoing inhomogeneous waves is upgoing in the upper half-space  $z>0$  and downgoing in the lower half-space; the incident wave in the upper half-space is then downgoing. The literature shows three criteria to characterize upgoing and downgoing waves. We denote the criteria by the letters P, A, F, and list them as follows.

- P) An inhomogeneous wave is upgoing (downgoing) if  $\Re\sigma>0$  ( $<0$ ).
- A) An inhomogeneous wave is upgoing (downgoing) if  $\Im\sigma>0$  ( $<0$ ).
- F) An inhomogeneous wave is upgoing (downgoing) if  $\mathcal{F}>0$  ( $<0$ ).

This means that, an inhomogeneous wave is upgoing if  $\Re\mathbf{k}$  has a positive vertical component when P is chosen, if the amplitude decreases exponentially along  $z$  when A is chosen, and if the energy flux has a positive vertical component when F is chosen. Useful references are, e.g., Ref. 11 for the P criterion, Refs. 7–10 for the A criterion (in elasticity), and Refs. 6–10 for the F criterion (in elasticity).

The values of  $\mathbf{k}_\parallel$  such that  $\mathcal{F}=0$  for an inhomogeneous wave, in the upper or lower half-space, are termed *critical*.

Compatibility and uniqueness in the reflection-transmission process hold provided that:<sup>6,8</sup> (a)  $\mathbf{k}_\parallel$  is real, (b) a plane interface between elastic anisotropic solids is consid-

ered; (c) the F criterion is adopted for homogeneous waves; (d) the A criterion is adopted for inhomogeneous waves, in that the related energy flux  $\mathcal{F}$  vanishes; (e) surface or interface waves do not occur. Hence a combination of criteria A and F yields a consistent mathematical model for reflection-transmission, if (a), (b), and (e) hold. This weights against the use of the P criterion in viscoelasticity. In addition, the P criterion is subject to the objection that, at an inhomogeneous wave,  $(\Re\mathbf{k})\cdot(\Im\mathbf{k})$  may be negative which means that the amplitude may increase in the claimed direction of propagation.<sup>20</sup> Instead, as a consequence of the second law of thermodynamics (or the principle of dissipation of energy)  $\mathcal{F}\cdot(\Im\mathbf{k})>0$ , which means that, according to the F criterion, the amplitude of an inhomogeneous wave necessarily decreases in the direction of propagation.<sup>20</sup> That is a reason why the F criterion appears to be the appropriate one to determine the direction of propagation.

If the F criterion is applied it may happen that the amplitude of an outgoing wave increases with distance from the interface. Such may be the case if  $\mathcal{F}$  is not collinear to  $\mathbf{n}$ . The decay of the wave along  $\mathcal{F}$  is a general property,<sup>20</sup> the decay with distance from the interface need not hold even for the outgoing waves that are selected through the sign of  $\mathcal{F}$  (cf. Ref. 15, Chap. 1).

According to Eqs. (6)–(8), the inhomogeneous waves of the basis may be collected into pairs of opposite mean energy flux  $\mathcal{F}$ . Therefore application of F yields three upgoing and three downgoing waves. Note in passing that the same pairs yield waves of opposite direction according to criterion A, or P, although the directions are not necessarily consistent with those determined by F. As a consequence the polarization mode is preserved at the interface, independent of the criterion.

By way of a numerical example, it is of interest to consider the first layer of the Berkeley crustal model.<sup>21</sup> Letting  $\omega=10^2$  Hz we find that the pertinent parameters are

$$\rho=2.1\text{ g/cm}^3, \quad \mu=(12.10-i\,0.40)\text{ GPa},$$

$$2\mu+\lambda=(37.04-i\,0.55)\text{ GPa};$$

$$k_T^2=\rho\omega^2/\mu=1.73\,10^{-7}+i\,5.73\,10^{-9}\text{ cm}^{-2},$$

$$k_L^2=\rho\omega^2/(\lambda+2\mu)=5.66\,10^{-8}+i\,8.41\,10^{-10}\text{ cm}^{-2}.$$

To make evidence of possible differences in applying the P, A, and F criteria we let the incident wave be transverse, with  $\Im\mathbf{k}_T$  parallel to  $\Re\mathbf{k}_T$ . For simplicity, vector  $\mathbf{k}_\parallel$  is taken in the  $x$ -direction, that is  $k_y=0$ . Hence we let

$$k_x=k_T\sin\theta, \quad (9)$$

and examine the consequences, on the reflected longitudinal waves, as the incidence angle  $\theta$  varies on  $[0,\pi/2]$ .

Figure 1 shows the dependence of  $\Re\sigma_L$ ,  $\Im\sigma_L$  and  $\mathcal{F}(\mathbf{q}_L^+)$  as a function of the incidence angle. It is evident that  $\Re\sigma_L>0$  and  $\Im\sigma_L,\mathcal{F}(\mathbf{q}_L^+)<0$  as  $\theta>46^\circ$  while  $\Re\sigma_L,\mathcal{F}(\mathbf{q}_L^+)>0$  and  $\Im\sigma_L<0$  as  $23^\circ<\theta<46^\circ$ . Hence Fig. 1 provides a clear example of the nonequivalence of the P, A and F criteria. Also,  $\Re\sigma_L$ ,  $\Im\sigma_L$ , and  $\mathcal{F}(\mathbf{q}_L^+)$  have the same sign ( $>0$ ) as  $\theta$  is small, here  $\theta<23^\circ$ .

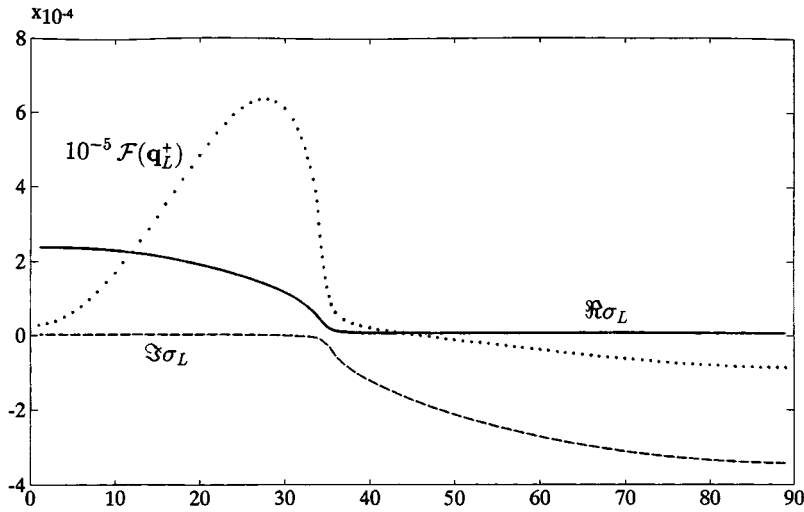


FIG. 1. Dependence of  $\Re\sigma_L$  (full line),  $\Im\sigma_L$  (dashed line), and  $10^{-5}\mathcal{F}(\mathbf{q}_L^+)$  (dotted line) on the incidence angle  $\theta$ .

The common sign of  $\Re\sigma$ ,  $\Im\sigma$ , and  $\mathcal{F}$  for small values of  $|\mathbf{k}_\parallel|$  that is, of the incidence angle is a general feature. Hence the P, A, and F criteria provide the same classification for small  $|\mathbf{k}_\parallel|$ . This feature is a consequence of the restrictions placed by thermodynamics, namely<sup>15</sup>

$$\Im\mu \leq 0, \quad \Im(2\mu + \lambda) \leq 0, \quad (10)$$

and the assumption that the elastic behavior is predominant, which implies that

$$\Re\mu > 0, \quad \Re(2\mu + \lambda) > 0 \quad (11)$$

as a consequence of  $\mu_0 > 0$ ,  $2\mu_0 + \lambda_0 > 0$ . It follows from Eq. (5) that

$$\Re\sigma_{L,T} = \sqrt{\frac{1}{2}[\Re h_{L,T} + |h_{L,T}|]},$$

$$\Im\sigma_{L,T} = \text{sgn}(\Im h_{L,T}) \sqrt{\frac{1}{2}[|h_{L,T}| - \Re h_{L,T}]},$$

while Eq. (8) yields

$$\mathcal{F}(\alpha \mathbf{q}_{TH}^+) = \frac{1}{2} |\alpha|^2 \omega |\mathbf{k}_\parallel|^2 (\Re\mu \Re\sigma_T - \Im\mu \Im\sigma_T) \times \exp[-2\Im(\mathbf{k}_\parallel) \cdot \mathbf{x}].$$

Now  $\Re\sigma_{L,T} > 0$ , for all values of  $\theta$ . Meanwhile by Eqs. (4) and (5) we have

$$\text{sgn}(\Im\sigma_T) = \text{sgn}\left(-\frac{\rho\omega^2\Im\mu}{|\mu|^2} - \Im k_\parallel^2\right), \quad (12)$$

$$\text{sgn}(\Im\sigma_L) = \text{sgn}\left(-\frac{\rho\omega^2\Im(2\mu + \lambda)}{|2\mu + \lambda|^2} - \Im k_\parallel^2\right). \quad (13)$$

It is apparent from Eq. (10) that, for small values of  $|\mathbf{k}_\parallel|$ , we have  $\Im\sigma_{T,L} > 0$ . Finally, by Eq. (8)

$$\text{sgn}(\mathcal{F}(\mathbf{q}_{TH}^+)) = \text{sgn}(\Re\mu \Re\sigma_T - \Im\mu \Im\sigma_T). \quad (14)$$

Because of Eq. (11), the condition  $\Im\sigma_T > 0$  for small values of  $|\mathbf{k}_\parallel|$  implies that also  $\mathcal{F}(\mathbf{q}_{TH}^+) > 0$  for small angles. Analogous results hold for  $\mathcal{F}(\mathbf{q}_L^+)$  and  $\mathcal{F}(\mathbf{q}_T^+)$ .

If  $\Im\mathbf{k}_\parallel = 0$  then we can choose the  $x$ -axis such that  $k_x = 0$ . Hence, in view of Eqs. (12) and (13), we have

$$\Im\sigma_T > 0, \quad \Im\sigma_L > 0.$$

Because of Eq. (14) it follows that

$$\mathcal{F}(\mathbf{q}_{TH}^+) > 0.$$

In the same way, by Eqs. (6) and (7) we obtain  $\mathcal{F}(\mathbf{q}_L^+), \mathcal{F}(\mathbf{q}_T^+) > 0$ . In conclusion, if  $\Im\mathbf{k}_\parallel = 0$  or  $|\mathbf{k}_\parallel|^2$  is sufficiently small we have, at the same time,

$$\Re\sigma > 0, \quad \Im\sigma > 0, \quad \mathcal{F} > 0;$$

the three criteria are then equivalent. This shows that the P, A, and F criteria are equivalent, in the lower-half space, when the incident wave is coming from an elastic half-space.

In general, an inhomogeneous wave is upgoing or downgoing depending on the criterion adopted and the value of  $\mathbf{k}_\parallel$ . Whatever criterion is adopted, the waves are partitioned into three pairs of opposite orientation; the three criteria provide the same results if  $|\mathbf{k}_\parallel|$  is small or  $\mathbf{k}_\parallel$  is real. Even though the F criterion is adopted, the same inhomogeneous wave, say  $\mathbf{q}_{TH}^+$ , may be upgoing or downgoing depending on the value of  $\mathbf{k}_\parallel$ . This implies that the selection of outgoing waves is affected by the value of  $\mathbf{k}_\parallel$ . In fact, as  $\mathbf{k}_\parallel$  varies, only a limited number of possibilities may occur, thus providing a few different bases of generators for reflection-transmission. This view is inevitable but, to our knowledge, has never been pointed out.

### III. REFLECTION AND TRANSMISSION

The six-dimensional vector  $\mathbf{p}$  is a superposition of the inhomogeneous waves  $\{\mathbf{q}\}$ . The continuity of  $\mathbf{u}$  and  $\mathbf{t}$  at the interface results into the continuity of  $\mathbf{p}$ , at  $z=0$ , and hence into the condition

$$A_L^+ \mathbf{q}_L^+ + A_L^- \mathbf{q}_L^- + A_T^+ \mathbf{q}_T^+ + A_T^- \mathbf{q}_T^- + A_{TH}^+ \mathbf{q}_{TH}^+ + A_{TH}^- \mathbf{q}_{TH}^-$$

$$= \hat{A}_L^+ \hat{\mathbf{q}}_L^+ + \hat{A}_L^- \hat{\mathbf{q}}_L^- + \hat{A}_T^+ \hat{\mathbf{q}}_T^+ + \hat{A}_T^- \hat{\mathbf{q}}_T^- + \hat{A}_{TH}^+ \hat{\mathbf{q}}_{TH}^+$$

$$+ \hat{A}_{TH}^- \hat{\mathbf{q}}_{TH}^-, \quad \text{at } z=0, \quad (15)$$

where a superposed  $\hat{\phantom{x}}$  labels quantities pertaining to the lower half-space. Upon the observation that Eq. (15) involves vectors parallel to  $\mathbf{k}_\parallel \times \mathbf{n}$  and vectors perpendicular to  $\mathbf{k}_\parallel \times \mathbf{n}$  we obtain the systems

$$A_{TH}^+ + A_{TH}^- = \hat{A}_{TH}^+ + \hat{A}_{TH}^-, \quad (16)$$

TABLE I. Incident, reflected, and transmitted waves for horizontal polarization.

Incident wave	Reflected wave	Transmitted wave	Reflected amplitude	Transmitted amplitude
$\mathbf{q}_{TH}^-$	$\mathbf{q}_{TH}^+$	$\hat{\mathbf{q}}_{TH}^-$	$A_{TH}^+ = \frac{\mu\sigma_T - \hat{\mu}\hat{\sigma}_T}{\mu\sigma_T + \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^-$	$\hat{A}_{TH}^- = \frac{2\mu\sigma_T}{\mu\sigma_T + \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^-$
$\mathbf{q}_{TH}^-$	$\mathbf{q}_{TH}^+$	$\hat{\mathbf{q}}_{TH}^+$	$A_{TH}^+ = \frac{\mu\sigma_T + \hat{\mu}\hat{\sigma}_T}{\mu\sigma_T - \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^-$	$\hat{A}_{TH}^+ = \frac{2\mu\sigma_T}{\mu\sigma_T - \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^-$
$\mathbf{q}_{TH}^+$	$\mathbf{q}_{TH}^-$	$\hat{\mathbf{q}}_{TH}^-$	$A_{TH}^- = \frac{\mu\sigma_T + \hat{\mu}\hat{\sigma}_T}{\mu\sigma_T - \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^+$	$\hat{A}_{TH}^- = \frac{2\mu\sigma_T}{\mu\sigma_T - \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^+$
$\mathbf{q}_{TH}^+$	$\mathbf{q}_{TH}^-$	$\hat{\mathbf{q}}_{TH}^+$	$A_{TH}^- = \frac{\mu\sigma_T - \hat{\mu}\hat{\sigma}_T}{\mu\sigma_T + \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^+$	$\hat{A}_{TH}^+ = \frac{2\mu\sigma_T}{\mu\sigma_T + \hat{\mu}\hat{\sigma}_T} \bar{A}_{TH}^+$

$$\mu\sigma_T(A_{TH}^+ - A_{TH}^-) = \hat{\mu}\hat{\sigma}_T(\hat{A}_{TH}^+ - \hat{A}_{TH}^-) \quad (17)$$

and

$$(A_L^+ + A_L^-) - \sigma_T(A_T^+ - A_T^-) = (\hat{A}_L^+ + \hat{A}_L^-) - \hat{\sigma}_T(\hat{A}_T^+ - \hat{A}_T^-), \quad (18)$$

$$b(A_L^+ + A_L^-) + c(A_T^+ - A_T^-) = \hat{b}(\hat{A}_L^+ + \hat{A}_L^-) + \hat{c}(\hat{A}_T^+ - \hat{A}_T^-), \quad (19)$$

$$\begin{aligned} \sigma_L(A_L^+ - A_L^-) + k_{\parallel}^2(A_T^+ + A_T^-) \\ = \hat{\sigma}_L(\hat{A}_L^+ - \hat{A}_L^-) + k_{\parallel}^2(\hat{A}_T^+ + \hat{A}_T^-), \end{aligned} \quad (20)$$

$$a(A_L^+ - A_L^-) - b(A_T^+ + A_T^-) = \hat{a}(\hat{A}_L^+ - \hat{A}_L^-) - \hat{b}(\hat{A}_T^+ + \hat{A}_T^-). \quad (21)$$

The system Eqs. (16)–(21) is equivalent to Eq. (15). The subsystem Eqs. (16)–(17) involves only the coefficients of horizontally polarized waves, whereas the subsystem Eqs. (18)–(21) involves the coefficients of vertically polarized waves; this means that vertical and horizontal polarizations are conserved at the interface.

Consider first the system Eqs. (16)–(17) for horizontally polarized waves. The incident wave is proportional to either  $\mathbf{q}_{TH}^-$  or  $\mathbf{q}_{TH}^+$ , according as  $\mathcal{F}(\mathbf{q}_{TH}^-) < 0$  or  $\mathcal{F}(\mathbf{q}_{TH}^-) > 0$ . In the former case  $A_{TH}^- = \bar{A}_{TH}^-$  is regarded as given and  $A_{TH}^+$  is unknown; in the latter case the roles of  $A_{TH}^-$  and  $A_{TH}^+$  are interchanged. Among the unknowns  $\hat{A}_{TH}^-$  and  $\hat{A}_{TH}^+$ , the coefficient of the incoming wave (upgoing) is taken to be zero. The results are summarized by Table I where the barred

quantities pertain to the incident wave.

If the direction of a wave is determined through the sign of  $\Re\sigma$  we say that  $\mathbf{q}_{TH}^-$  is incident,  $\mathbf{q}_{TH}^+$  is the reflected wave, and  $\hat{\mathbf{q}}_{TH}^-$  is the transmitted wave. Hence the first line of Table I collects the description and results of reflection and transmission based on the P criterion, which is the customary picture for homogeneous waves in elasticity.

As an example, the data for the upper half-space are taken to be those of the first layer (of the Berkeley crustal model) while, in the lower half-space, they are those of the second layer

$$\hat{\rho} = 2.6 \text{ g/cm}^3, \quad \hat{\mu} = (31.85 - i 0.71) \text{ GPa},$$

$$2\hat{\mu} + \hat{\lambda} = (96.74 - i 0.96) \text{ GPa}.$$

The value of  $\mathbf{k}_{\parallel} = k_x \mathbf{e}_x$  is parameterized by the incidence angle  $\theta$  in that  $k_x = k_T \sin \theta$ . Direct calculations show that  $\mathcal{F}(\mathbf{q}_{TH}^-) < 0$  and  $\mathcal{F}(\mathbf{q}_{TH}^+) > 0$  while

$$\mathcal{F}(\hat{\mathbf{q}}_{TH}^+) = -\mathcal{F}(\hat{\mathbf{q}}_{TH}^-) > 0, \quad \text{as } 0 \leq \theta \leq 73^\circ,$$

$$\mathcal{F}(\hat{\mathbf{q}}_{TH}^+) = -\mathcal{F}(\hat{\mathbf{q}}_{TH}^-) < 0, \quad \text{as } 74^\circ \leq \theta \leq 90^\circ$$

and  $\mathcal{F}(\hat{\mathbf{q}}_{TH}^+) = \mathcal{F}(\hat{\mathbf{q}}_{TH}^-) = 0$  at the critical angle  $\theta \approx 73.5^\circ$ . This implies that the outgoing waves are  $\mathbf{q}_{TH}^+$  and  $\hat{\mathbf{q}}_{TH}^-$  as  $\theta \leq 73^\circ$ ;  $\mathbf{q}_{TH}^+$  and  $\hat{\mathbf{q}}_{TH}^+$  as  $\theta \geq 74^\circ$ . Accordingly, as  $\theta$  changes around  $73.5^\circ$ , the pair of outgoing waves changes. Hence the result of the reflection-transmission process is given by the first line of Table I as  $\theta \leq 73^\circ$  and by the second line as  $\theta \geq 74^\circ$ . Figure 2 shows that, because of the change of outgo-

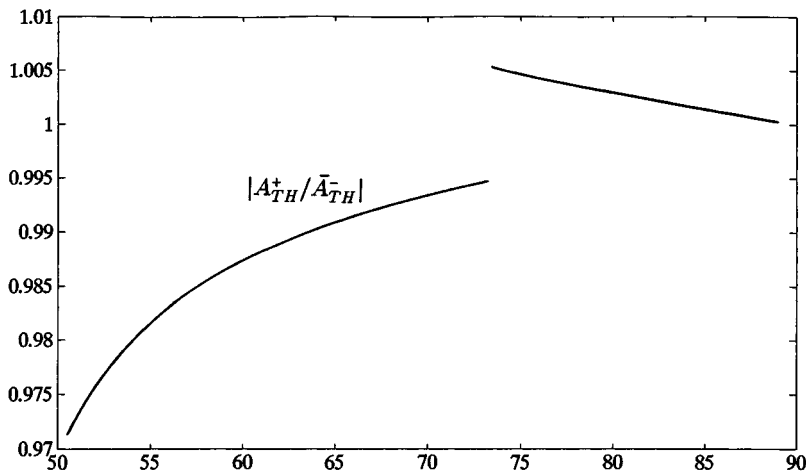


FIG. 2. The reflection coefficient  $|A_{TH}^+ / \bar{A}_{TH}^-|$  versus the incidence angle  $\theta$ .

ing waves, the reflection coefficient suffers a jump discontinuity. A jump discontinuity occurs for the transmission coefficient as well. As is the case in this example, the discontinuity of the reflection-transmission coefficients is likely to be relatively small and to occur for large values of the incidence angle.

We now consider vertically polarized waves. Equations (18)–(21) can be reduced to the equivalent form

$$A_L^+ = \alpha_{LL}^+ \hat{A}_L^+ + \alpha_{LL}^- \hat{A}_L^- + \alpha_{LT}^+ \hat{A}_T^+ + \alpha_{LT}^- \hat{A}_T^-, \quad (22)$$

$$A_L^- = \alpha_{LL}^- \hat{A}_L^+ + \alpha_{LL}^+ \hat{A}_L^- - \alpha_{LT}^+ \hat{A}_T^+ - \alpha_{LT}^- \hat{A}_T^-, \quad (23)$$

$$A_T^+ = \alpha_{TL}^+ \hat{A}_L^+ + \alpha_{TL}^- \hat{A}_L^- + \alpha_{TT}^+ \hat{A}_T^+ + \alpha_{TT}^- \hat{A}_T^-, \quad (24)$$

$$A_T^- = -\alpha_{TL}^- \hat{A}_L^+ - \alpha_{TL}^+ \hat{A}_L^- + \alpha_{TT}^- \hat{A}_T^+ + \alpha_{TT}^+ \hat{A}_T^-, \quad (25)$$

where

$$\alpha_{LL}^\pm = \frac{\sigma_L[\hat{\rho}\omega^2 + 2k_{\parallel}^2(\mu - \hat{\mu})] \pm \hat{\sigma}_L[\rho\omega^2 - 2k_{\parallel}^2(\mu - \hat{\mu})]}{2\rho\omega^2\sigma_L},$$

$$\alpha_{LT}^\pm = \frac{\mp 2\sigma_L\hat{\sigma}_T(\mu - \hat{\mu}) + [(\rho - \hat{\rho})\omega^2 - 2k_{\parallel}^2(\mu - \hat{\mu})]k_{\parallel}^2}{2\rho\omega^2\sigma_L},$$

$$\alpha_{TL}^\pm = \frac{-(\rho - \hat{\rho})\omega^2 + 2k_{\parallel}^2(\mu - \hat{\mu}) \pm 2\hat{\sigma}_L\sigma_T(\mu - \hat{\mu})}{2\rho\omega^2\sigma_T},$$

$$\alpha_{TT}^\pm = \frac{\pm \hat{\sigma}_T[\rho\omega^2 - 2k_{\parallel}^2(\mu - \hat{\mu})] + \sigma_T[\hat{\rho}\omega^2 + 2k_{\parallel}^2(\mu - \hat{\mu})]}{2\rho\omega^2\sigma_T}.$$

As an example, let  $\mathbf{p}^{\text{inc}} = \bar{A}_T^- \mathbf{q}_T^- + \bar{A}_L^- \mathbf{q}_L^-$  be the incident wave and hence  $\mathbf{p}^{\text{ref}} = A_T^+ \mathbf{q}_T^+ + A_L^+ \mathbf{q}_L^+$  is the reflected wave. If  $\hat{\mathbf{q}}_T^-$  and  $\hat{\mathbf{q}}_L^-$  are downgoing then  $\mathbf{p}^{\text{trans}} = \hat{A}_T^- \hat{\mathbf{q}}_T^- + \hat{A}_L^- \hat{\mathbf{q}}_L^-$  is the transmitted wave and  $\hat{A}_{L,T}^+ = 0$ . First we solve Eqs. (23) and (25) in the unknowns  $\hat{A}_L^-, \hat{A}_T^-$ . Next we replace in Eqs. (22) and (24) to find  $A_T^+$  and  $A_L^+$ . This particular case is the standard one if the P criterion is adopted.

In general, we find that

$$\begin{bmatrix} \hat{A}_L^\pm \\ \hat{A}_T^\pm \end{bmatrix} = \frac{1}{\alpha_{LL}^\mp \alpha_{TT}^\mp - \alpha_{TL}^\mp \alpha_{LT}^\mp} \begin{bmatrix} \alpha_{TT}^\mp & \alpha_{LT}^\mp \\ \alpha_{TL}^\mp & \alpha_{LL}^\mp \end{bmatrix} \begin{bmatrix} \bar{A}_L^\mp \\ \bar{A}_T^\mp \end{bmatrix},$$

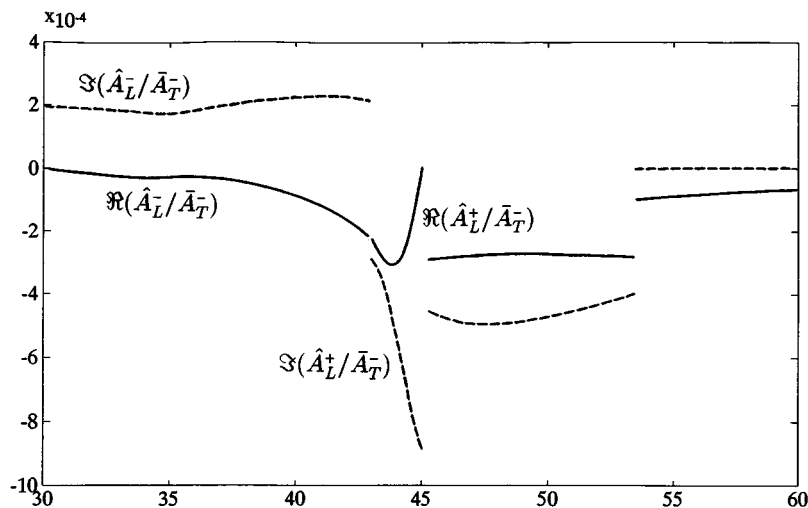


FIG. 3. Real (full line) and imaginary (dotted line) parts of the transmission coefficient for the longitudinal wave versus the incidence angle  $\theta$ .

TABLE II. Reflected and transmitted waves for transverse, vertically polarized, incident waves.

Incidence angle	Reflected $L$	Reflected $T$	Transmitted $L$	Transmitted $T$
$1^\circ \leq \theta \leq 42^\circ$	$\mathbf{q}_L^+$	$\mathbf{q}_T^+$	$\hat{\mathbf{q}}_L^-$	$\hat{\mathbf{q}}_T^-$
$43^\circ \leq \theta \leq 45^\circ$	$\mathbf{q}_L^+$	$\mathbf{q}_T^+$	$\hat{\mathbf{q}}_L^+$	$\hat{\mathbf{q}}_T^-$
$46^\circ \leq \theta \leq 53^\circ$	$\mathbf{q}_L^-$	$\mathbf{q}_T^+$	$\hat{\mathbf{q}}_L^+$	$\hat{\mathbf{q}}_T^-$
$54^\circ \leq \theta \leq 89^\circ$	$\mathbf{q}_L^-$	$\mathbf{q}_T^+$	$\hat{\mathbf{q}}_L^+$	$\hat{\mathbf{q}}_T^+$

$$\begin{bmatrix} \hat{A}_L^\pm \\ \hat{A}_T^\pm \end{bmatrix} = \frac{1}{\alpha_{LL}^\mp \alpha_{TT}^\mp + \alpha_{TL}^\mp \alpha_{LT}^\mp} \begin{bmatrix} \alpha_{TT}^\mp & \alpha_{LT}^\mp \\ -\alpha_{TL}^\mp & \alpha_{LL}^\mp \end{bmatrix} \begin{bmatrix} \bar{A}_L^\mp \\ \bar{A}_T^\mp \end{bmatrix},$$

$$\begin{bmatrix} \hat{A}_L^\pm \\ \hat{A}_T^\pm \end{bmatrix} = \frac{1}{\alpha_{LL}^\mp \alpha_{TT}^\mp + \alpha_{TL}^\mp \alpha_{LT}^\mp} \begin{bmatrix} \alpha_{TT}^\mp & -\alpha_{LT}^\mp \\ \alpha_{TL}^\mp & \alpha_{LL}^\mp \end{bmatrix} \begin{bmatrix} \bar{A}_L^\mp \\ \bar{A}_T^\mp \end{bmatrix},$$

$$\begin{bmatrix} \hat{A}_L^\pm \\ \hat{A}_T^\pm \end{bmatrix} = \frac{1}{\alpha_{LL}^\mp \alpha_{TT}^\mp - \alpha_{TL}^\mp \alpha_{LT}^\mp} \begin{bmatrix} \alpha_{TT}^\mp & -\alpha_{LT}^\mp \\ -\alpha_{TL}^\mp & \alpha_{LL}^\mp \end{bmatrix} \begin{bmatrix} \bar{A}_L^\mp \\ \bar{A}_T^\mp \end{bmatrix}.$$

Substitution in the pertinent equations of Eqs. (22)–(25) yields the reflected amplitudes  $A_{L,T}^\pm$ .

Table II gives a scheme of outgoing waves, associated with a vertically polarized transverse incident wave, as  $k_x = k_T \sin \theta$  varies.

Figure 3 shows the real and imaginary parts of the transmission coefficient for the longitudinal wave as  $\theta \in [30^\circ, 60^\circ]$ . Discontinuities of the coefficient occur at the critical angles (about  $42.7^\circ$ ,  $45.1^\circ$ ,  $53.2^\circ$ ). Discontinuities occur also for the other reflection and transmission coefficients at the same incidence angles.

It is worth appending a comment on the occurrence of jump discontinuities for the reflection and transmission coefficients. The jumps occur because there are values of the incidence angle such that the intensity  $\mathcal{F}$  associated with an inhomogeneous wave vanishes and hence the set of outgoing waves changes. The system in the unknown reflection and transmission coefficients changes too and hence the coefficients will generally suffer a jump. The reflected and transmitted intensities  $\mathcal{F}$  need not suffer jump discontinuities. For instance, look at transverse horizontal waves. The reflection and transmission coefficients are discontinuous at the critical

angle  $\theta_c = 73.5^\circ$  (cf. Fig. 2) such that  $\mathcal{F}(\hat{\mathbf{q}}_{TH}^-) = \mathcal{F}(\hat{\mathbf{q}}_{TH}^+) = 0$ . The transmitted intensity  $\hat{\mathcal{F}}$ , instead, is continuous. This is so because  $\hat{\mathcal{F}} = |\hat{A}_{TH}^-|^2 \mathcal{F}(\hat{\mathbf{q}}_{TH}^-)$ , as  $\theta \in (0, \theta_c)$ , and  $\hat{\mathcal{F}} = |\hat{A}_{TH}^+|^2 \mathcal{F}(\hat{\mathbf{q}}_{TH}^+)$ , as  $\theta \in (\theta_c, \pi/2)$ , depend continuously on  $\theta$ .

#### IV. CONCLUSIONS

For any fixed value of  $\mathbf{k}_\parallel$ , a basis of six inhomogeneous waves for fields obeying Snell's law is determined for each half-space. Reflected and transmitted waves are represented as superpositions of three outgoing basis elements. The selection of the outgoing waves is made through the sign of the averaged intensity. The other criteria (through  $\Im \mathbf{k}$  or  $\Re \mathbf{k}$ ) are generally nonequivalent. However, the three criteria become equivalent when  $|\mathbf{k}_\parallel|$  is sufficiently small or  $\mathbf{k}_\parallel$  takes real values. In such cases, the usual schemes for reflection and transmission are recovered.

More involved situations occur for large values of  $|\mathbf{k}_\parallel|$ . In correspondence with a vector  $\mathbf{k}_\parallel$  that makes one of the energy intensities Eqs. (6)–(8) be zero, an upgoing inhomogeneous wave becomes downgoing and viceversa. Hence the partition of the inhomogeneous waves into outgoing and incoming depends on  $\mathbf{k}_\parallel$ . Accordingly, as  $\mathbf{k}_\parallel$  varies, the set of outgoing inhomogeneous waves may change thus affecting the description of the reflection-transmission process. The analytic expressions for the reflection and transmission coefficients have been determined in correspondence with each admissible set of outgoing waves. The incident wave selects the appropriate set through the parallel component  $\mathbf{k}_\parallel$  of the wave vector.

Because of the change of outgoing waves, both horizontally polarized and vertically polarized waves may give rise to a jump discontinuity in the reflection and transmission coefficients. As shown by some examples, the discontinuity is likely to be quite small and perhaps difficult to detect. Furthermore, the energy flux in any half-space need not suffer jump discontinuities as  $\mathbf{k}_\parallel$  varies.

While this work was under review we have become aware of a paper on reflection and transmission of antiplane shear waves at a plane boundary between monoclinic viscoelastic media.<sup>22</sup> It is a common feature with our approach that the direction of the wave is taken to be that of the average energy flux. Instead, the consequences on the selection of the inhomogeneous waves in reflection-transmission processes are investigated in this paper for the first time.

#### ACKNOWLEDGMENTS

The authors are grateful to one of the referees for helpful remarks on a previous version. The research leading to this

work was supported by the MURST98 Research Project "Mathematical Models for Materials Science."

- <sup>1</sup>A. Atalar, "Reflection of ultrasonic waves at a liquid-cubic-solid interface," J. Acoust. Soc. Am. **73**, 435–440 (1983).
- <sup>2</sup>B. Mandal, "Reflection and transmission properties of elastic waves on a plane interface for general anisotropic media," J. Acoust. Soc. Am. **90**, 1106–1118 (1991).
- <sup>3</sup>O. Arikan, E. Teletar, and A. Atalar, "Reflection coefficient null of acoustic waves at a liquid-anisotropic-solid interface," J. Acoust. Soc. Am. **85**, 1–10 (1989).
- <sup>4</sup>A. H. Nayfeh, "The general problem of elastic wave propagation in multilayered anisotropic media," J. Acoust. Soc. Am. **89**, 1521–1531 (1991).
- <sup>5</sup>A. H. Nayfeh and D. E. Chimenti, "Elastic wave propagation in fluid-loaded multiaxial anisotropic media," J. Acoust. Soc. Am. **89**, 542–549 (1991).
- <sup>6</sup>T. W. Wright, "A note on oblique reflections in elastic crystals," Q. J. Mech. Appl. Math. **29**, 15–24 (1976).
- <sup>7</sup>J. Bazer and R. Burridge, "Energy partition in the reflection and refraction of plane waves," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **34**, 78–92 (1978).
- <sup>8</sup>A. N. Stroh, "Steady-state problems in anisotropic elasticity," J. Math. Phys. **41**, 77–103 (1962).
- <sup>9</sup>P. Lancelleur, H. Ribeiro, and J. F. DeBelleval, "The use of inhomogeneous waves in the reflection-transmission problem at a plane interface between two anisotropic media," J. Acoust. Soc. Am. **93**, 1882–1892 (1993).
- <sup>10</sup>S. I. Rokhlin, T. K. Bolland, and L. Adler, "Reflection and refraction of elastic waves on a plane interface between two generally anisotropic media," J. Acoust. Soc. Am. **79**, 906–918 (1986).
- <sup>11</sup>H. F. Cooper, "Reflection and transmission of oblique plane waves at a plane interface between viscoelastic media," J. Acoust. Soc. Am. **42**, 1064–1069 (1967).
- <sup>12</sup>M. Schoenberg, "Reflection and transmission of plane waves at an elastic viscoelastic interface," Geophys. J. R. Astron. Soc. **25**, 35–47 (1971).
- <sup>13</sup>M. Deschamps and B. Hosten, "The effects of viscoelasticity on the reflection and transmission of ultrasonic waves by an orthotropic plate," J. Acoust. Soc. Am. **91**, 2007–2015 (1992).
- <sup>14</sup>R. D. Borchardt, "Reflection and refraction of elastic waves for stratified materials," Wave Motion **10**, 333–348 (1988).
- <sup>15</sup>G. Caviglia and A. Morro, *Inhomogeneous Waves in Solids and Fluids* (World Scientific, Singapore, 1992).
- <sup>16</sup>G. Caviglia and A. Morro, "Waves at a bonded interface between dissipative solids," Q. J. Mech. Appl. Math. **47**, 305–322 (1994).
- <sup>17</sup>J. L. Synge, "Flux of energy for elastic waves in anisotropic media," Proc. R. Ir. Acad. A, Math. Phys. Sci. **58**, 13–23 (1956).
- <sup>18</sup>G. Caviglia and A. Morro, "Nonuniqueness of reflected and transmitted waves," Mech. Res. Commun. **23**, 571–576 (1996).
- <sup>19</sup>G. Caviglia and A. Morro, "On the modelling of reflection-transmission problems," Math. Models Methods Appl. Sci. **8**, 875–896 (1998).
- <sup>20</sup>G. Caviglia and A. Morro, "Inhomogeneous waves in anisotropic dissipative bodies," Continuum Mech. Thermodyn. **7**, 231–248 (1995).
- <sup>21</sup>W. Silva, "Body waves in a layered anelastic solid," Bull. Seismol. Soc. Am. **66**, 1539–1554 (1976).
- <sup>22</sup>J. M. Carcione, "Reflection and refraction of antiplane shear waves at a plane boundary between viscoelastic anisotropic media," Proc. R. Soc. London, Ser. A **453**, 919–942 (1997).



# Backscattering enhancements due to reflection of meridional leaky Rayleigh waves at the blunt truncation of a tilted solid cylinder in water: Observations and theory

Karen Gipson<sup>a)</sup> and Philip L. Marston<sup>b)</sup>

*Department of Physics, Washington State University, Pullman, Washington 99164-2814*

(Received 6 August 1998; revised 21 June 1999; accepted 22 June 1999)

Leaky surface waves once launched on an elastic object in water will undergo partial reflection at the truncations of the object. Under certain conditions, such reflections can radiate locally flat acoustic wavefronts with a reversal of the acoustic wave vector, and thus a large backscattered signal is produced. A stainless-steel circular rod with flat ends was ensonified through an angular region appropriate for launching meridional Rayleigh waves, and the resulting backscattering enhancement is consistent with the hypothesized end-reflection process. The observed backscattering peak is in general agreement with predictions from a convolution formulation [P. L. Marston, *J. Acoust. Soc. Am.* **102**, 358–369 (1997)]. The tilt of the cylinder was varied to explore the peak's width, and the results are consistent with an extension of the theory for the dephasing of the Rayleigh wave given here. The amplitude measurements used ultrasonic tone bursts having  $ka$  of 50 and 83, and the enhancement is predicted to be significant over a wide range of  $ka$ . © 1999 Acoustical Society of America. [S0001-4966(99)02910-0]

PACS numbers: 43.20.Fn, 43.35.Pt, 43.30.Gv [AN]

## INTRODUCTION

Guided wave contributions are known to contribute significantly to scattering from solid elastic spheres and cylinders.<sup>1–7</sup> Therefore, when contributions to scattering cannot be explained by specular reflection, it is sensible to consider the conditions under which guided waves supported by the elastic object can strongly couple with the sound field and radiate in the direction of the observer. The present study investigates a large backscattering contribution which manifests when the tilt angle of a finite cylinder is in the vicinity of the flat-surface coupling angle for launching leaky surface waves on the solid–liquid interface. This coupling angle is such that the incident wave vector's projection along an axis parallel to the interface matches the wave vector associated with the surface wave's propagation along the interface and can thus be expressed as  $\theta_l = \sin^{-1}(c/c_l)$ , where  $c$  is the speed of sound in the liquid and  $c_l$  is the phase velocity of the surface wave, which exceeds  $c$  for leaky waves.<sup>8,9</sup> Because the phase velocity is supersonic with respect to the surrounding fluid, the surface wave will leak radiation back into the fluid, and by reciprocity the angle for the radiation leakage is the coupling angle  $\theta_l$ . The backscattering enhancement observed for tilt angles in the vicinity of this coupling angle is thought to result from the launching of leaky surface rays along the meridian of the cylinder which lies in the plane defined by the incident wave vector and the cylinder's axis; upon reaching the truncation, such a ray would be partially reflected so that the leaky wave vector would be reversed and the leaky radiation would thus be in the backscattering direction. The process is illustrated in Fig. 1, where for simplicity

only a few rays of specific orientation are shown. The actual physical situation includes both incoming and outgoing radiation for all of the rays shown, and there are many more such rays along the entire length of the cylinder.

High-frequency backscattering enhancements from meridional rays on bluntly truncated cylindrical shells have been observed when the tilt angle is in the vicinity of the coupling angles for certain Lamb waves.<sup>10,11</sup> The present study describes experimental observations of similar backscattering enhancements by solid cylinders tilted near the coupling angle for Rayleigh waves (hereafter referred to as the Rayleigh angle). Unlike the experiments in Refs. 10 and 11, the present experiments provide a quantitative test of the method of analysis introduced in Ref. 12 to approximate the backscattering pressure amplitudes. It is found that the far-field backscattered pressure exceeds that of specular reflection from a sphere having the same radius as the cylinder, so that this scattering mechanism greatly improves the visibility of the truncation and may also be important in the initial detection of objects with sharp truncations. Meridional-ray backscattering enhancements may be evident in ultrasonic NDT as well as in sonar measurements.

It is worth mentioning that for tilt angles  $\gamma$  less than the coupling angle  $\theta_l$ , there will be off-meridional points at which the incident wave locally satisfies the coupling condition. These points lie on a coupling curve on the ensonified surface of the cylinder, and it is expected that leaky helical rays will generally be launched for tilts less than  $\theta_l$ . Such waves may reflect from the truncation and leak radiation in the backscattering direction,<sup>6,13</sup> however, because of increased attenuation due to a longer path along the cylinder, they are expected to produce smaller backscattering amplitudes than the mechanism of interest in the present investigation. Backscattering contributions from leaky helical rays are therefore not specifically addressed in this work, al-

<sup>a)</sup>Present address: Physics Department, Grand Valley State University, Allendale, MI 49401.

<sup>b)</sup>Electronic mail: marston@wsu.edu

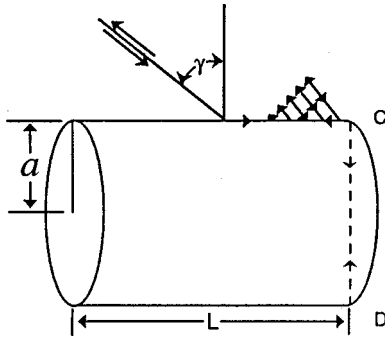


FIG. 1. Leaky Rayleigh rays are launched along the meridian defined by the incident acoustic wave vector and the axis of the cylinder when the tilt angle  $\gamma$  corresponds to the coupling angle  $\theta_l$ . The rays will be partially reflected at the truncation (point C) and the reflected rays will leak radiation in the backscattering direction. The ray transmitted onto the face of the cylinder at point C will subsequently be partially reflected at point D and transmitted back onto the meridian at point C; such rays may give rise to an additional time-delayed backscattering contribution which is generally much weaker.

though they are evident in the data scan displayed in Sec. II. For an ensemble of randomly oriented cylinders, the meridional-ray reflection considered here will be a significant mechanism for producing large backscattering.

Following the formulation discussed in Ref. 12, the leaky wave backscattered pressure is approximated by determining the contribution to the pressure at the surface of the cylinder due to leaky wave coupling, accounting for reflection of the leaky wave at the truncation, and describing the outgoing pressure due to the leaky wavefront in a suitable exit plane so that the Rayleigh–Sommerfeld propagation integral may be employed to calculate the pressure in the far field. The leaky wave contribution to the pressure at a point  $S$  on the cylinder's surface,  $p_l(S)$ , is determined by convolving the incident pressure  $p_i$  with an appropriate two-dimensional response function:

$$p_l(S) \approx \int_D p_i(S') [\kappa H_0^{(1)}(k_p s)] dA', \quad (1)$$

where  $dA'$  is the differential area of the contributing surface patch at the illuminated point  $S'$ ,  $D$  is a directionally weighted domain,<sup>12,13</sup>  $H_0^{(1)}$  is the zero-order Hankel function of the first kind, having an argument proportional to the leaky wave number  $k_p$  and the geodesic distance  $s$  between  $S$  and  $S'$ , and the factor  $\kappa$  can be written in terms of the leaky wave parameters. Note that Eq. (1) gives a representation of the leaky wave on the surface as a superposition of cylindrical waves emanating from surface points illuminated by the incident wave. The leaky wave number can be written as  $k_p = k_l + i\alpha$ , where  $k_l = k(c/c_l)$ , with  $k$  being the wave number in water, is associated with the propagation of the leaky wave and  $\alpha$  is associated with its attenuation due to leaky radiation. The factor  $\kappa$  can be written  $\kappa \approx -\alpha k_l \exp(i\varphi_{bl})$ , where  $\varphi_{bl}$  is the background phase.<sup>4,12,13</sup> At the high frequencies used in our experiments,  $\alpha$  generally is large enough that global resonances are unimportant, since  $\alpha L \gg 1$  (where  $L$  is the length of the cylinder). At the same time, however, the fluid loading is weak enough that  $\alpha \ll k_l$  and  $k_p \approx k_l$  for some of the simplifications used in evaluating Eq. (1).

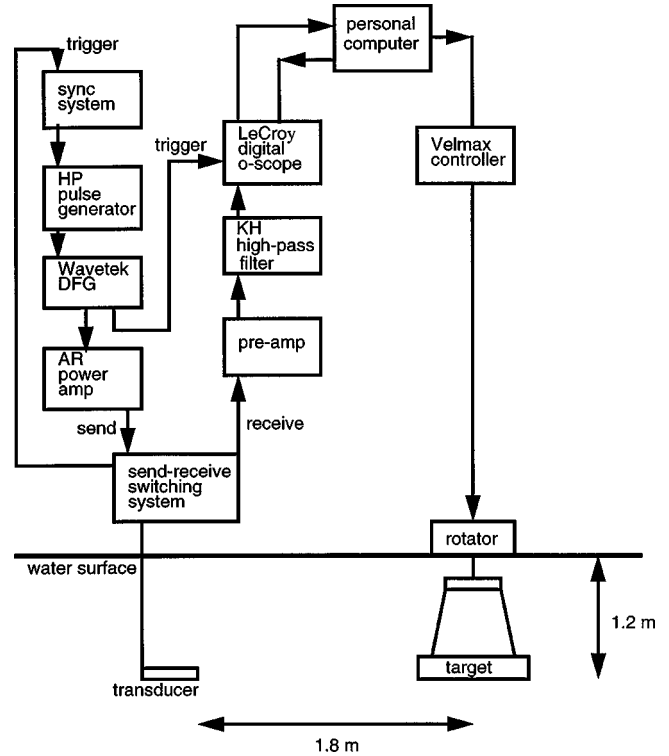


FIG. 2. Experimental setup for measuring backscattering at various tilt angles.

Once the leaky wave amplitude on the surface is determined from Eq. (1), a modified method of images can be used to account for the reflection of the leaky wave at the truncation. The curvature of the backscattered leaky wavefront is related to the spatial phase dependence in an exit plane oriented tangent to the cylinder along the meridian of interest. The far-field backscattering amplitude can then be calculated using the Rayleigh–Sommerfeld propagation integral.<sup>12</sup> The method of analysis is conceptually quite straightforward, but its implementation may involve the evaluation of complicated integrals. Various approximations which simplify the evaluation of the necessary integrals are discussed in Appendix A, which extends the results in Ref. 12 to the case where  $\gamma \neq \theta_l$ ; the reader is referred to Ref. 12 for other details of the calculation. When  $\gamma \neq \theta_l$ , the meridional leaky wave launched on the surface becomes dephased relative to the incident acoustic wave.

## I. EXPERIMENTAL DESIGN

Pulse–echo scattering experiments were conducted in a redwood tank of 12 ft diameter and 8 ft depth which held approximately 6500 gallons of water. A 2.54-cm diameter piston transducer (Panametrics model V302) was used as both source and receiver. The experimental setup and associated electronics are shown in Fig. 2. The transducer and target were centered in the tank and separated by a distance of 1.8 m. This separation distance approximated plane wave incidence on the cylinder while ensuring that reflections from the tank walls were temporally distinct from the signals backscattered by the cylinder. (A correction to plane wave incidence due to the finite nature of this separation distance

TABLE I. Material parameters for the experiments. The parameters for SS 304 and water are needed to determine the velocity of a leaky Rayleigh wave on the cylinder, and the parameters for 440c and water are needed to determine the magnitude of the specular contribution to the form function of the calibration sphere.

	Density, $\rho$ (in $\text{g/cm}^3$ )	$c_{\text{Longitudinal}}$ (in $\text{mm}/\mu\text{s}$ )	$c_{\text{shear}}$ (in $\text{mm}/\mu\text{s}$ )
Cylinder (SS 304)	7.57	5.675	3.141
Sphere (SS 440c)	7.455	6.110	3.354
Water	0.998	1.483	

is discussed in Appendix C.) Using a single transducer in a send–receive configuration facilitated the alignment of the experiment, and the time window of the experiment was chosen so as to isolate the electrical noise introduced by the switching mechanism from the signals of interest. In the send mode, the source was driven by multiple-cycle sinusoidal tone bursts generated by the Wavetek digital function generator and amplified by the Amplifier Research power amplifier. Frequencies of 620 kHz, 800 kHz, and 1.03 MHz were used, corresponding to the specified or experimentally determined optimum frequencies of the transducer, and the lengths of the tone bursts were chosen so as to ensure a build-up to steady-state conditions for the scattering mechanism of interest at a given frequency. The received signal was bandpass filtered and averaged over many pings by the digital LeCroy oscilloscope before being recorded.

The target was a solid cylinder of length 254 mm and radius 19.0 mm composed of stainless-steel 304. The radius was chosen to correspond to high  $ka$  for the frequency range of the experiments, and the length was chosen to satisfy the requirement of being much greater than the radius. The material properties of SS304 and those of the surrounding water are given in Table I and are needed to determine the Rayleigh velocity  $c_R$  and hence the Rayleigh angle  $\theta_R$ . The cylinder was oriented with its axis horizontal and was suspended from the rotary stage by two nylon fishing lines. (See Appendix C.) The nylon lines were attached to the cylinder by threading them through holes drilled through small set screws; the set screws were counter-sunk below the surface of the cylinder and positioned such that the suspension lines would be angled so as to contribute as little as possible to the backscattered signal. The rotary stage enabled the tilt of the cylinder with respect to the incident acoustic wave to be varied by  $0.025^\circ$ -steps through a full  $360^\circ$  range if desired. In practice, smaller scans at coarser steps were implemented as appropriate for observing global features or fine details of the scattered field, as discussed subsequently for two specific cases. The experiments were automated through a computer, which drove the rotary stage through the Velmax controller and recorded the averaged signals sent to it by the LeCroy. The high inertia of the cylinder required a fairly lengthy time delay (approximately 5 min) between turning the target and recording the data in order to minimize destructive interference effects from the varying spatial positions of a swinging target.

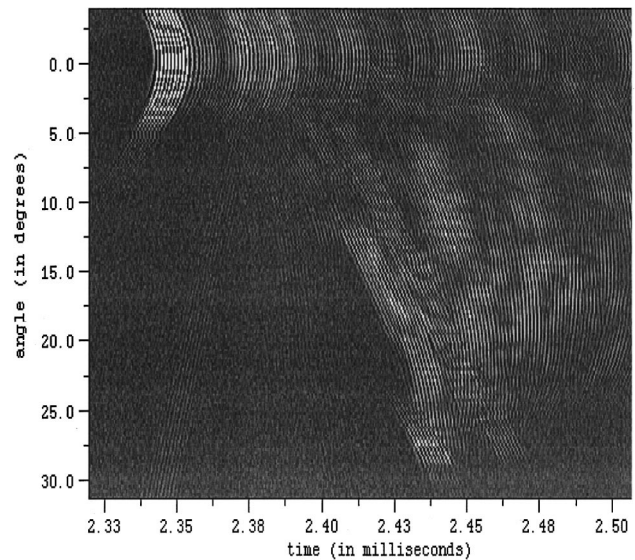


FIG. 3. Scan in tilt for eight-cycle tone bursts at 800 kHz. Individual time records for a  $35^\circ$  scan at  $0.2^\circ$ -increments have been cascaded in order to display the angular evolution of the backscattered pressure. Broadside incidence corresponds to  $0^\circ$  and the meridional feature of interest is evident near  $30^\circ$ .

## II. EXPERIMENTAL OBSERVATIONS

The initial experiments were performed at the transducer's specified optimum frequency of 800 kHz, corresponding to  $ka$  of 65. The parameters in Table I predict a flat-surface Rayleigh velocity<sup>9</sup> of  $2.904 \text{ mm}/\mu\text{s}$ , and therefore a Rayleigh angle of  $30.7^\circ$ . Hence, angular scans of  $35^\circ$  at steps of  $0.2^\circ$  were performed in order to observe the presence of the meridional-ray reflection and explore general features of the field for angles less than the Rayleigh angle. (Flat-surface values for  $k_l$  and  $\alpha$  are sufficient for the present analysis because the parameters associated with leaky wave propagation on curved surfaces are usually found to converge to the flat-surface values at the high frequencies used.<sup>12,14–16</sup>) In Fig. 3, individual time records for eight-cycle tone bursts have been cascaded to display the general dependence of the far-field backscattered pressure on the tilt angle of the cylinder. The zero angle is determined from the tilt corresponding to the longest arrival time for specular reflection near broadside incidence, because any tilt of the cylinder away from broadside will bring some portion of the cylinder closer to the source–receiver position. The subsequent signals at the zero angle correspond to the launching and repeated circumnavigation of circumferential leaky Rayleigh waves and other leaky waves,<sup>14,15</sup> and the curves arcing from the broadside region to the region around  $30^\circ$  are suggestive of the evolution of the circumferential Rayleigh waves into helical waves. (The identification of the helical waves, based on an analysis of the timing of the contributions,<sup>6</sup> is considered elsewhere.<sup>17</sup>) The feature of interest in this investigation is the prominent signal observed for incidence in the angular region near  $30^\circ$ , in general agreement with the predicted value of the Rayleigh angle. It is noteworthy that the arrival time of this signal (approximately  $88 \mu\text{s}$  after the specular



return) also supports its interpretation as a reflected leaky wave propagating at the Rayleigh velocity along the meridian.

The smaller time-delayed return observed in the angular region of  $30^\circ$  in Fig. 3 suggests the possibility of a similar reflection process for a ray which is partially transmitted across the face of the cylinder. When the meridional leaky ray reflects at the truncation indicated by point C in Fig. 1, a portion of it will be partially transmitted; such a ray could propagate across the face of the cylinder to be subsequently retroreflected at point D and thus result in a reversal of the wave vector in a manner similar to the meridional ray's reflection at point C. The amplitude of such a face-traversing ray would be significantly reduced due to extra radiation damping  $\exp(-4aa)$  as well as the multiplicative effects of two partial transmissions, and its time delay relative to the retroreflected meridional-ray return would be  $4a/c_R$ . Although the observed time delay of  $26 \mu\text{s}$  for the signal under discussion agrees with the theoretical prediction, the amplitude reduction factor is somewhat complicated to predict since it will depend not only on attenuation damping but also on possible focusing effects. The observed amplitude reduction factor was 0.20, but the experiments of this study are inconclusive for identifying the time-delayed return as a face-traversing ray, and further investigation of that signal is deferred. However, in the process of attempting to investigate this effect by using long tone bursts of various frequencies to drive a resonance of the cylinder's face, it was discovered that the transducer actually responded inadequately in the 800-kHz frequency range. (That is, the tone burst generated by the transducer in this frequency range failed to have an adequately square envelope for quantitative measurements.) Although the global scans at 800 kHz are adequate for identifying qualitative features of the backscattered pressure's dependence on tilt, subsequent experiments were conducted at the experimentally determined optimum frequencies of the transducer, 620 kHz and 1.03 MHz, corresponding to  $ka$  of 50 and 83, respectively.

To substantiate the identification of the prominent feature observed for incidence near  $30^\circ$  as a retroreflected leaky meridional ray, a comparison of its peak amplitude with the amplitude predicted by the approximate theoretical analysis of Ref. 12 is necessary. In addition, the angle corresponding to the peak as well as the dephasing away from the peak should corroborate theoretical expectations. In order to obtain better resolution near the peak, steps of  $0.1^\circ$  were used for  $8^\circ$ -scans approximately centered on the angle of maximum backscattering. Figures 4 and 5 are examples of the backscattered far-field pressure measured in two such experiments, for frequencies of 620 kHz and 1.03 MHz, respectively. These experiments were performed using 50-cycle tone bursts (motivated by the attempts to drive a resonance of the face of the cylinder as previously discussed), and the example time traces displayed in these figures correspond to the angle of the maximum backscattered pressure (i.e., the experimentally determined Rayleigh angle). The magnitude of the pressure at each angular step throughout the scan region was determined by averaging over three peak-to-peak cycles of an early portion of the signal as indicated by the

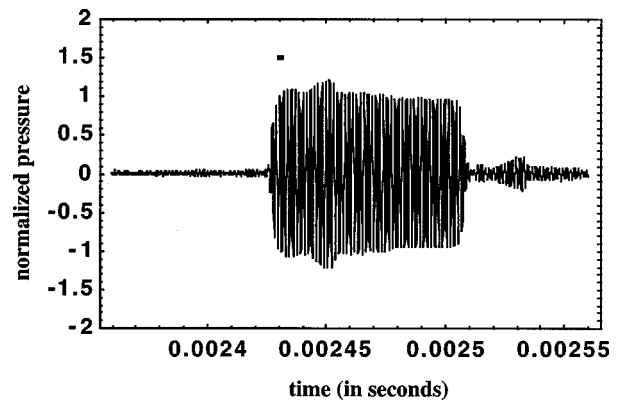


FIG. 4. Time trace of backscattered pressure for  $\gamma=28.2^\circ$  using 50-cycle tone bursts at 620 kHz. The data in this plot were normalized with respect to specular reflection from a calibration sphere, so that the vertical axis is dimensionless. The horizontal axis has units of seconds. The form function assigned to this angle was determined by averaging the pressure over three cycles after the signal reached steady state (as indicated by the horizontal mark in the figure), but before interference effects from other waves set in, and calibrated as discussed in Appendix B. The process was repeated for each  $0.1^\circ$ -step in an  $8^\circ$  scan centered around the predicted Rayleigh angle to generate the solid curve in Fig. 6.

horizontal marks in Figs. 4 and 5. This early interval was chosen in an attempt to reduce the complication of interference from helical ray contributions which arrive slightly later in time.<sup>6,17</sup>

### III. DISCUSSION

Figures 6 and 7 show the angular dependence of the backscattered pressure for tilts near the Rayleigh angle. The normalization of Appendix B was applied to the data displayed in Figs. 6 and 7, and comparison between theory and data are expressed in terms of the amplitude of the form function. The smooth curves are the normalized peak pressures measured in the experiments corresponding to Figs. 4 and 5. The structure evident in the data curves is thought to be the result of interference with other weaker waves, such as helical Rayleigh waves. This view is motivated by the observed presence of interfering waves evident in Figs. 3, 4, and 5 and is supported by the frequency dependence of the structure seen by comparing Figs. 6 and 7. The dashed curves in Figs. 6 and 7 are the theoretical approximate

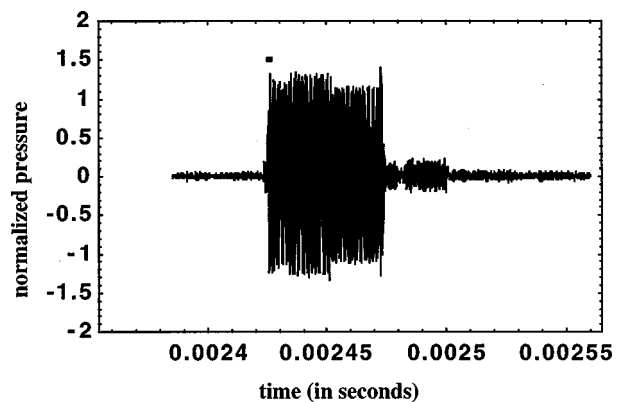


FIG. 5. Same as Fig. 4 except for  $\gamma=28.0^\circ$  at 1.03 MHz. The corresponding dephasing relationship is shown by the solid curve in Fig. 7.

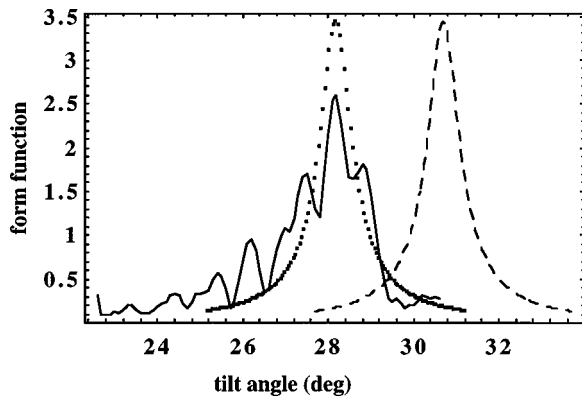


FIG. 6. Dephasing curves for 620 kHz. The solid curve is data from the experiment corresponding to Fig. 4, the dashed curve is a theoretical approximation given by Eq. (A2), and the dotted curve is a shifted version of the theoretical curve. The structure in the experimental curve is believed to arise from interference with other types of backscattered waves, and the justification for the angular shift of the theory curve is discussed in Appendix C.

dephasing curves, normalized by putting the far-field backscattered pressure for the tilted cylinder into the standard form of Eq. (B1). Marston's predictions for the approximate dephasing curves based on Eq. (1) are derived in Appendix A.

Taking the tilt angle as an adjustable parameter facilitates comparison of the experimental and predicted features of the dephasing curves; the dotted curves in Figs. 6 and 7 are shifted versions of the theoretical curves. The justification of the angular shift stems from the finite distance of the target from the source–receiver. As demonstrated in Appendix C, this finite distance results in the angle to the edge of the cylinder being slightly larger than the angle to the center of the cylinder. The angular shifts required to match the theory to the data are in the correct direction and well within the range of uncertainty. In addition, the experiments confirm the theoretical prediction that the peak signal due to the retroreflected Rayleigh wave can significantly exceed the specular reflection from a sphere of the same radius of the cylinder. (The latter would correspond to unity in the normalized units of Figs. 6 and 7.) Although the amplitudes corresponding to the peak signals are not in perfect agreement with theoretical predictions, the measured peak ampli-

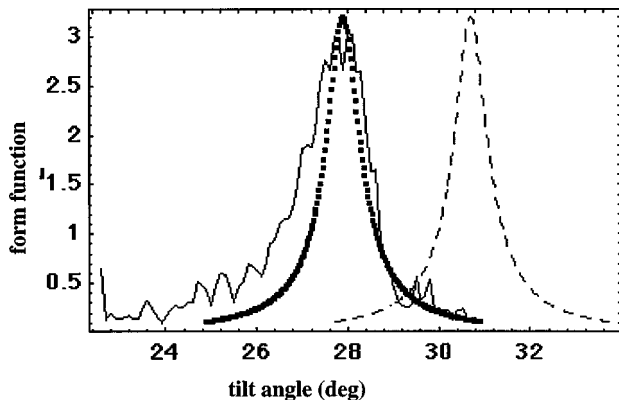


FIG. 7. Same as Fig. 6 except for 1.03 MHz, corresponding to the experiment of Fig. 5.

tudes are 25% and 5% below the predictions for the experiments corresponding to Figs. 6 and 7, respectively. Inspection of Fig. 7 indicates that the discrepancy for 1.03 MHz may actually be even lower than 5%, as the data exhibit interference suggestive of helical wave contributions. The greater discrepancy evident in Fig. 6 may arise at least in part from the lower rate of radiation damping at lower frequency:  $\alpha=0.0155 \text{ mm}^{-1}$  at 620 kHz compared to  $\alpha=0.0258 \text{ mm}^{-1}$  at 1.03 MHz. The lower radiation damping results in a greater attenuation length, which increases the distance required for far-field conditions to be met for the mechanism of interest. (See Appendix A.) It is also noteworthy that the widths of the theoretical and measured dephasing curves are seen to be in relatively good agreement. The discrepancies in the widths of the dephasing curves are difficult to quantify because the interference effects serve to obscure the absolute width of the experimental dephasing curves. However, both the peak amplitude and angular width comparisons are sufficient to support the interpretation of the large backscattering enhancement observed for a finite solid cylinder tilted near the Rayleigh angle as the meridional-ray retroreflection process depicted in Fig. 1, and to substantiate the usefulness of the convolution method described in Ref. 12 to approximate the amplitude of the backscattered pressure for tilts in the vicinity of the Rayleigh angle.

## ACKNOWLEDGMENTS

This research was supported by grants from the Office of Naval Research and Washington State University.

## APPENDIX A: MERIDIONAL BACKSCATTERING ENHANCEMENT FROM REFLECTION AT A BLUNT END WHEN CYLINDER TILT DIFFERS FROM THE RAYLEIGH ANGLE

The purpose of this appendix is to summarize the derivation of the theory plotted in Figs. 6 and 7. The derivation follows from an extension of Sec. VII of Ref. 12 to include the case when the tilt  $\gamma \neq \theta_l$ . For brevity, specific equations from that reference are cited using a prefix M and the same notation as explained there is used. Certain approximations in the description of the leaky Rayleigh wave are also introduced. Some of these approximations concern the curvature of the leaky wavefronts on the cylinder. An important test case confirmed in Ref. 12 (and in an analogous study for an aluminum cylinder in Ref. 18) is the approximation of the meridional-ray amplitude for a tilted infinite cylinder starting with Eq. (1). Those comparisons with partial wave series results show that for the purpose of approximating the meridional contribution, the curvature of the leaky wavefront on the cylinder for the case when the cylinder is tilted such that  $\gamma \neq \theta_l$  may be approximated using the value for the case  $\gamma = \theta_l$ . This approximation gives acceptable results because the decrease in the magnitude of the meridional contribution as  $\gamma$  is shifted away from  $\theta_l$  is dominated by the dephasing of the incident acoustic field with the meridional ray.<sup>12,18</sup> The same approximation is made here in the description of the curvature of the reflected leaky wavefront on the cylinder. An additional approximation is made in that it is assumed



that the phase of the amplitude reflection coefficient at the blunt end varies only weakly with Rayleigh wave angle of incidence for near-normal incidence at the truncation. If the phase variation were appreciable, it would be necessary to modify the curvature of the reflected wavefront from the result given in Eq. (M29).

We now turn to the specific extension of the far-field backscattering amplitude, Eq. (M34), to allow for  $\gamma \neq \theta_l$ , where in the present case  $\theta_l$  is the Rayleigh wave coupling angle  $\theta_R$ . First, consider the case where the amplitude reflection coefficient  $B$  at the truncation is taken to be unity. The outgoing leaky wave amplitude from Eq. (1) is given by Eq. (M12) with the integral  $I$  replaced by  $K$  from Eq. (M27). [From a transcription error noted in the erratum,<sup>12</sup> a factor of 2 should be inserted on the right side of Eq. (M12).] To allow for  $\gamma \neq \theta_l$ , the previous values of the arguments  $\nu$  and  $\beta$  in  $K(\nu, \beta, \tau, \eta)$  are replaced by  $\nu = \alpha + i(k \sin \gamma - k_l)$  and  $\beta \approx k_l a / (k \cos \gamma)$ , where  $\beta$  is defined by Eq. (M11) and the approximation (also used in Ref. 12) assumes that  $\alpha \ll k_l$  and therefore  $\alpha$  may be omitted in Eq. (M11). Consequently, Eq. (M30) for  $p_l$  is replaced by

$$p_l(\eta, y) \approx -2B p_C e^{i\varphi_{bl}} (\pi \mu \Delta_C)^{1/2} \times (\Delta_S)^{-1/2} e^{i\chi y^2} e^{i\eta k \sin \gamma} e^{\mu \Delta_C \Delta_S} \times \operatorname{erfc}[(\mu \Delta_S \Delta_C + \alpha \eta + i \eta (k \sin \gamma - k_l))^{1/2}] H(\eta), \quad (\text{A1})$$

where  $\Delta_C = \cos \theta_l / \cos \gamma$ ,  $\Delta_S = 1 + i[(\sin \gamma / \sin \theta_l - 1) / (\alpha / k_l)]$ ,  $\chi$  and the step function  $H$  are defined as below Eq. (M30),  $B$  is the leaky wave amplitude reflection coefficient at the truncation,  $\mu = \alpha a \tan \theta_l$ ,  $\eta$  and  $y$  are longitudinal and transverse coordinates as in Ref. 12, and  $p_C$  is the incident wave amplitude at the corner  $C$  in Fig. 1. The factors  $\Delta_C$  and  $\Delta_S$  become unity when  $\gamma = \theta_l$ , and it is easily verified that Eq. (M30) is recovered in the appropriate limit. The leaky wave amplitude in a reference plane tangent to the cylinder is then evaluated giving a modified form of Eq. (M31). The far-field amplitude is given by evaluating a Rayleigh–Sommerfeld propagation integral<sup>4</sup> over the aforementioned plane as in Eq. (M32). The resulting far-field backscattering amplitude is conveniently normalized like Eq. (M34), in the same way that the amplitude for a sphere of radius  $a$  is related to a form function. (See Appendix B.) The magnitude of the form function which generalizes Eq. (M34) is

$$|f_l| \approx 2|B| \left( \frac{k_l}{\alpha} \right)^{1/2} \left| \frac{1}{\sqrt{\Delta_C \Delta_S}} F_{e\Delta} \right|, \quad (\text{A2})$$

$$F_{e\Delta}(\mu, \gamma) = F_\rho \exp(\mu \Delta_C) \int_{\mu \Delta_C}^{\infty} \operatorname{erfc}[\sqrt{w \Delta_S}] dw, \quad (\text{A3})$$

where the complementary error function is

$$\operatorname{erfc}(\nu) = (2/\pi^{1/2}) \int_{\nu}^{\infty} e^{-u^2} du, \quad (\text{A4})$$

and  $F_\rho$  is given by Eq. (M35b). The integral in Eq. (A3) was evaluated numerically at  $0.025^\circ$  steps for a  $6^\circ$  angular scan centered on  $\gamma = \theta_R$  in order to generate the theoretical

dephasing curves in Figs. 6 and 7. When  $\gamma$  is close to  $\theta_R$ , the contribution predicted by Eq. (A2) is significantly larger than the edge diffraction amplitude given by GTD for a rigid cylinder.<sup>12</sup>

Since the density of the cylinder greatly exceeds that of the surrounding water and the frequency is sufficiently high that  $k_l a \gg 1$ , it is reasonable to approximate  $B$  for the present situation by using the corresponding value for the reflection of Rayleigh waves at the edge of an unloaded elastic quarter-space.<sup>19,20</sup> Since  $B$  is found to vary relatively slowly for near-normal incidence,<sup>20,21</sup> the localization of the Fresnel coupling patch around the meridian<sup>12</sup> justifies using  $B$  for normal incidence and treating it as a constant throughout the integration region. Though the effects of fluid loading are also expected to be small, it is worth noting that using the normal incidence free-surface value of  $B$  from Gautesen's analysis<sup>20</sup> probably yields slightly elevated results. The value of  $|B|$  used was 0.34, as estimated by interpolating Gautesen's results for Poisson's ratios of 0.25 and 0.33 with the Poisson ratio of 0.279 for the stainless steel used.

It is noteworthy that, according to Ref. 12, the meridional enhancement under consideration here is not highly localized in frequency. Since the ratios  $k/k_l$  and  $\alpha/k_l$  are approximated as independent of  $k$  for Rayleigh waves at high frequencies, the principal dependence on  $k$  is through the factor  $F_{e\Delta}$  in Eq. (A3), which reduces to the dashed curve in Fig. 8 of Ref. 12.

In the evaluation of the propagation integral leading to Eq. (A2), a far-field assumption is used as in Eq. (M32). This assumption requires that the distance from point  $C$  exceeds  $(1/2)kW^2$ , where  $W$  is the effective half-width of the target.<sup>4</sup> In the present application,  $W$  may be approximated as  $(2\alpha)^{-1} \cos \gamma$ , which gives a distance of 0.6 m at 1.03 MHz.

As in the case of the meridional coupling peak for an infinite cylinder,<sup>12,13</sup> the angular width of the peaks in Figs. 6 and 7 is such that the amplitude decreases when  $|\gamma - \theta_R|$  is sufficiently large that the dephasing distance along the meridian becomes as small as the attenuation length  $\alpha^{-1}$ . It is noteworthy that when Eqs. (A2) and (A3) are used to predict the meridional-ray backscattering contribution due to a leaky antisymmetric Lamb wave on a thick, tilted cylindrical shell with a blunt end,<sup>22</sup> the resulting magnitude and tilt dependence are in general agreement with an approximate partial wave series model where both the model and the ray theory use  $|B| = 1$ . While circumferential and helical rays have long been recognized as important contributors to radiation and scattering,<sup>1-6,11,13,23</sup> especially in the case of thin shells,<sup>24-26</sup> meridional contributions of the type considered here are relatively unexplored. Such contributions appear to be relatively significant at high frequencies for backscattering by solid cylinders and thick shells.<sup>10,11</sup> Diffraction by the ends of a tilted bluntly truncated rigid cylinder (reviewed at the end of Sec. 7 of Ref. 12) is predicted to be much weaker than the peak contribution given by Eq. (A2).

## APPENDIX B: CALIBRATION OF AMPLITUDE MEASUREMENTS

In practice, a direct measurement of  $p_C$  may be complicated, so experimental measurements of  $p_{\text{lscA}}$  are typically

normalized with respect to the measured specular reflection from a solid elastic sphere to eliminate the dependence on  $p_C$ . The generic equation for far-field scattering from a sphere of radius  $a_{\text{sph}}$  is conventionally expressed as<sup>4</sup>

$$p_{\text{sph}} = \frac{p_{\text{inc}} f_{\text{sph}} a_{\text{sph}}}{2R_S} \exp(ikR_S), \quad (\text{B1})$$

where  $R_S$  is the distance from the sphere and  $f_{\text{sph}}$  is a dimensionless form function describing the angular characteristics of the scattering pattern. The magnitude of the high-frequency specular contribution to the form function for backscattering from a sphere is approximated as<sup>4,7</sup>  $|\rho_E c_L - \rho c| / |\rho_E c_L + \rho c|$ , where  $\rho_E$  and  $c_L$  are the density and longitudinal wave speed of the elastic solid and  $\rho$  and  $c$  are the density and speed of sound in the surrounding water, respectively. Using the parameters in Table I, we find that  $|f_{\text{sph}}| \cong 0.937$  for the calibration sphere used in these experiments. If we assume that the far-field scattering from the cylinder is also expressible in the form of Eq. (B1), then the normalization for the case under investigation gives the form function of the cylinder in terms of measurable quantities of the cylinder and the sphere

$$|f_{\text{cyl}}| = \left| \frac{p_{\text{cyl}}}{p_{\text{sph}}} \right| \cdot \left( \frac{a_{\text{sph}}}{a_{\text{cyl}}} \right) \cdot \left( \frac{R_C}{R_S} \right)^2 \cdot |f_{\text{sph}}|, \quad (\text{B2})$$

where the notation  $p_{\text{tsca}}$  has been changed to  $p_{\text{cyl}}$  and appropriate subscripts have been used to distinguish between parameters associated with the cylinder and the calibration sphere.

The quadratic dependence on the positions of the scatterers arises because of spreading of both the incident and backscattered waves. Because the alignment of the experiment can be time consuming, it was desirable that the experiment not be disturbed by the calibration measurements. Therefore, the calibration sphere was placed somewhat in front of the cylinder (rather than removing the cylinder to place the sphere at the cylinder's location) so that the pressure incident on the sphere was slightly larger than the pressure incident on the cylinder. Assuming inverse distance spreading of the incident wave, the incident pressures are related by  $|p_{\text{inc(sph)}}| / |p_{\text{inc(cyl)}}| = R_C / R_S$ . Thus, one factor of  $R_C / R_S$  in Eq. (B2) is seen to arise from the spreading of the backscattered wave, and the other factor of  $R_C / R_S$  arises from the spreading of the incident wave.

In practice, time measurements were more accurate than direct measurements of the distances, so the ratio  $R_C / R_S$  was replaced by  $t_C / t_S$ . (The velocity of sound in the water was determined to be constant throughout the experiments, so that the ratio of distances is converted to a ratio of times.) The actual calibration applied to the data displayed in Figs. 6 and 7 was thus

$$|f_{\text{cyl}}| = \left| \frac{p_{\text{cyl}}}{p_{\text{sph}}} \right| \cdot \left( \frac{a_{\text{sph}}}{a_{\text{cyl}}} \right) \cdot \left( \frac{t_C}{t_S} \right)^2 \cdot |f_{\text{sph}}|, \quad (\text{B3})$$

where  $t_S$  is the time of the specular return from the sphere as determined from the time record of its backscattered pressure, and  $t_C$  is the time of the peak signal backscattered by corner C of the cylinder as read from individual data records at each step in the angular scan. (Although the measured

times are two-way transit times, the factor of 1/2 obviously cancels from the ratio.)

### APPENDIX C: ESTIMATION OF ANGULAR UNCERTAINTY DUE TO FINITE DISTANCE

This appendix analyzes the effect that the finite distance between the source–receiver and target has on the angle at which the incident wave strikes the end of the cylinder. Because of radiation damping, the largest contribution to backscattering will come from leaky rays with a detachment point very near the launch point, i.e., those launched very near the reflection point at the end of the cylinder. Therefore, the peak backscattered pressure will occur at the angle associated with incidence on the cylinder's end rather than at its center. The situation is illustrated in Fig. C1, where it can be seen that the angle of incidence on the end of the cylinder is larger than the angle of incidence at the cylinder's center by  $\delta$ . The cylinder was suspended as shown in Fig. C2 and was rotated about its center of mass. Since the experimental angle  $\gamma$  is measured at the center of mass of the cylinder, the peak backscattered signal is expected to occur for  $\gamma$  smaller than the Rayleigh angle by  $\delta$ . The present analysis approximates the source–receiver as a point, so that the results give only an upper bound on  $\delta$  rather than an exact value.

Analysis of the geometry allows  $\delta$  to be determined in terms of the distance to the center of the cylinder  $D$  and the length of the cylinder  $L$ . It is found that  $\tan \delta = A' / d = A / D$ .

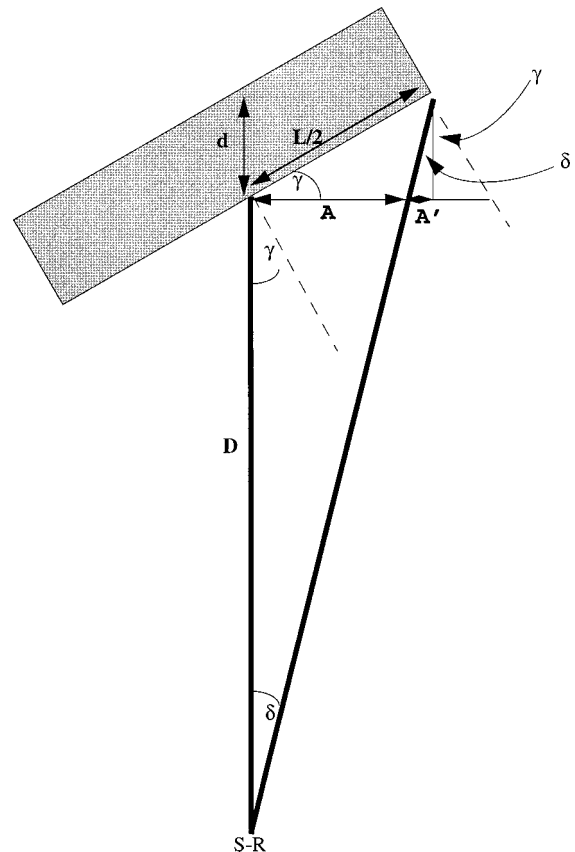


FIG. C1. Diagram showing how finite source distance affects the local angle of incidence. The transducer is modeled as point S-R in this sketch. From geometry, it can be seen that the angle of incidence on the end of the cylinder is  $\delta + \gamma$ .

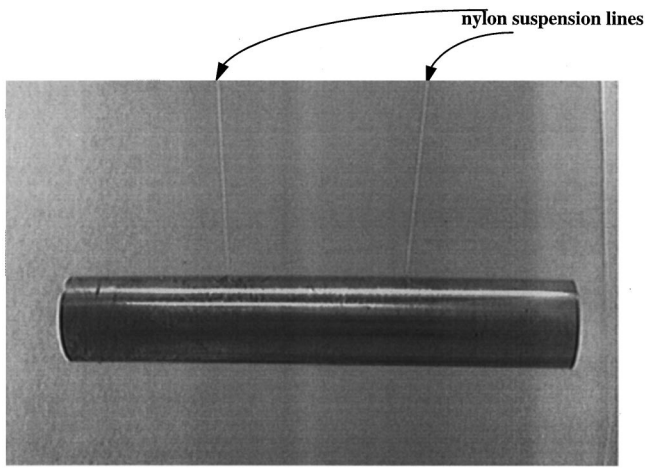


FIG. C2. Photograph of cylindrical target. The target was a solid stainless-steel cylinder of radius 19.0 mm and length 254 mm. It was suspended by two nylon lines which were attached to the cylinder via counter-sunk set screws and angled so that the mounting system would give negligible contribution to the backscattered pressure.

Reference to Fig. C2 shows that  $d = (L/2)\sin \gamma$  and  $A + A' = (L/2)\cos \gamma$ . Solving this system of equations yields  $\delta = \arctan[(\tan \gamma)^{-1}(1 + (2D)/(L \sin \gamma))^{-1}]$ . Substituting in the measured values of  $D = 174$  cm,  $L = 25.4$  cm, and  $\gamma_{620 \text{ kHz}} = 28.2^\circ$ ,  $\gamma_{1.03 \text{ MHz}} = 28.0^\circ$  from the experiments gives a predicted shift of  $-3.5^\circ$  for both experiments. The measured shifts for the experiments shown in Figs. 6 and 7 are  $-2.5^\circ$  and  $-2.7^\circ$ , respectively, well within this upper bound. The angular shift of the theoretical dephasing curves in Figs. 6 and 7 is thus justified on the basis of the finite distance between the source–receiver and target.

- <sup>1</sup>L. Flax, V. K. Varadan, and V. V. Varadan, "Scattering of an obliquely incident acoustic wave by an infinite cylinder," *J. Acoust. Soc. Am.* **68**, 1832–1835 (1980).
- <sup>2</sup>F. Hanovar and A. N. Sinclair, "Acoustic wave scattering from transversely isotropic cylinders," *J. Acoust. Soc. Am.* **100**, 57–63 (1996).
- <sup>3</sup>N. D. Veksler, "Scattering of a plane acoustic wave obliquely incident on a solid elastic cylinder," *Acustica* **71**, 111–120 (1990).
- <sup>4</sup>P. L. Marston, "Geometrical and catastrophe optics methods in scattering," in *Physical Acoustics*, edited by R. N. Thurston and A. D. Pierce (Academic, Boston, 1992), Vol. 21, pp. 1–234.
- <sup>5</sup>R. H. Hackman, "Acoustic scattering from elastic solids," in *Physical Acoustics*, edited by R. N. Thurston and A. D. Pierce (Academic, Boston, 1992), Vol. 22, pp. 1–194.
- <sup>6</sup>X. L. Bao, "Echoes and helical surface waves on a finite elastic cylinder excited by sound pulses in water," *J. Acoust. Soc. Am.* **94**, 1461–1466 (1993).
- <sup>7</sup>K. L. Williams and P. L. Marston, "Backscattering from an elastic sphere: Sommerfeld–Watson transformation and experimental confirmation," *J. Acoust. Soc. Am.* **78**, 1093–1102 (1985).

- <sup>8</sup>L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980).
- <sup>9</sup>H. L. Bertoni and T. Tamir, "Unified theory of Rayleigh-angle phenomena for acoustic beams at liquid–solid interfaces," *Appl. Phys.* **2**, 157–172 (1973).
- <sup>10</sup>G. Kaduchak, C. M. Wassmuth, and C. M. Loeffler, "Elastic wave contributions in high resolution acoustic images of fluid-filled cylindrical shells in water," *J. Acoust. Soc. Am.* **100**, 64–71 (1996).
- <sup>11</sup>S. F. Morse, P. L. Marston, and G. Kaduchak, "High frequency backscattering enhancements by thick cylindrical shells in water at oblique incidence: Experiments, interpretation, and calculations," *J. Acoust. Soc. Am.* **103**, 785–794 (1998).
- <sup>12</sup>P. L. Marston, "Approximate meridional leaky ray amplitudes for tilted cylinders: End-backscattering enhancements and comparisons with exact theory for infinite solid cylinders," *J. Acoust. Soc. Am.* **102**, 358–369 (1997); *erratum* **103**, 2236 (1998).
- <sup>13</sup>P. L. Marston, "Spatial approximation of leaky wave surface amplitudes for three-dimensional high frequency scattering: Fresnel patches and application to edge-excited and regular helical waves on cylinders," *J. Acoust. Soc. Am.* **102**, 1628–1638 (1997).
- <sup>14</sup>P. L. Marston, "GTD for backscattering from elastic spheres and cylinders and the coupling of surface elastic waves with the acoustic field," *J. Acoust. Soc. Am.* **83**, 25–37 (1988).
- <sup>15</sup>G. V. Frisk, J. W. Dickey, and H. Uberall, "Surface wave modes on elastic cylinders," *J. Acoust. Soc. Am.* **58**, 996–1008 (1975); J. W. Dickey, G. V. Frisk, and H. Uberall, "Whispering gallery modes on elastic cylinders," *ibid.* **59**, 1339–1346 (1976).
- <sup>16</sup>R. N. Thurston, "Elastic waves in rods and clad rods," *J. Acoust. Soc. Am.* **64**, 1–37 (1978).
- <sup>17</sup>K. Gipson, "Leaky Rayleigh wave ultrasonic backscattering enhancements: Experimental tests of theory for tilted solid cylinders and cubes," Ph.D. thesis, Washington State University, 1998.
- <sup>18</sup>P. L. Marston, "Approximations for leaky wave amplitudes in acoustic imaging: Applications to high frequency sonar," in *Acoustical Imaging*, edited by S. Lees and L. A. Ferrari (Plenum, New York, 1997), Vol. 23, pp. 369–374.
- <sup>19</sup>A. K. Gautesen, "Scattering of a Rayleigh wave by an elastic quarter-space," *J. Appl. Mech.* **52**, 664–668 (1985).
- <sup>20</sup>A. K. Gautesen, "Scattering of an obliquely incident Rayleigh wave in an elastic quarter-space," *Wave Motion* **8**, 27–41 (1986).
- <sup>21</sup>Z. L. Li, J. D. Achenbach, I. Komsky, and Y. C. Lee, "Reflection and transmission of obliquely incident surface waves by the edge of a quarter space: Theory and experiment," *J. Appl. Mech.* **59**, 349–355 (1992).
- <sup>22</sup>P. L. Marston and S. F. Morse, "Angle dependence of the meridional leaky-ray backscattering enhancement from the end of a tilted finite cylinder: Convolution analysis and a numerical test for shells," *J. Acoust. Soc. Am.* **102**, 3073(A) (1997).
- <sup>23</sup>A. Nagl, H. Uberall, P. P. Delsanto, J. D. Alemar, and E. Rosario, "Refraction effects in the generation of helical surface waves on a cylindrical obstacle," *Wave Motion* **5**, 235–246 (1983).
- <sup>24</sup>P. W. Smith, Jr., "Phase velocities and displacement characteristics of free waves in a thin cylindrical shell," *J. Acoust. Soc. Am.* **27**, 1065–1072 (1955).
- <sup>25</sup>L. B. Felsen, J. M. Ho, and T. Lu, "Three-dimensional Green's function for fluid-loaded thin elastic cylindrical shell: Alternative representations and ray acoustic forms," *J. Acoust. Soc. Am.* **87**, 554–569 (1990).
- <sup>26</sup>A. N. Norris and D. A. Rebinsky, "Acoustic coupling to membrane waves on elastic shells," *J. Acoust. Soc. Am.* **95**, 1809–1829 (1994).



# An acoustic model of a laminar sand bed

Nicholas P. Chotiros and Dennis J. Yelton

*Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029*

Morris Stern

*Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas 78712*

(Received 12 March 1998; revised 6 May 1999; accepted 24 May 1999)

It was postulated that a laminar sand bed may be modeled as an ensemble of randomly layered Biot media. The thickness of each layer was approximately half a grain diameter. The porosity variations in the vertical direction were matched to the mean and standard deviation of that of a structure of packed spherical grains. The effect of large-scale lateral variations in porosity was simulated by performing a coherent ensemble average of the acoustic output from several realizations of the randomly layered medium. The medium parameters were chosen to represent water-saturated sand. Specifically, the sand bed was modeled as bounded by a homogeneous water halfspace above, and a homogeneous poroelastic halfspace of equal average porosity below. Reflected and transmitted signals were computed. Coherent and random components of the reflected signal were calculated. The coherent parts were directly related to the reflected and the transmitted waves. Results showed significant differences between the modeled sand bed and an equivalent uniform Biot medium. In the modeled sand bed, the fast wave attenuation was found to be anisotropic, and a propagating slow wave was excited at most incident angles, except at normal incidence, which may explain the apparent failure to detect the slow wave of certain experiments. © 1999 Acoustical Society of America. [S0001-4966(99)01110-8]

PACS numbers: 43.20.Gp, 43.30.Ma [DLB]

## INTRODUCTION

The goal of this study was to investigate the extent to which sediment granularity and bedding can influence sound propagation in ocean sediments, particularly water-saturated sand. Typically, a couple of devices are employed to match theory to experimental measurements of attenuation and scattering—the attenuation is accounted for in terms of a complex bulk modulus, and the scattering is modeled as a random field of point scatterers in which the scattering strength is adjusted to match the measured data. Our hypothesis is that both attenuation and scattering might be explained in terms of acoustic interaction with structures associated with granularity and bedding. The viscoelastic theory of acoustic propagation in a solid is not a suitable starting point since it does not possess a mechanism that can account for the interaction between pore fluid and solid particles. Our starting point is Biot's theory of acoustic propagation in a poroelastic medium. It contains the basic mechanisms of acoustic interaction between the solid matrix and pore liquid as far as forward propagation is concerned, but it does not have any mechanism to account for scattering and any associated losses. In this study, Biot's theory is extended to include the effects of granularity and bedding, using a direct, numerical simulation approach.

## I. BACKGROUND

Current models of acoustic bottom backscatter from sandy sediments are based on composite roughness and volume scattering effects from sediments modeled as fluids.<sup>1,2</sup> However, recent bottom penetration experiments by

Chotiros<sup>3,4</sup> have shown that Biot's theory<sup>5,6</sup> of sound propagation through a fluid-filled porous solid matrix better explains the refraction of acoustic energy into the sediment. Biot's theory treats acoustic propagation through a poroelastic medium as a coupled wave motion within the solid and the pore fluid. The model may be visualized as a solid with tubular pores in the direction of particle motion, as illustrated in Fig. 1. The propagating wave, in the direction of the pore tubes, can be decomposed into three components—fast and slow compression waves, and a shear wave. The equations of motion, in a form used by Stern, Bedford, and Millwater<sup>7</sup> (correcting for typographical errors), are

$$\mu \nabla^2 \mathbf{u} + (H - \mu) \nabla (\nabla \cdot \mathbf{u}) - C \nabla (\nabla \cdot \mathbf{w}) = \rho \ddot{\mathbf{u}} - \rho_f \ddot{\mathbf{w}}, \quad (1)$$

$$C \nabla (\nabla \cdot \mathbf{u}) - M \nabla (\nabla \cdot \mathbf{w}) = \rho_f \ddot{\mathbf{u}} - \frac{c \rho_f}{\beta} \ddot{\mathbf{w}} - \frac{F^* \eta}{\kappa} \dot{\mathbf{w}}, \quad (2)$$

where  $\mathbf{u}$  is the displacement vector of the solid frame,  $\mathbf{w}$  is the negative porosity times the displacement vector of the pore fluid relative to the solid frame,  $\beta$  is the porosity of the solid frame,  $\rho_f$  is the mass density of the pore fluid,  $\rho$  is the mass density of the saturated sediment,  $\mu$  is the shear modulus of the solid frame,  $c$  is the virtual mass coefficient of fluid motion,  $\eta$  is the viscosity of the fluid, and  $\kappa$  is the permeability of the solid frame.  $C$ ,  $H$ , and  $M$  are constitutive coefficients depending on  $\beta$ ,  $\mu$ , and the bulk moduli of the pore fluid, grain material, and the saturated sediment.  $F$  is a dynamic correction term describing the frictional force due to the relative motion of the solid and fluid, as defined by Biot.<sup>6</sup> The relationships between the parameters of Eqs. (1) and (2) are given in Appendix A. In reality the pores are significantly



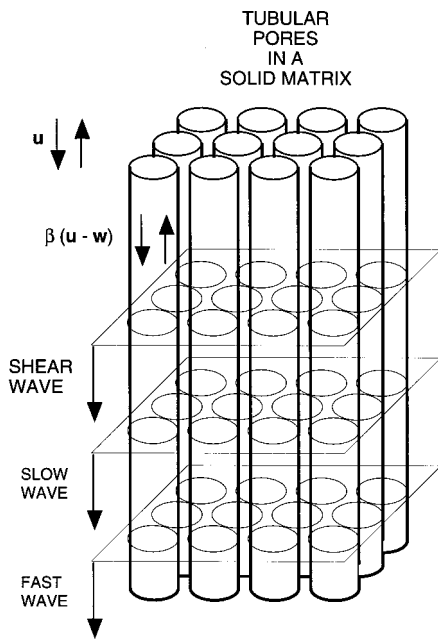


FIG. 1. The Biot medium: Tubular pores in a solid matrix.

more tortuous and twisted than the ideal, colinear tubular structures upon which the model is based. These deviations from the ideal are accounted for in the permeability and virtual mass terms,  $\kappa$  and  $c$ , respectively, as drag and inertia. These terms, in combination with the log decrement terms, allow the model to represent the macroscopic forward propagation of acoustic waves in a wide range of porous media. However, the theory has no mechanism for scattering because the pores are idealized as parallel smooth walled tubes, perfectly aligned with the direction of wave propagation.

When  $\mathbf{u}$  and  $\mathbf{w}$  are written in terms of vector and scalar potentials, the above two equations separate into four coupled partial differential equations. For homogeneous media, the vector and scalar potentials have plane wave solutions. At a fluid/poroelastic interface, there are four boundary conditions imposed on these solutions—continuity of fluid pressure, shear traction, normal traction, and normal fluid displacement. At a poroelastic/poroelastic interface there are two additional boundary conditions—continuity of tangential and normal solid displacements. The mathematical formulation of the differential equations and boundary conditions governing the vector and scalar potentials is illustrated by Stern, Bedford, and Millwater,<sup>7</sup> and outlined in Appendix A.

## II. MODEL

### A. Granular medium

The term “poroelastic” refers to a medium that follows Biot’s theory of acoustic propagation as described by Eqs. (1) and (2), in which the pores are initially modeled as parallel tubes as described in Fig. 1. However, since the underlying tubular structure is colinear with the direction of wave propagation, it is unable to model the effects of pore size variations that are undoubtedly present in granular sediments.

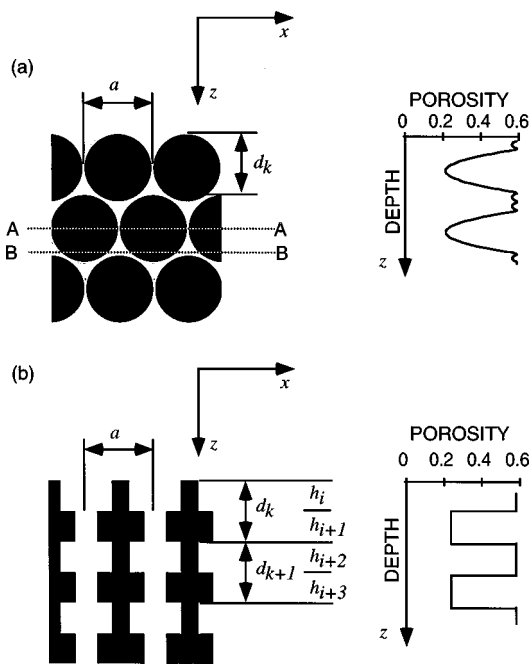


FIG. 2. Local porosity variations (a) in tetrahedral packing of spheres, and (b) its approximation in terms of a layered Biot medium.

In the context of this study, the term “granular” refers to a medium that is more complicated than the poroelastic medium as defined in Biot’s theory. In the granular medium, the local pore size may vary as a function of position, leading to local variations in porosity, permeability, and virtual mass constant. These variations produce local changes in impedance and sound speed which interact with propagating waves. This may be illustrated in the case of a lattice structure of spherical grains, a cross section of which is shown in Fig. 2(a). It is apparent that the local pore area in any vertical cross section is dependent on the vertical distance  $z$ . In this example, the local porosity in the plane AA is evidently less than that in the plane BB. The porosity varies periodically as a function of  $z$  as indicated in the graph. The period  $d_k$  is of the order of the grain diameter  $a$ .

The underlying hypothesis in this study is that local porosity variations play a significant role in both attenuation and scattering of acoustic waves, and therefore, there is a need to model them. Two approaches are possible. On the one hand, one can start at the microscopic level, and construct a very detailed finite-element model and work up toward the total acoustic response. This approach promises to be computationally intensive even if the physics of grain-grain and grain-fluid interactions were completely understood, which they are not. On the other hand, one can start at the macroscopic level with Biot’s equation, and extend it to include local variations, in steps of increasing complexity. The latter is the adopted approach.

“Bedding” is a term that is used to describe the structure of a sediment. It is the product of different combinations of grain size, shape, orientation, and packing.<sup>8</sup> Let us assume that the local structure of a sand bed may be divided into distinct layers, each layer being approximately one grain diameter thick, and termed a “monolayer.” The regular lattice structure in Fig. 2(a) certainly fits this description, where

“ $k$ ” is the number of a monolayer. Laminated bedding, characterized by very fine stratification and moderately common in occurrence, may be approximately described as a series of monolayers. Cross bedding, characterized by regions of laminar bedding of varying orientations, may be similarly described but on a local scale. Other types of bedding, such as massive bedding, are unlikely to fit this description and are excluded from this study.

Within a monolayer, the pore area changes through one complete cycle, as shown in Fig. 2(a). Let us approximate the pore area variation as a periodic two-step cycle, as illustrated in Fig. 2(b). Within each step the medium is modeled as a uniform Biot medium. The boundary between adjacent steps must satisfy all the boundary conditions for Biot media. In principle, the period, mean, and standard deviation of the variations may be matched to that of any laminar bedding. The propagation of sound through such a medium may be computed using an adaptation of Biot’s equations for stratified media.

It is expected that the sound field in a sandy sediment may be considered in terms of a deterministic or coherent component, and a random or incoherent component resulting from variations in the acoustic properties as a function of position. In the case of laminar or cross bedding, the randomness is expected to be governed by a spatial correlation function that is wide in the horizontal dimensions, but as narrow as a monolayer in the vertical dimension. Therefore, on a local scale, the sound field may be modeled as a randomly stratified Biot medium.

## B. Sand

The model has two critical components, the probabilistic definition of the layered structure and the solution of the wave propagation problem through numerous layers. Specifically, the model is fashioned after the sand samples used in Nolle’s measurements.<sup>9</sup>

In any structural arrangement of spherical grains, the porosity of the structure is simply the ratio of the interstitial volume to the total volume. The volume of each sphere is proportional to the cube of its diameter. On a per-sphere basis, the volume occupied by the structure is proportional to the cube of the distance between spheres. For any given structure, the distance between spheres is proportional to the separation between monolayers. Therefore, the average porosity,  $\beta_{\text{avg}}$ , may be expressed in terms of the cube of the diameter,  $a$ , divided by the layer separation,  $d_k$ , and a constant  $c_\beta$  that is a function of the structure,

$$\beta_{\text{avg}} = 1 - c_\beta (a/d_k)^3. \quad (3)$$

While grains are generally described as randomly packed, the arrangement may not be entirely structureless. Finney and Wallace,<sup>10</sup> in a computer modeling study of noncrystalline packed structures, state that “the structure can be considered as an assembly of distorted octahedra and tetrahedra,” and the resulting interstitial sphere statistics would indicate that, in a relaxed structure, the tetrahedra tend to be more dominant. In an earlier experimental study, using spheres of various materials, Westman and Hugill<sup>11</sup> concluded that “it was noticed, particularly with large spheres,

that a vertically cubic and horizontally hexagonal packing was the prevailing tendency,” which suggests tetrahedral packing. It would seem that a tetrahedral model has some merit, and, without loss of generality, it will be used as an illustrative example.

In an ideal tetrahedral packing, each sphere touches 12 adjacent spheres. The centers of any group of four adjacent spheres form corners of the basic tetrahedron. From simple trigonometry, the height of the tetrahedron can be shown to be equal to  $(a/3)\sqrt{6}$ . In a tetrahedral structure, the height of the tetrahedron is also equal to the monolayer separation  $d_k$ . It is also easily shown that the volume of solid per-unit area associated with each monolayer is  $(a\pi/9)\sqrt{6}$ . This, divided by the layer separation, gives the solid volume per unit volume,  $(\pi/6)\sqrt{2}$ , which is identical to one minus the porosity,  $(1 - \beta_{\text{avg}})$ . The resulting value of porosity is then  $1 - (\pi/6)\sqrt{2}$ , which is approximately 0.26. Substituting into Eq. (3) for  $\beta_{\text{avg}}$  and  $d_k$ , the value of  $c_\beta$  is found to be approximately 0.40.

In granular materials, such as sand, the porosity is higher than that of the ideal tetrahedral structure. It is postulated that the separation  $d_k$  is increased. In practice, this may occur for a number of reasons. The grains may be imperfect, with protruding jagged edges that hold them apart at distances that are, on average, greater than the mean, volumetric grain diameter, i.e., the diameter of a sphere of equal volume. In addition, one or more forces, including electrostatic and capillary forces, may also be forcing the grains apart. The result is that the separation between the centers of adjacent grains may be greater than one grain diameter.

For the purposes of modeling water-saturated sand, the average separation  $d_k$  was adjusted to match the measured value of porosity. For example, to match a measured porosity of 0.36, using Eq. (3), the monolayer separation  $d_k$  would have to be  $0.86a$ . Modeling each monolayer as a pair of uniform Biot layers of equal thickness, the thickness of each layer would be exactly one-half of  $d_k$ , i.e.,  $0.43a$ . In general, the grain diameter  $a$  is not a constant, but a random variable, and the distribution of the log of the grain diameter is usually approximated by a normal distribution. Therefore, the thickness  $h_i$  is also assumed to be similarly distributed.

In this model, the porosity  $\beta_i$  of the Biot layers alternates between two nominal values, as illustrated in Fig. 2(b). The mean  $\beta_{\text{avg}}$  and standard deviation  $\sigma_\beta$  of local porosity variations as a function of depth may be adjusted to fit any set of reasonable values. For the tetrahedral structure, with  $d_k$  adjusted to give an average porosity of 0.36, the standard deviation was computed to be 0.15. This was computed numerically, by formulating the void area in a unit horizontal plane as a function of depth, as illustrated in Fig. 2(a), and calculating the standard deviation of the resulting function.

Going from a perfect tetrahedral structure to a more random, close-packed structure, a degree of randomness must be introduced into the porosity. Measurements by Finney and Wallace<sup>10</sup> indicate that the pore volume distribution may be approximated by a log-normal distribution with a standard deviation of 0.6 Nepers. Thus the porosity of the Biot layers is modeled as

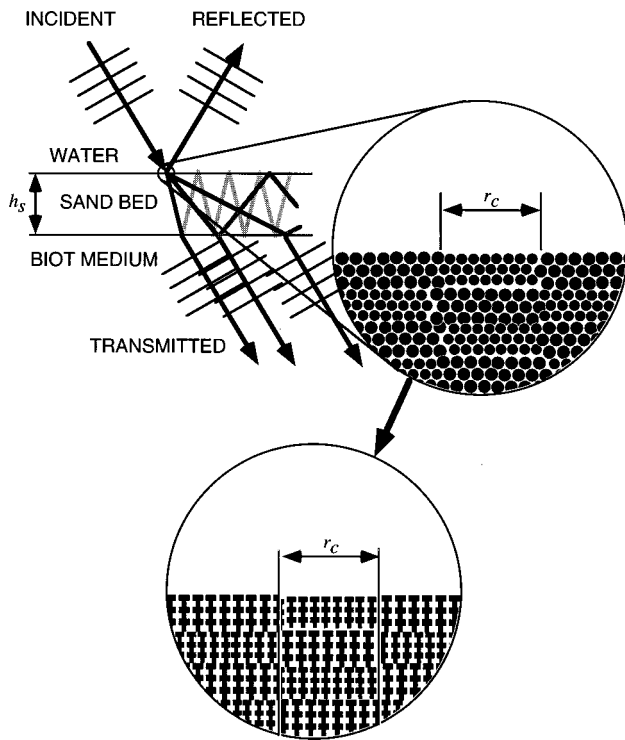


FIG. 3. Modeling of reflection and propagation through a sand bed.

$$\beta_i = (\beta_{\text{avg}} + (-1)^i \sigma_\beta) b, \quad (4)$$

where the log of  $b$  follows a normal distribution with a mean value of zero, and a standard deviation of 0.6 Nepers.

### C. Reflection and propagation

In modeling the propagation of a sound wave incident on a sand bed, and the subsequent reflection and transmission, consider a sand bed sandwiched between an upper half-space of water and lower halfspace of a uniform Biot medium, as illustrated in Fig. 3. Both the sand bed and the lower halfspace are saturated with water. The lower halfspace is a uniform Biot medium in which the parameter values are set equal to the average parameter values in the sand bed. The incident sound wave from the water would be partially reflected at the water–sand bed interface, and transmitted into the sand bed as fast, slow, and shear waves. At the lower interface, there will be transmission into the lower halfspace, and perhaps further reflection back into the sand bed. Of particular interest is the reflected wave at the water–sand bed interface and the transmitted waves into the lower halfspace. Ideally, if the granularity of the sand bed has a negligible effect on the propagating waves, then, the sand bed would be indistinguishable from the Biot medium in the lower halfspace, and the water–sand bed interface would be the only interface of any significance.

Before proceeding further, it is necessary to address the problem of lateral structure. There are many scales of lateral structure. The small-scale structure, on the order of a grain diameter, may have some similarities to the vertical structure, and may exert a similar effect on the lateral component

of sound propagation. Unfortunately, there is no efficient way to model it without resorting to finite-element methods, therefore it will be ignored for the present.

The large-scale structures may be associated with the sand bed deposition process. A laminated bedding comes closest to the model that is being developed in this study. It usually occurs in low current areas in which suspended particles settle onto the sediment bed in an orderly fashion. The nature of the sediment transport and precipitation process in coastal waters is such that particles of similar size tend to settle out at the same level. This naturally leads to inhomogeneities in particle size and porosity that are lenslike, i.e., pancake shaped, with large lateral extent but very short vertical extent. The lateral variations would likely mirror the large-scale, slow moving eddies within the currents that transport and deposit the sediments, leading to large-scale lateral correlation lengths that are very long. The correlation function of such a medium would have a very short correlation length in the vertical direction perhaps as small as a grain diameter, and relatively long correlation lengths in the horizontal directions.

To account for lateral variations of the sand bed, a locally stratified medium is assumed with a lateral correlation length  $r_c$ , as illustrated in Fig. 3. The granular structure is assumed to be horizontally stratified within the correlation length. On a local scale, the reflection coefficient  $R$  and transmission coefficients  $T_f$ ,  $T_s$ , and  $T_t$  of the fast, slow, and shear waves, respectively, are predicted by the solution of Eqs. (1) and (2) for a stratified medium, which will be described in the following section.

Of particular interest is the macroscopic sound field, both in the water and in the substrate. If the surface of the sand were rough, the roughness would need to be taken into consideration, but in this study, for simplicity, the surface roughness of the sand bed will be assumed to be negligible. Since the lateral correlation length is large, it is permissible to apply the Kirchhoff approximation, i.e., the local reflection and transmission coefficients are assumed to be equal to that of an infinite medium with the same horizontal stratification. Finally, if only thin beds are considered, then it is permissible to approximate the macroscopic, coherent component of the reflection and transmission sound fields simply as the ensemble average of the local sound fields. Thus the coherent or average reflection coefficient  $R_a$  for plane waves is simply equal to the average of all the local reflection coefficients. Similarly, for the coherent transmission coefficients  $T_{fa}$ ,  $T_{sa}$ , and  $T_{ta}$ , for shallow depths of penetration,

$$R_a = \langle R \rangle, \quad (5)$$

$$T_{fa} = \langle T_f \rangle, \quad (6)$$

$$T_{sa} = \langle T_s \rangle, \quad (7)$$

$$T_{ta} = \langle T_t \rangle. \quad (8)$$

Therefore, the coherent reflection and transmission coefficients may be numerically computed as the simple average of several realizations of a randomly stratified Biot medium. The result is a randomized, layered Biot medium which, to some extent, simulates the acoustic behavior of a laminar

sand bed for the purpose of modeling the propagation and attenuation of acoustic waves.

### D. Implementation

To implement the model, a numerical process for computing the reflection coefficient of a finely layered medium was developed, based on the interface boundary equations as defined by Stern, Bedford, and Millwater.<sup>7</sup> At a fluid-to-Biot medium interface, there are four boundary conditions that must be satisfied. At each interface between adjacent Biot layers there are six boundary conditions. Therefore, to model an  $n$ -layer medium would require solving a  $(6n+4)$  by  $(6n+4)$  banded matrix equation. Using this approach, it was quickly found that problems with more than a few hundred layers would become impractical due to the enormous computer time required. A different method was found that greatly reduced computation resource requirements. It was found that the  $(6n+4)$  by  $(6n+4)$  banded matrix was reducible to a  $4 \times 4$  matrix calculation of the form

$$(\mathbf{M}|\mathbf{Q})\mathbf{x}=\mathbf{b}, \quad (9)$$

where the  $4 \times 4$  matrix  $(\mathbf{M}|\mathbf{Q})$  consists of a  $4 \times 1$  matrix  $\mathbf{M}$ , which depends solely on the properties of the semi-infinite fluid overlayer, augmented with a  $4 \times 3$  matrix  $\mathbf{Q}$ , which is the product of three matrices: a  $4 \times 6$  prefix matrix  $\mathbf{P}$  that is independent of fluid, sediment, or semi-infinite Biot layer properties, a  $6 \times 6$  matrix  $\mathbf{S}_N$  that depends solely on the layered sediment properties, and a  $6 \times 3$  matrix  $\mathbf{C}$  that depends solely on the semi-infinite Biot layer properties. The four-vector  $\mathbf{x}$  contains the solution for the reflection coefficient and the transmission coefficients through the multilayered sediment for the Biot fast, slow, and shear waves. The four-vector  $\mathbf{b}$  depends solely on the properties of the fluid overlayer. Since  $\mathbf{S}_N$  contains all the layered sediment information, the problem reduces largely to studying the properties of this matrix. It has the following simple form,

$$\mathbf{S}_N=\mathbf{L}_1\mathbf{L}_2\cdots\mathbf{L}_N, \quad (10)$$

TABLE I. Sand samples.

Sample #		1	2	3	4
Mesh range		120–140	80–100	40–45	25–30
Mean diam.	mm	0.12	0.17	0.40	0.64
Standard dev.	Neper	0.28	0.17	0.15	0.11

where the  $6 \times 6$  matrix  $\mathbf{L}_i$  depends solely on the properties of the  $i$ th Biot layer.  $\mathbf{L}_i$ , in turn, is defined as

$$\mathbf{L}_i=\mathbf{B}_i\mathbf{D}_i^{-1}\mathbf{B}_i^{-1}, \quad (11)$$

where the  $6 \times 6$  matrix  $\mathbf{B}_i$  depends solely on the material properties of the  $i$ th Biot layer and  $\mathbf{D}_i$  is a diagonal  $6 \times 6$  matrix depending on both the material properties and the layer thickness. Details are provided in Appendix B.

### III. MODELING OF WATER-SATURATED SAND

Of particular interest was the modeling of water-saturated sand, because the sediment in many shallow-water regions at the periphery of the oceans is sandy. The modeling effort was directed toward the four samples of well-sorted quartz sand used by Nolle *et al.*<sup>9</sup> In each sample, the grain size distribution, as shown in Fig. 1 of Ref. 9, may be approximated by a log-normal distribution. The log-normal approximation permits the statistics of the grain size to be represented by just two parameters: mean and standard deviation. The best-fit values are shown in Table I. The fast wave speed and attenuation, and the reflection loss at normal incidence, were measured by Nolle *et al.*,<sup>9</sup> and the values obtained at 500 kHz are reproduced in Table II. No corresponding measurements were given for the shear wave speed and attenuation. Typically, shear waves are measured at frequencies below 1 kHz. Bell<sup>12</sup> made some measurements in the 1 to 10 kHz band and gave average values of speed and attenuation at 5 kHz in the region of 100 m/s and 150 dB/m, respectively. The speed was independent of frequency, but attenuation increased with frequency. Hamilton<sup>13</sup> stated that

TABLE II. Model parameters.

Sample #		1	2	3	4
From Nolle <sup>a</sup> at 500 kHz					
Reflection	dB	-11	-11	-11	-11
Fast wave speed	m/s	1720	1720	1720	1720
Fast wave attenuation	dB/m	139	98	42	26
Rough estimate at 500 kHz					
Shear wave speed	m/s	100	100	100	100
Shear wave attenuation	dB/m	1500	1500	1500	1500
Biot model parameters					
Fluid viscosity ( $\eta$ )	kg/m s	0.001	0.001	0.001	0.001
Fluid mass density ( $\rho_f$ )	kg/m <sup>3</sup>	1000	1000	1000	1000
Fluid bulk modulus ( $K_f$ )	GPa	2.25	2.25	2.25	2.25
Grain mass density ( $\rho_g$ )	kg/m <sup>3</sup>	2650	2650	2650	2650
Grain bulk modulus ( $K_r$ )	GPa	7.00	7.00	7.00	7.00
Frame shear modulus ( $\mu$ )	GPa	0.02	0.02	0.02	0.02
Frame bulk modulus ( $K_b$ )	GPa	5.60	5.58	5.70	5.69
Frame porosity ( $\beta$ )	...	0.36	0.36	0.36	0.36
Frame shear log dec ( $\delta_s$ )	...	0.5049	0.4402	0.2814	0.1872
Frame bulk log dec ( $\delta$ )	...	0.0532	0.0459	0.0094	0.0017

<sup>a</sup>Reference 9.



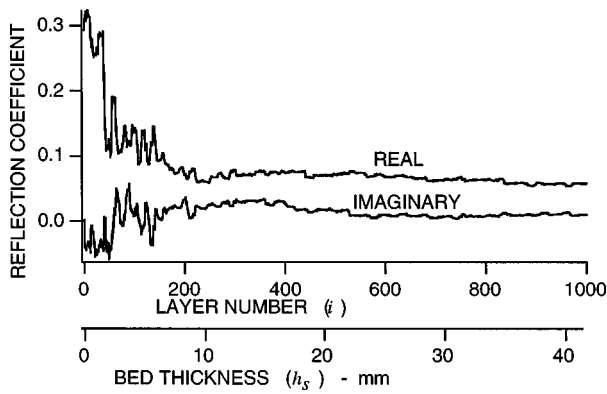


FIG. 4. Example reflection coefficient as a function of sand bed thickness from one realization of the randomly stratified medium as a function of sand bed thickness,  $h_s$ , and number of random Biot layers,  $i$ , for sand sample #1.

shear wave attenuation increases as  $f^1$  or  $f^2$  at low frequencies, or at  $f^{1/2}$  at high frequencies. Extrapolating from Bell's measurements to 500 kHz at the rate of  $f^{1/2}$ , rough estimates of shear speed and attenuation were made, as shown in Table II. These values of reflection loss, and fast and shear wave speeds and attenuations, serve as the acoustic properties that, when combined with the grain size distribution, porosity, and the physical properties of the constituent materials, define the porous medium to be modeled. Following the procedure given by Chotiros,<sup>4</sup> the Biot model parameters of the medium were estimated. The results are also shown in Table II. It should be noted that these parameter values are significantly different from those of Stoll and Kan,<sup>14</sup> which have been referred to as the "historical" values. The arguments for and against may be found in a couple of Letters to the Editor.<sup>15,16</sup>

### A. Reflection

The reflection coefficient at 500 kHz was computed as a function of the number of layers. One realization as a function of increasing sand bed thickness (or number of random layers), for sand sample #1, is shown in Fig. 4. Initially, the reflection coefficient varied wildly with the addition of each new layer due to interference effects caused by multiple scattering within the layered medium and its boundaries. As the number of layers steadily increased, the reflection coefficient appeared to converge toward an asymptotic value. A point is reached, at about 600 layers, beyond which additional layers have insignificant effect.

Using 1000 layers, the normal incidence reflection coefficient of several realizations of the randomly stratified medium were computed. In each case, the mean and standard deviation were measured from ten random realizations. The mean represents the coherent part, and the standard deviation the incoherent part, of the reflection from the sand bed. The result for sample #1 is shown in Fig. 5. There is some oscillation of the coherent component at very small thicknesses, followed by a convergence toward an asymptotic value. The asymptotic value is in the region of  $-11$  dB, consistent with that of the equivalent uniform Biot medium. The incoherent component is about 10 dB below the coherent component. Similar results were obtained from the simulations of the

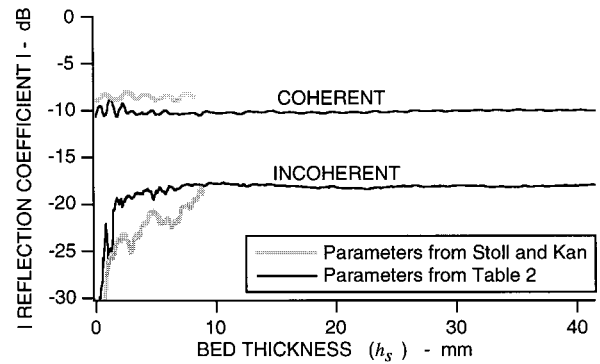


FIG. 5. Coherent and incoherent components of reflection loss as a function of sand bed thickness at normal incidence, for sample #1.

remaining three sand samples. The magnitude of the asymptotic reflection coefficient as a function of grazing angle, computed for all four sand samples, is shown in Fig. 6. Compared to the predicted reflection loss from a uniform Biot medium with the same average parameter values, they are in approximate agreement. The spread is likely due to the relatively small number of random realizations used to compute the averages. For comparison, an equivalent simulation was made using parameter values from Stoll and Kan.<sup>14</sup> Both sets of results are compared in Fig. 5. It is evident that, although the absolute mean levels are different, the effects of stratification are similar, i.e., the initial random variations and the convergence toward asymptotic values of the coherent and incoherent parts.

### B. Transmission

In addition to the reflection coefficient at the water-sand bed interface, the pressure levels of the fast, slow, and shear waves transmitted into the Biot substrate were also computed. If the sand bed were a perfectly uniform Biot medium, it would be indistinguishable from the substrate, and the fast and slow wave pressure levels, as a function of bed thickness, would be expected to decrease exponentially. The rate of decay of the slow wave would be greater due to its greater attenuation coefficient. The result for normal incidence would be as shown in Fig. 7, which was computed using the

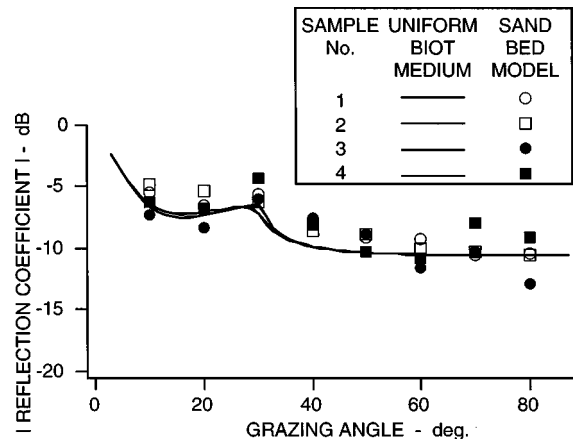


FIG. 6. Magnitude of coherent component of reflection coefficient as a function of grazing angle for all four sand samples, and corresponding curves for a uniform Biot medium.

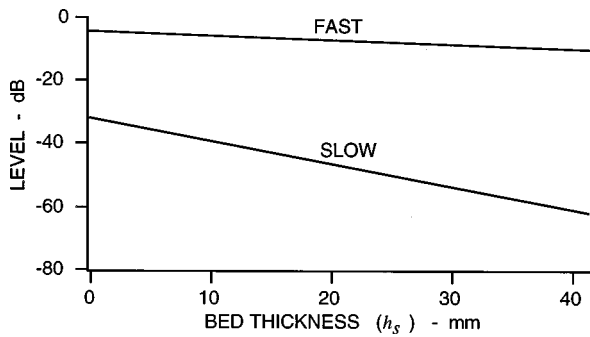


FIG. 7. Expected pressure levels of the fast and slow waves transmitted into the substrate as a function of sand bed thickness at normal incidence, if the sand bed were a uniform Biot medium, for sample #1.

same computer code as the layered medium, but with the Biot parameters in all layers set equal to that of the substrate. At normal incidence there is no shear wave generation. Modeling the sand bed as an ensemble of randomly layered Biot media, again at normal incidence, quite a different result was obtained, as shown in Fig. 8. It appears that the slow wave pressure level transmitted into the substrate goes through a number of oscillations in the first 20 mm, and then settles down to a decay rate that is identical to that of the fast wave.

The phase speed  $c_s$  at which a sinusoidal acoustic wave, at a frequency  $f$ , is propagated through the sand bed and delivered to the substrate was computed from the phase change as a function of thickness,  $\partial\phi/\partial z$ , using the wave vector relationship,

$$\left(2\pi\frac{f}{c_w}\cos(\theta)\right)^2 + \left(\frac{\partial\phi}{\partial z}\right)^2 = \left(2\pi\frac{f}{c_s}\right)^2, \quad (12)$$

where  $\theta$  is the grazing angle,  $c_w$  the sound speed in the water, and  $\phi$  the phase angle, defined as the arctangent of the imaginary and real parts of the sinusoidal signal.

The phase speed computations show that the energy that was transmitted into the substrate, in the form of both fast and slow waves, had traveled through the sand bed at the speed of the fast wave, at approximately 1700 m/s. Therefore, it must be concluded that the slow wave that was launched at the water-sand bed interface was short lived, probably going no further than the first couple of oscillations of the slow wave pressure level curve in Fig. 8. Effectively, all the energy propagating through the sand bed traveled at

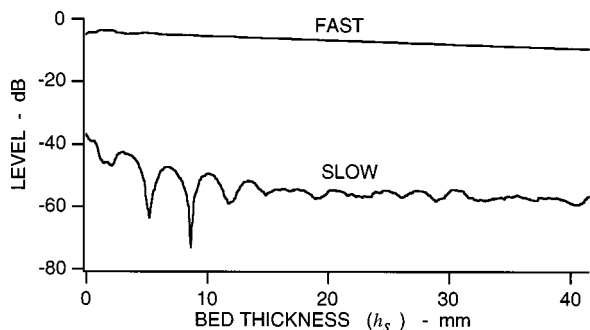


FIG. 8. Coherent components of the fast and slow waves transmitted into the substrate as a function of sand bed thickness at normal incidence, and indicated wave speeds, for sample #1.

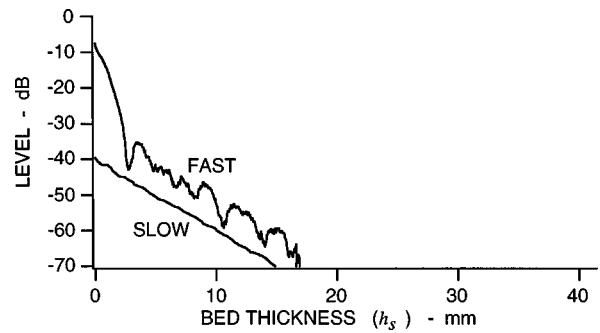


FIG. 9. Coherent components of the fast and slow waves in the substrate as a function of sand bed thickness at  $10^\circ$  grazing angle, and indicated wave speeds, using sample #1.

the fast wave speed. At the sand bed-substrate interface, part of this energy was converted into slow wave and transmitted into the substrate. This accounts for both wave types having the same decay rate. This phenomenon may account for the failure of experimental attempts to detect the slow wave in water-saturated sand by direct transmission at normal incidence, over distances greater than about a centimeter.

A similar computation at a grazing angle of  $10^\circ$ , which is well below the critical grazing angle of the fast wave, produces quite the opposite result. The pressure levels of the fast and slow waves transmitted into the Biot substrate are shown in Fig. 9. It shows that, beyond the evanescent layer, the fast wave level dropped rapidly and then decayed at the same rate as the slow wave. Analysis of the phase rates as a function of the sand bed thickness indicates that the energy, in both waves, traveled through the sand bed at the slow wave speed, at approximately 1000 m/s. This result is quite the opposite of that obtained at normal incidence.

The speed at which acoustic energy travels through the sand bed into the substrate, for each type of wave, is a very useful indicator of the wave propagation mechanism within the sand bed. The results for the fast, slow, and shear waves as a function of grazing angle are shown in Fig. 10. For the fast wave, it is seen that the effective speed through the sand bed is indeed the speed of the fast wave for grazing angles greater than  $40^\circ$ . Below this grazing angle, the energy that appears as the fast wave in the substrate actually travels through the sand bed at the slow wave speed. In the case of the slow wave, it is seen that at  $90^\circ$ , the energy that is transmitted into the substrate as a slow wave traveled through the sand bed at the speed of the fast wave. At shallow grazing angles, the slow wave energy travels through the sand bed at the appropriate slow wave speed. There is an apparent decrease in slow wave speed with increasing grazing angle. In calculating the wave speed, it had been assumed that the wave speed was isotropic. This apparent variation with grazing angle indicates that the slow wave speed is anisotropic; it appears to be faster in the horizontal direction than in the vertical direction. For the shear wave, in only a few cases does the energy travel through the sand bed at the appropriate shear wave speed; in most cases the shear wave energy is delivered at the speed of the fast wave above  $40^\circ$ , or at the speed of the slow wave at lower angles.

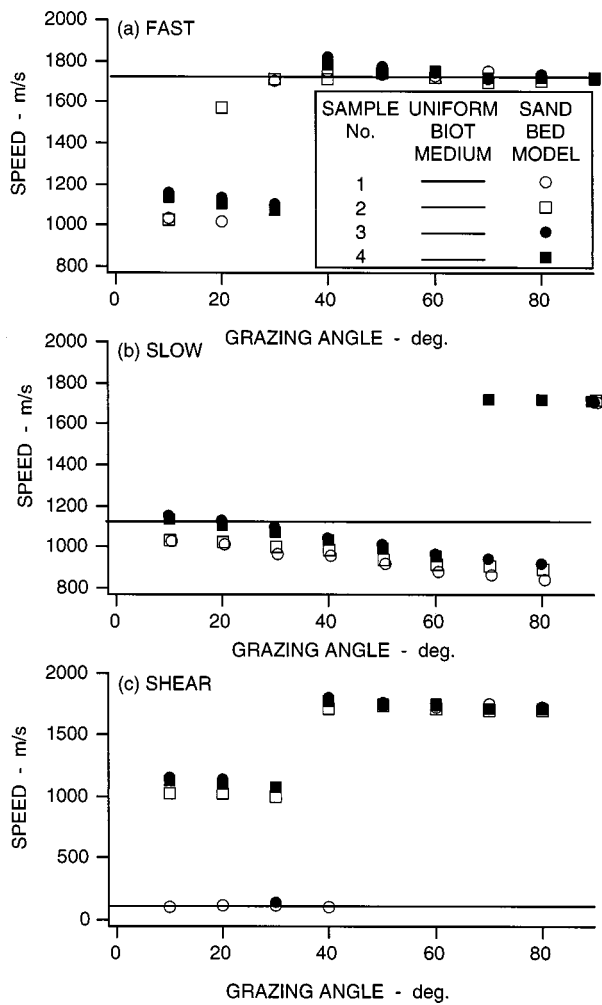


FIG. 10. Indicated transit speeds through the sand bed of (a) fast, (b) slow, and (c) shear waves transmitted into the substrate, as a function of grazing angle, computed for all sand samples.

### C. Attenuation

With reference to Fig. 10, wherever the calculated wave speed matched the indicated wave type, the attenuation coefficient of the wave within the sand bed was computed. The values are compared with those of a uniform Biot medium of the same average properties. The results are shown in Fig. 11. The fast wave attenuation is found to be at a minimum at normal incidence, with values comparable to those of the uniform Biot medium, and increases quite noticeably as the grazing angle is reduced, indicating that the medium is anisotropic. The slow wave attenuation is relatively insensitive to the grazing angle, but it is significantly higher than that of the uniform Biot medium, indicating that the random layering has caused an increase in attenuation. The shear wave attenuation is significantly different than that of the uniform Biot medium, which was designed to be 1500 dB/m at 500 kHz. Since there were very few data points of shear wave attenuation, these results should be taken as very preliminary.

### IV. CONCLUSIONS

The Biot model has been extended to include the effects of granularity in a sand bed by a numerical simulation. A

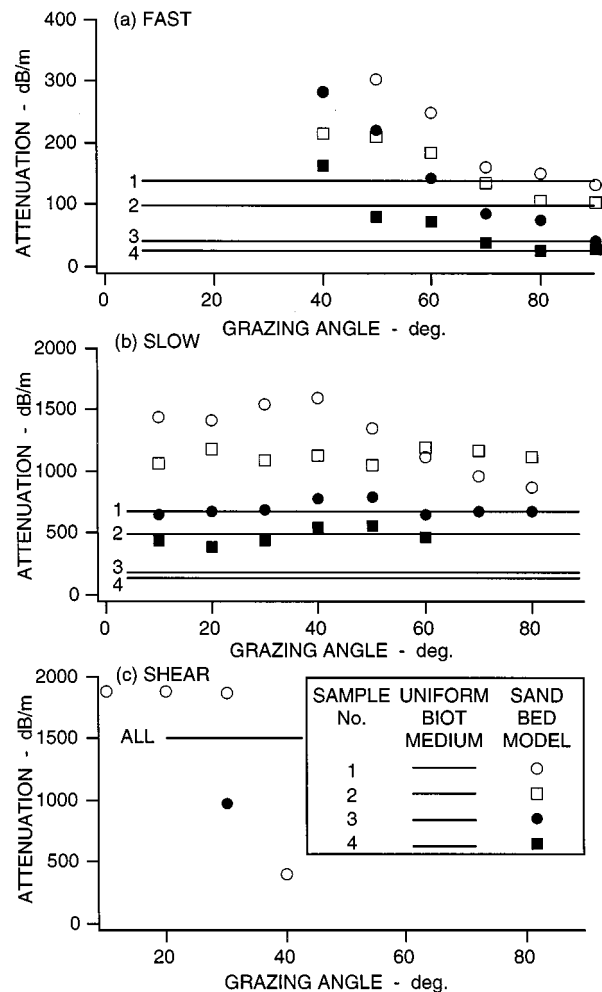


FIG. 11. Indicated transit attenuations through the sand bed of (a) fast, (b) slow, and (c) shear waves transmitted into the substrate, as a function of grazing angle, computed for all sand samples.

laminar sand bed has been treated as homogeneous Biot layers on the order of half a grain diameter thick. There are variations in porosity between layers that contain both deterministic and random components. The deterministic component is designed to have the same mean and standard deviation as that of tetrahedrally packed spherical grains. The random component attempts to account for the random departure from the tetrahedral structure and changes in grain size.

The effects of lateral variations in sediment structure on the mean sound field were simulated by coherently averaging the results for several random realizations. The reflection loss predicted by this model was found to be in good agreement with that of the equivalent uniform Biot medium. The attenuation of the fast wave was found to be anisotropic, with the minimum occurring at normal incidence. The attenuation of the slow wave was relatively insensitive to grazing angle, but elevated relative to that of the equivalent uniform Biot medium, indicating that the granular structure has significantly increased the slow wave attenuation.

Most interesting of all was the speed at which the acoustic energy traversed the modeled sand bed. At normal incidence, all of the energy transmitted into the substrate, both

fast and slow wave, arrived at the substrate at the fast wave speed. While at shallow grazing angles, practically all of the energy arrived at the slow wave speed. This phenomenon is not predicted for a uniform Biot medium, but it is the direct result of the variations in local porosity. It may explain why the experimental search for the slow wave in water-saturated sand, using normally incident sound waves, has been unsuccessful, yet the slow wave has been observed at shallow grazing angles.

This approach gives some insight into the processes that govern the reflection and propagation of sound within a granular medium such as water-saturated sand. The results are directly applicable to ocean sediments. In future work, simulation of monostatic and bistatic scattering are planned.

## ACKNOWLEDGMENTS

A large portion of the computer code used for the simulations was written and tested by Theodore J. Kim, at Applied Research Laboratories, the University of Texas at Austin. This work is supported by the Office of Naval Research, Code 321 OA.

## APPENDIX A: PROBLEM SETUP AND SOLUTION

Most of the equations used are the same as those given by Stern, Bedford, and Millwater.<sup>7</sup> There are a few typographical errors in the equations from this reference, which have been corrected. The corrected equations in this Appendix are Eqs. (A5), (A6), (A29), (A45), and (A46).

The granular ocean sediments, such as water-saturated sand, are represented as a finite granular medium bounded above by a homogeneous fluid halfspace (water) and below by a homogeneous Biot halfspace. The granular medium is approximated as homogeneous layers of uniform Biot media, differing in properties and thicknesses, as depicted in Fig. 3.

The layers are numbered 1 to  $N$ , from top to bottom. The homogeneous Biot halfspace is layer number  $N+1$ , and the fluid medium is number 0. All  $N+1$  Biot layers have identical values of the fluid viscosity ( $\eta$ ), fluid mass density ( $\rho_f$ ), fluid bulk modulus ( $K_f$ ), grain mass density ( $\rho_g$ ), grain bulk modulus ( $K_r$ ), frame shear modulus ( $\mu_0$ ), and frame bulk modulus ( $K_{b0}$ ). The fluid medium has the same fluid mass density ( $\rho_f$ ) and fluid bulk modulus ( $K_f$ ) as the Biot media. These values are given in Table II.

Each layer, numbered from 1 to  $N$ , is treated as a uniform Biot medium with a variable grain size,  $a_i$ , and porosity,  $\beta_i$ . The choice of grain size and porosity distributions determines the statistical properties of the layer.

In this model there is some correlation between the grain size,  $a$ , and layer thickness,  $h_i$ , of the  $i$ th layer. The lattice spacing of the tetrahedral structure in Fig. 2(a) may be adjusted to give any arbitrary value of mean porosity  $\beta_{\text{avg}}$ . Each pair of layers is designed to represent a monolayer of granular material of uniform grain size, as illustrated in Fig. 2(b). For a given grain size distribution, the dimensionless layer thickness distribution is assumed to be identical to that of the grain size distribution, according to the relationship,

$$\beta_{\text{avg}} = 1 - c_\beta (a/2h_i)^3. \quad (\text{A1})$$

It was arranged that the mean porosity of the layers 1 to  $N$  be the same as that of the homogeneous Biot halfspace,  $\beta_{N+1}$ . The porosity  $\beta_i$  alternates between two nominal values, to achieve the same mean and standard deviation of the local porosity in a granular structure,

$$\beta_i = (\beta_{\text{avg}} + (-1)^i \sigma_\beta) b, \text{ and } \beta_{N+1} = \beta_{\text{avg}}, \quad (\text{A2})$$

where  $b$  is a random variable with a log mean of zero and a log standard deviation of 0.6 Nepers, representing random departures from the tetrahedral structure.

The skeletal frame's bulk and shear logarithmic decrements were adjusted to give the same attenuation as the fast wave measured by Nolle and the shear wave attenuation extrapolated from the measurements of Bell. The values are as shown in Table II.

$$\delta_i = \delta_{N+1}, \quad (\text{A3})$$

$$\delta_{si} = \delta_{s(N+1)}, \quad (\text{A4})$$

$$K_{bi} = K_{b0} \left( 1 - i \frac{\delta_i}{\pi} \right), \quad (\text{A5})$$

$$\mu_i = \mu_0 \left( 1 - i \frac{\delta_{si}}{\pi} \right). \quad (\text{A6})$$

The mass density of the saturated sediment is simply

$$\rho_i = (1 - \beta_i) \rho_s + \beta_i \rho_f. \quad (\text{A7})$$

The constitutive coefficients given in Eqs. (1) and (2) are

$$M_i = \frac{K_r}{1 - \frac{K_{bi}}{K_r} + \beta_i \left( \frac{K_r}{K_f} - 1 \right)}, \quad (\text{A8})$$

$$C_i = \left( 1 - \frac{K_{bi}}{K_r} \right) M_i, \quad (\text{A9})$$

$$H_i = \left( 1 - \frac{K_{bi}}{K_r} \right) C_i + K_{bi} + \frac{4}{3} \mu_i. \quad (\text{A10})$$

A plane wave of angular frequency  $\omega$  is incident upon this sediment configuration from above with grazing angle  $\theta$ . The component of the wave number parallel to the fluid/sediment interface is

$$k = \omega \sqrt{\frac{\rho_f}{K_f}} \cos(\theta), \quad (\text{A11})$$

where  $\rho_f$  and  $K_f$  are the fluid mass density and fluid bulk modulus, as given in Table I.

For the  $i$ th layer, the virtual mass coefficient,  $c$ , and the permeability,  $\kappa$ , of Eq. (2) are given by

$$c_i = 1 + 0.5 \frac{1 - \beta_i}{\beta_i} \quad (\text{A12})$$

and

$$\kappa_i = \frac{\beta_i p_i^2}{20}, \quad (\text{A13})$$

where  $p_i$  is the pore size parameter, given by



$$p_i = \frac{\beta_i a_i}{3(1-\beta_i)}, \quad i = 1, 2, 3, \dots, N, \quad (\text{A14})$$

$$P_{N+1} = \frac{\beta_{N+1} \langle a \rangle}{3(1-\beta_{N+1})}, \quad (\text{A15})$$

and  $\langle a \rangle$  is the mean grain size.

The dynamic correction term,  $F$ , of Eq. (2), depends on the incident frequency. For the  $i$ th layer, it can be written as

$$F_i = \frac{\xi_i T(\xi_i)}{4 \left[ 1 - \frac{2T(\xi_i)}{i\xi_i} \right]}, \quad (\text{A16})$$

where

$$T(\xi_i) = \frac{\text{ber}'(\xi_i) + i \text{bei}'(\xi_i)}{\text{ber}(\xi_i) + i \text{bei}(\xi_i)}, \quad (\text{A17})$$

and

$$\xi_i = p_i \sqrt{\frac{\omega \rho_f}{\eta}}. \quad (\text{A18})$$

The horizontal position coordinate is denoted by  $x$ . The vertical position coordinate increases downward, and for the  $i$ th layer is denoted by  $z_i$ . At the upper boundary of the  $i$ th Biot layer,  $z_i = 0$ . The first vertical coordinate,  $z_0$ , within the fluid medium, is set equal to zero at the fluid/sediment interface.

The solid displacement term,  $\mathbf{u}$ , and the relative solid/fluid displacement term,  $\mathbf{w}$ , from Eqs. (1) and (2) can be expressed in terms of scalar and vector potentials,

$$\mathbf{u} = \nabla \Phi_s + \nabla \times \Psi_s, \quad (\text{A19})$$

$$\mathbf{w} = \nabla \Phi_f + \nabla \times \Psi_f. \quad (\text{A20})$$

The scalar and vector potentials for the  $i$ th layer can be written as

$$\Phi_{si} = \phi_{si}(z_i) e^{i(kx - \omega t)}, \quad (\text{A21})$$

$$\Phi_{fi} = \phi_{fi}(z_i) e^{i(kx - \omega t)}, \quad (\text{A22})$$

$$\Psi_{si} = \Psi_{si}(z_i) e^{i(kx - \omega t)} \mathbf{j}, \quad (\text{A23})$$

$$\Psi_{fi} = \psi_{fi}(z_i) e^{i(kx - \omega t)} \mathbf{j}. \quad (\text{A24})$$

Inserting these into Biot's equations of motion, Eqs. (1) and (2), yields a set of four coupled equations for the  $\phi$ s and  $\psi$ s,

$$\phi_{si}'' + \nu_{si}^2 \phi_{si} - \zeta_{si}^2 \phi_{fi} = 0, \quad (\text{A25})$$

$$\phi_{fi}'' + \chi_{fi}^2 \phi_{fi} - \lambda_{fi}^2 \phi_{si} = 0, \quad (\text{A26})$$

$$\psi_{si}'' + l_{si}^2 \psi_{si} = 0, \quad (\text{A27})$$

$$\psi_{fi} = \gamma_i \psi_{si}, \quad (\text{A28})$$

where

$$\alpha_i = i \frac{F_i^* \eta \omega}{\kappa_i}, \quad (\text{A29})$$

$$\bar{\rho}_i = \frac{c_i \rho_f}{\beta_i}, \quad (\text{A30})$$

$$\gamma_i = \frac{\rho_f \omega^2}{\bar{\rho}_i \omega^2 + \alpha_i}, \quad (\text{A31})$$

$$d_i = M_i H_i - C_i^2, \quad (\text{A32})$$

$$\kappa_{si}^2 = \frac{(M_i \rho_i - C_i \rho_f)}{d_i} \omega^2, \quad (\text{A33})$$

$$\kappa_{fi}^2 = \frac{(H_i \bar{\rho}_i - C_i \rho_f)}{d_i} \omega^2, \quad (\text{A34})$$

$$\beta_{si}^2 = \frac{(\rho_i - \gamma_i \rho_f)}{\mu_i} \omega^2, \quad (\text{A35})$$

$$\lambda_{si}^2 = \frac{(M_i \rho_f - C_i \bar{\rho}_i)}{d_i} \omega^2, \quad (\text{A36})$$

$$\lambda_{fi}^2 = \frac{(H_i \rho_f - C_i \rho_i)}{d_i} \omega^2, \quad (\text{A37})$$

$$\chi_{fi}^2 = \kappa_{fi}^2 - k^2 + \frac{H_i \alpha_i}{d_i}, \quad (\text{A38})$$

$$\zeta_{si}^2 = \lambda_{si}^2 - \frac{C_i \alpha_i}{d_i}, \quad (\text{A39})$$

$$\nu_{si}^2 = \kappa_{si}^2 - k^2, \quad (\text{A40})$$

$$l_{si}^2 = \beta_{si}^2 - k^2. \quad (\text{A41})$$

The solutions to Eqs. (A25)–(A27) are plane waves of the form

$$\begin{aligned} \phi_{si}(z_i) = & A_{+si} e^{il_{1i} z_i} + B_{+si} e^{il_{2i} z_i} \\ & + A_{-si} e^{-il_{1i} z_i} + B_{-si} e^{-il_{2i} z_i}, \end{aligned} \quad (\text{A42})$$

$$\begin{aligned} \phi_{fi}(z_i) = & \delta_{1i} A_{+si} e^{il_{1i} z_i} + \delta_{2i} B_{+si} e^{il_{2i} z_i} \\ & + \delta_{1i} A_{-si} e^{-il_{1i} z_i} + \delta_{2i} B_{-si} e^{-il_{2i} z_i}, \end{aligned} \quad (\text{A43})$$

$$\psi_{si}(z_i) = C_{+si} e^{il_{si} z_i} + C_{-si} e^{-il_{si} z_i}, \quad (\text{A44})$$

where

$$l_{1i}^2 = \frac{1}{2} (\chi_{fi}^2 + \nu_{si}^2) - \sqrt{\frac{1}{4} (\chi_{fi}^2 - \nu_{si}^2)^2 + \lambda_{fi}^2 \zeta_{si}^2}, \quad (\text{A45})$$

$$l_{2i}^2 = \frac{1}{2} (\chi_{fi}^2 + \nu_{si}^2) + \sqrt{\frac{1}{4} (\chi_{fi}^2 - \nu_{si}^2)^2 + \lambda_{fi}^2 \zeta_{si}^2}, \quad (\text{A46})$$

$$\delta_{1i} = \frac{\nu_{si}^2 - l_{1i}^2}{\zeta_{si}^2}, \quad (\text{A47})$$

$$\delta_{2i} = \frac{\nu_{si}^2 - l_{2i}^2}{\zeta_{si}^2}, \quad (\text{A48})$$

The terms  $l_{1i}$ ,  $l_{2i}$ , and  $l_{si}$  are the wave numbers of the Biot fast, slow, and shear waves within the  $i$ th layer, respectively. The “+” and “-” coefficients are the amplitudes of the downward-moving (transmitted) waves and the upward-moving (reflected) waves within the  $i$ th layer, respectively.

Let  $R$  be the reflection amplitude of the wave within the fluid medium. The unknowns of the problem are  $R$ , and the  $A_s$ ,  $B_s$ , and  $C_s$  of Eqs. (A42)–(A44). Since there are no reflections within the homogeneous semi-infinite Biot medium, we have

$$A_{-sN+1} = B_{-sN+1} = C_{-sN+1} = 0. \quad (\text{A49})$$

Thus, there are a total of  $6N + 4$  unknowns, which satisfy the  $6N + 4$  linear algebraic equations generated by the boundary conditions.

The boundary conditions pertain to continuity of fluid pressure, traction, fluid displacement, and solid displacement. Within the Biot media, the fluid (pore) pressure is

$$p_i = [M_i(\phi_{fi}'' - k^2 \phi_{fi}) - C_i(\phi_{si}'' - k^2 \phi_{si})] e^{i(kx - \omega t)},$$

$$i = 1, 2, \dots, N+1, \quad (\text{A50})$$

and within the fluid medium it is

$$p_0 = \rho_f \omega^2 (1 + R) e^{i(kx - \omega t)}. \quad (\text{A51})$$

The shear traction within the Biot media is given by

$$(\sigma_i)_{xz} = \mu_i [2ik\phi_{si}' - (\psi_{si}'' + k^2 \psi_{si})] e^{i(kx - \omega t)},$$

$$i = 1, 2, \dots, N+1, \quad (\text{A52})$$

and within the fluid medium it is

$$(\sigma_0)_{xz} = 0. \quad (\text{A53})$$

The normal traction is

$$(\sigma_i)_{zz} = [H_1(\phi_{si}'' - k^2 \phi_{si}) + 2ik\mu_i(\psi_{si}' - ik\phi_{si})$$

$$- C_i(\phi_{fi}'' - k^2 \phi_{fi})] e^{i(kx - \omega t)},$$

$$i = 1, 2, \dots, N+1, \quad (\text{A54})$$

$$(\sigma_0)_{zz} = -p_0 = -\rho_f \omega^2 (1 + R) e^{i(kx - \omega t)}, \quad (\text{A55})$$

and the normal fluid displacement is given by

$$(u_i)_z - (w_1)_z = [(\phi_{si}' + ik\psi_{si}) - (\phi_{fi}' + ik\psi_{fi})] e^{i(kx - \omega t)},$$

$$i = 1, 2, \dots, N+1, \quad (\text{A56})$$

$$(u_0)_z - (w_0)_z = ik(1 - R) \tan(\theta) e^{i(kx - \omega t)}. \quad (\text{A57})$$

Finally, within the  $i$ th Biot medium, the normal solid displacement is given by

$$(u_i)_z = (\phi_{si}' + ik\psi_{si}) e^{i(kx - \omega t)}, \quad (\text{A58})$$

and the tangential solid displacement is

$$(u_i)_x = (ik\phi_{si} - \psi_{si}') e^{i(kx - \omega t)}. \quad (\text{A59})$$

Neither of these last two quantities has counterparts within the fluid medium.

With  $h_0 = 0$ , all  $N + 1$  interfaces are subject to the boundary conditions

$$p_i|_{z_i=0} = p_{i-1}|_{z_{i-1}=h_{i-1}}, \quad i = 1, 2, \dots, N+1, \quad (\text{A60})$$

$$(\sigma_i)_{xz}|_{z_i=0} = (\sigma_{i-1})_{xz}|_{z_{i-1}=h_{i-1}}, \quad i = 1, 2, \dots, N+1, \quad (\text{A61})$$

$$(\sigma_i)_{zz}|_{z_i=0} = (\sigma_{i-1})_{zz}|_{z_{i-1}=h_{i-1}}, \quad i = 1, 2, \dots, N+1, \quad (\text{A62})$$

$$[(u_i)_z - (w_i)_z]|_{z_i=0}$$

$$= [(u_{i-1})_z - (w_{i-1})_z]|_{z_{i-1}=h_{i-1}}, \quad i = 1, 2, \dots, N+1, \quad (\text{A63})$$

and the Biot/Biot interfaces are subject to the additional two boundary conditions given by

$$(u_i)_z|_{z_i=0} = (u_{i-1})_z|_{z_{i-1}=h_{i-1}}, \quad i = 2, 3, \dots, N+1, \quad (\text{A64})$$

$$(u_i)_x|_{z_i=0} = (u_{i-1})_x|_{z_{i-1}=h_{i-1}}, \quad i = 2, 3, \dots, N+1. \quad (\text{A65})$$

The transmitted and reflected fast, slow, and shear wave amplitudes within each medium are determined by substituting Eqs. (A42)–(A44) and (A49), into the above six boundary conditions and solving the resulting set of linear algebraic equations for  $R$ ,  $A_s$ ,  $B_s$ , and  $C_s$ .

## APPENDIX B: A SIMPLE, AND VERY EFFICIENT, NUMERICAL APPROACH

The  $N$ -layer problem described in Appendix A requires solving a  $(6N + 4) \times (6N + 4)$  banded matrix equation. However, if we limit our attention to the reflection term from the sediment into the fluid, and to the transmission terms from the inhomogeneous layered sediment into the homogeneous sediment halfspace, it is possible to greatly simplify the computational complexity of the problem. The simplified approach is very computationally efficient, both in terms of memory usage and compute time, and is described below.

Given the definition of the reflection and transmission coefficients for the  $i$ th layer, as displayed in Eqs. (A42)–(A44), we define the vectors  $\mathbf{v}_i$  by

$$\mathbf{v}_i = \begin{pmatrix} A_{+si} \\ B_{+si} \\ C_{+si} \\ A_{-si} \\ B_{-si} \\ C_{-si} \end{pmatrix}, \quad \mathbf{i} = 1, 2, 3, \dots, N, \quad (\text{B1})$$

$$\mathbf{v}_{N+1} = \begin{pmatrix} A_{+sN+1} \\ B_{+sN+1} \\ C_{+sN+1} \end{pmatrix}. \quad (\text{B2})$$

The values of  $\mathbf{v}_i$  and  $\mathbf{v}_{i+1}$  are related by the boundary conditions supplied in Eqs. (A60)–(A65). These equations can be recast in matrix-vector form as

$$\mathbf{A}_i \mathbf{v}_i = \mathbf{B}_{i+1} \mathbf{v}_{i+1}, \quad \mathbf{i} = 1, 2, 3, \dots, N-1, \quad (\text{B3})$$

$$\mathbf{A}_N \mathbf{v}_N = \mathbf{C} \mathbf{v}_{N+1}, \quad (\text{B4})$$

where  $\mathbf{A}_i$  and  $\mathbf{B}_{i+1}$  are  $6 \times 6$  matrices describing the boundary conditions at the lower surface of the  $i$ th layer and the upper surface of the  $(i + 1)$ th layer, respectively, and  $\mathbf{C}$  is a  $6 \times 3$  matrix describing the boundary conditions at the surface of the Biot halfspace.

Applying the coordinate system described in the paragraph following Eq. (A18), the matrices for the  $i$ th layer,  $\mathbf{A}_i$  and  $\mathbf{B}_i$ , with layer thickness  $h_i$ , are related by

$$\mathbf{A}_i = \mathbf{B}_i \mathbf{D}_i, \quad (\text{B5})$$

where

$$\mathbf{B}_i = \begin{pmatrix} \frac{-\Theta_{1i}}{\omega^2} & \frac{-\Theta_{2i}}{\omega^2} & \frac{\Pi_{si}}{\omega^2} & \frac{\Theta_{1i}}{\omega^2} & \frac{\Theta_{2i}\mu_i}{\omega^2} & \frac{\Pi_{si}}{\omega^2} \\ \frac{\Omega_{1i}\mu_i}{\omega^2} & \frac{\Omega_{2i}}{\omega^2} & \frac{-\Theta_{si}}{\omega^2} & \frac{\Omega_{1i}}{\omega^2} & \frac{\Omega_{2i}}{\omega^2} & \frac{\Theta_{si}}{\omega^2} \\ il_{1i}\Delta_{1i} & il_{2i}\Delta_{2i} & ik\Gamma_i & -il_{1i}\Delta_{1i} & -il_{2i}\Delta_{2i} & ik\Gamma_i \\ \frac{\Sigma_{1i}}{\omega^2} & \frac{\Sigma_{2i}}{\omega^2} & 0 & \frac{\Sigma_{1i}}{\omega^2} & \frac{\Sigma_{2i}}{\omega^2} & 0 \\ il_{1i} & il_{2i} & ik & -il_{1i} & -il_{2i} & ik \\ ik & ik & -il_{si} & ik & ik & il_{si} \end{pmatrix}, \quad (\text{B6})$$

$$\mathbf{D}_i = \begin{pmatrix} e^{il_1 h_i} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{il_2 h_i} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{il_{si} h_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-il_1 h_i} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{-il_2 h_i} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-il_{si} h_i} \end{pmatrix}, \quad (\text{B7})$$

and

$$\Theta_{ij} = 2kl_{ij}\mu_i, \quad (\text{B8})$$

$$\Pi_{ij} = (l_{ij}^2 - k^2)\mu_i, \quad (\text{B9})$$

$$\Omega_{ij} = (l_{ij}^2 + k^2)(\delta_{ij}C_j - H_j) + 2k^2\mu_i, \quad (\text{B10})$$

$$\Delta_{ij} = 1 - \delta_{ij}, \quad (\text{B11})$$

$$\Gamma_i = 1 - \gamma_i, \quad (\text{B12})$$

$$\Sigma_{ij} = (C_j - \delta_{ij}M_j)(l_{ij}^2 + k^2), \quad (\text{B13})$$

and where  $M_i$ ,  $C_i$ ,  $H_i$ ,  $k$ ,  $\gamma_i$ ,  $l_{si}$ ,  $l_{1i}$ ,  $l_{2i}$ ,  $\delta_{1i}$  and  $\delta_{2i}$  are defined by Eqs. (A8)–(A11), (A31), (A41), and (A45)–(A48), respectively. Furthermore,  $\mathbf{C}$  is given by

$$\mathbf{C} = \mathbf{B}_{N+1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (\text{B14})$$

Inserting Eq. (B5) into Eq. (B3), and rearranging terms, produces

$$\mathbf{v}_i = \mathbf{D}_i^{-1} \mathbf{B}_i^{-1} \mathbf{B}_{i+1} \mathbf{v}_{i+1}, \quad i = 1, 2, 3, \dots, N-1. \quad (\text{B15})$$

Applying this recursively, we have

$$\mathbf{v}_1 = \mathbf{D}_1^{-1} \mathbf{B}_1^{-1} \left[ \prod_{j=2}^{N-1} (\mathbf{B}_j \mathbf{D}_j^{-1} \mathbf{B}_j^{-1}) \right] \mathbf{B}_N \mathbf{v}_N. \quad (\text{B16})$$

Inserting Eqs. (B4) and (B5) gives

$$\mathbf{v}_1 = \mathbf{D}_1^{-1} \mathbf{B}_1^{-1} \left[ \prod_{j=2}^N (\mathbf{B}_j \mathbf{D}_j^{-1} \mathbf{B}_j^{-1}) \right] \mathbf{C} \mathbf{v}_{N+1}. \quad (\text{B17})$$

Consider the matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \quad (\text{B18})$$

Note that right-multiplying this matrix by any  $6 \times 6$  matrix truncates the bottom two rows of the  $6 \times 6$  matrix. Thus since the rows of  $\mathbf{B}_1$ , each of which corresponds to one of Eqs. (A60)–(A65), are ordered in the same sequence as those equations, the  $4 \times 6$  matrix  $\mathbf{PB}_1$  describes the four boundary conditions at the fluid/Biot interface of the first sediment layer. Hence, the boundary conditions at the fluid/Biot interface can be written as

$$\mathbf{R}\mathbf{M} + \mathbf{P}\mathbf{B}_1 \mathbf{v}_1 = \mathbf{b}, \quad (\text{B19})$$

where  $R$  is the reflection coefficient,  $\mathbf{M}$  is a  $4 \times 1$  matrix depending solely on the fluid and incident wave properties, and  $\mathbf{b}$  is a four-vector which also depends solely on these properties.  $\mathbf{M}$  and  $\mathbf{b}$  are determined from the boundary conditions at the fluid/Biot interface, as given in Eqs. (A60)–(A63),

$$\mathbf{M} = \begin{pmatrix} 0 \\ \rho_f \\ i\omega \sin(\theta) \sqrt{\frac{\rho_f}{K_f}} \\ -\rho_f \end{pmatrix}, \quad (\text{B20})$$

$$\mathbf{b} = \begin{pmatrix} 0 \\ -\rho_f \\ i\omega \sin(\theta) \sqrt{\frac{\rho_f}{K_f}} \\ \rho_f \end{pmatrix}, \quad (\text{B21})$$

Inserting Eq. (B17) into Eq. (B19) gives

$$RM + \mathbf{P}\mathbf{S}_N\mathbf{C}\mathbf{v}_{N+1} = \mathbf{b}, \quad (\text{B22})$$

where

$$\mathbf{S}_N = \prod_{j=1}^N (\mathbf{B}_j \mathbf{D}_j^{-1} \mathbf{B}_j^{-1}). \quad (\text{B23})$$

Note that Eq. (B22) can be rewritten as

$$(\mathbf{M}|\mathbf{P}\mathbf{S}_N\mathbf{C})\mathbf{x} = \mathbf{b}, \quad (\text{B24})$$

where the  $4 \times 4$  matrix  $(\mathbf{M}|\mathbf{P}\mathbf{S}_N\mathbf{C})$  consists of the  $4 \times 1$  matrix  $\mathbf{M}$  augmented with the  $4 \times 3$  matrix  $\mathbf{P}\mathbf{S}_N\mathbf{C}$ , and

$$\mathbf{x} = \begin{pmatrix} R \\ A_{+sN+1} \\ B_{+sN+1} \\ C_{+sN+1} \end{pmatrix}. \quad (\text{B25})$$

Thus the problem of Appendix A can be solved quite simply and efficiently by forming the  $4 \times 4$  matrix  $(\mathbf{M}|\mathbf{P}\mathbf{S}_N\mathbf{C})$  and then solving for  $\mathbf{x}$  in Eq. (B18). When compared to the original  $(6N+4) \times (6N+4)$  banded matrix algorithm, with  $N = 500$ , this method was very fast and consistently produced the same results to 14 out of 16 significant figures. However, the advantage of this approach is not purely computational. Note that all the layered sediment information is contained in the  $6 \times 6$  matrix  $\mathbf{S}_N$ , so the behavior of an ensemble of randomly layered sediments could, in principle, be investigated by studying the statistical properties of this matrix.

<sup>1</sup>D. R. Jackson, S. P. Winebrenner, and A. Ishimaru, "Application of composite roughness model to high-frequency bottom backscattering," *J. Acoust. Soc. Am.* **79**(5), 1410–1422 (1986).

<sup>2</sup>A. N. Ivakin and Yu. P. Lysanov, "Underwater sound scattering by volume inhomogeneities of a bottom bounded by a rough surface," *Sov. Phys. Acoust.* **27**(3), 212–215 (1981).

<sup>3</sup>A. Altenburg and N. P. Chotiros, "Plane-wave analysis of acoustic signals in a sandy sediment," *J. Acoust. Soc. Am.* **89**(1), 165–170 (1991).

<sup>4</sup>N. P. Chotiros, "Biot model of sound propagation in water saturated sand," *J. Acoust. Soc. Am.* **97**(1), 199–214 (1995).

<sup>5</sup>M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid, I. Low-frequency range," *J. Acoust. Soc. Am.* **28**, 168–178 (1956).

<sup>6</sup>M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid, II. Higher frequency range," *J. Acoust. Soc. Am.* **28**, 179–191 (1956).

<sup>7</sup>M. Stern, A. Bedford, and H. R. Millwater, "Wave reflection from a sediment layer with depth dependent properties," *J. Acoust. Soc. Am.* **77**(5), 1781–1788 (1985).

<sup>8</sup>F. J. Pettijohn, P. E. Potter, and R. Siever, *Sand and Sandstone* (Springer-Verlag, New York, 1987).

<sup>9</sup>A. W. Nolle, W. A. Hoyer, J. F. Mifsud, W. R. Runyan, and M. B. Ward, "Acoustical properties of water-filled sands," *J. Acoust. Soc. Am.* **35**(9), 1394–1408 (1963).

<sup>10</sup>J. L. Finney and J. Wallace, "Interstitial correlation functions; a new, sensitive characterization of non-crystalline packed structures," *J. Non-Cryst. Solids* **43**, 165–187 (1981).

<sup>11</sup>A. E. R. Westman and H. R. Hugill, "The packing of particles," *J. Am. Ceram. Soc.* **13**, 767 (1930).

<sup>12</sup>D. W. Bell, *Shear Wave Propagation in Unconsolidated Fluid Saturated Porous Media*, ARL-TR-79-31 (Applied Research Labs., Univ. of Texas, Austin, 1979).

<sup>13</sup>E. L. Hamilton, "Geoacoustic modeling of the sea floor," *J. Acoust. Soc. Am.* **68**(5), 1313–1340 (1980).

<sup>14</sup>R. D. Stoll and T. K. Kan, "Reflection of acoustic waves at a water-sediment interface," *J. Acoust. Soc. Am.* **70**, 149–156 (1981).

<sup>15</sup>R. D. Stoll, "Comments on 'Biot model of sound propagation in water-saturated sand' [*J. Acoust. Soc. Am.* **97**, 199–214 (1995)]," *J. Acoust. Soc. Am.* **103**(5), 2723–2725 (1998).

<sup>16</sup>N. P. Chotiros, "Reply to: 'Biot slow waves in sands near the seafloor, by R. D. Stoll'," *J. Acoust. Soc. Am.* **103**(5), 2726–2729 (1998).



# On the phase speed and attenuation of an interface wave in an unconsolidated marine sediment

Michael J. Buckingham<sup>a)</sup>

Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0213

(Received 23 March 1998; accepted for publication 7 June 1999)

The phase speed and attenuation of the interface wave at the seawater–sediment boundary are obtained by solving the characteristic equation for one of its complex roots. The characteristic equation itself is derived on the basis of a recently developed theory of wave propagation in porous media. Central to the theory is the stress relaxation that occurs when mineral grains slide against one another during the passage of a seismic wave. This type of stress relaxation is characterized by material response functions for compressional and shear waves of the form  $h(t) \propto t^{-n}$ , where  $t$  is time since the sliding began and  $n$  is a small positive number. The phase speed of the interface wave relative to that of the shear wave depends weakly on the grain size, increasing from about 85% for fine-grained silts and clays to 90% for coarse sands. The loss tangent of the interface wave,  $\beta_i$ , is found to be independent of the mechanical properties (grain size, porosity, and density) of the sediment, and is the same as that for the shear wave:  $\beta_i \approx 0.04$ . Since the loss tangent and phase speed are, in effect, independent of frequency, the attenuation coefficient of the interface wave scales as the first power of frequency. It turns out that the characteristic equation for the interface wave, as derived from the intergranular stress-relaxation mechanism, is exactly the same as if the sediment had been treated as an elastic solid. However, the elastic description fails to account for the grain-size dependencies exhibited by the compressional and shear waves. These dependencies emerge naturally from the stress-relaxation model. © 1999 Acoustical Society of America. [S0001-4966(99)00710-9]

PACS numbers: 43.20.Jr, 43.30.Ma, 43.30.Ky, 43.35.Mr. [DLB]

## LIST OF SYMBOLS

$j$	$\sqrt{-1}$	$s = k_s^2/p^2$	dimensionless variable in characteristic equation
$c_p$	compressional wave speed in sediment	$\eta = \sqrt{k^2 - p^2}$	vertical compressional wave number in sediment
$c_s$	shear wave speed in sediment	$\eta_s = \sqrt{k_s^2 - p^2}$	vertical shear wave number in sediment
$c_0$	compressional wave speed in sediment in absence of intergranular stress relaxation	$\eta_w = \sqrt{k_w^2 - p^2}$	vertical compressional wave number in seawater
$c_w$	compressional wave speed in water column	$t$	time
$c_i$	interface wave speed	$h(t)$	compressional material response function
$\eta_h$	coefficient of shear stress relaxation	$h_s(t)$	shear material response function
$\lambda_h$	coefficient of compressional stress relaxation	$H(j\omega)$	Fourier transform of $h(t)$
$\mu_c$	compressional stress-relaxation rigidity modulus	$H_s(j\omega)$	Fourier transform of $h_s(t)$
$\mu_s$	shear stress-relaxation rigidity modulus	$\psi_w$	velocity potential, compressional wave in water column
$\omega$	angular frequency	$\psi$	velocity potential, compressional wave in sediment
$k_0 = \omega/c_0$	compressional wave number in sediment in absence of intergranular stress relaxation	$\psi_s$	velocity potential, shear wave in sediment
$k = \omega/c_p$	compressional wave number in sediment	$n$	exponent of compressional material response function ( $0 < n < 1$ )
$k_s = \omega/c_s$	shear wave number in sediment	$m$	exponent of shear material response function ( $0 < m < 1$ )
$k_w = \omega/c_w$	compressional wave number in seawater	$u_g$	mean grain diameter, micrometers
$p$	horizontal wave number	$\rho_0$	density of sediment
		$\rho_w = 1024 \text{ kg/m}^3$	density of seawater
		$\rho_g = 2700 \text{ kg/m}^3$	density of mineral grains
		$b = \rho_w/\rho_0$	density ratio
		$\kappa_0$	bulk modulus of sediment

<sup>a)</sup>Also at: Institute of Sound and Vibration Research, The University, Southampton SO17 1BJ, England. Electronic mail: mjb@mpl.ucsd.edu

$\kappa_w = 2.25 \times 10^9$ Pa	bulk modulus of seawater
$\kappa_g = 1.47 \times 10^{10}$ Pa	bulk modulus of mineral grains
$\Delta = 3$ $\mu$ m	rms grain roughness
$u_0 = 1000$ $\mu$ m	reference grain diameter
$\mu_0 = 2 \times 10^9$ Pa	compressional stress-relaxation scaling modulus
$\mu_1 = 5.1 \times 10^7$ Pa	shear stress-relaxation scaling modulus
$N$	porosity of sediment
$P = 0.63$	packing factor of randomly packed, uniform spheres

## INTRODUCTION

In a recent series of papers, Buckingham<sup>1-4</sup> has developed a theory of seismic wave propagation in unconsolidated and consolidated saturated granular media. Central to the theory are material response (or memory) functions of the form  $h(t) \propto t^{-n}$ , where the index  $n$  is positive and much less than unity. This type of material response function represents a particular form of stress relaxation, proposed as arising from the microscopic sliding that occurs at grain-to-grain contacts during the passage of a wave through the medium. According to the theory, the stress relaxation introduces rigidity into the medium,<sup>2</sup> which allows shear waves to propagate even though the elastic rigidity (or “frame”) modulus of the material is taken to be identically zero. Moreover, the predicted seismic-wave attenuation is proportional to the first power of frequency, which is consistent with the measured attenuation of compressional and shear waves in both unconsolidated marine sediments<sup>5-8</sup> and consolidated, sedimentary rocks such as sandstone or limestone.<sup>9</sup>

A material that supports shear, whether associated with elasticity or stress relaxation, may also be expected to support an interface wave. The speed and attenuation of the interface wave will be governed by a characteristic equation that is determined from the boundary conditions. The purpose of this article is to derive the characteristic equation for an interface wave propagating along the boundary between an unconsolidated, porous sediment of the type discussed by Buckingham<sup>1-4</sup> and an overlying fluid (seawater). Complex solutions of this equation are developed which yield the interface wave speed and the attenuation coefficient.

Before proceeding with the discussion, it is appropriate to address a question of semantics. By the term “interface wave,” we mean the classical type of evanescent surface wave that propagates along the boundary between a homogeneous solid and: (1) a vacuum (Rayleigh wave); or (2) a fluid (Scholte wave); or (3) another solid (Stoneley wave). Such waves exist when the solid supports shear through elasticity. When the shear arises from stress relaxation rather than from elasticity, as is taken to be the case for the granular media considered below, the interface wave will be identified by the prefix “pseudo.” Thus, at the seawater–sediment boundary we shall refer to the “pseudo-Scholte” wave. This usage is not to be confused with terms such as “pseudo-Rayleigh” wave used by some authors<sup>10</sup> in connection with an interface wave seen at the boundary between an elastic solid and a thin layer of fluid.

Following Buckingham,<sup>1-4</sup> it is assumed in deriving the

characteristic equation that the unconsolidated sediment possesses no elastic frame; that is to say, the elastic modulus of the material is taken to be zero. Stress relaxation, arising from grain-to-grain sliding, as discussed by Buckingham,<sup>4</sup> is considered to be the mechanism that gives rise to shear, dissipation, and the existence of the interface wave. The characteristic equation itself is a function of the variable  $s$ , which is defined as the ratio of the complex wave number of the shear wave to the complex wave number of the interface wave. Underpinning the analysis are the usual boundary conditions; that is, continuity of total stress and of normal component of particle velocity across the seawater–sediment interface. In the absence of attenuation in the granular medium, one root of the characteristic equation would be real. This root becomes complex when attenuation is included in the analysis. The attenuation coefficient and the phase speed of the pseudo-Scholte wave are derived by solving the characteristic equation for this complex root. As with the attenuation of the compressional and shear waves, the attenuation coefficient of the interface wave is found to scale linearly with frequency. Although experimental data on the attenuation of interface waves in marine sediments are scarce, a comparison is made with the few data sets that are available, from which it is evident that the theoretically predicted attenuation is consistent with the observations.

## I. WAVE EQUATIONS AND MATERIAL RESPONSE FUNCTIONS

Figure 1 illustrates the geometry of the problem. A horizontal boundary representing the seafloor separates two infinite, homogeneous, isotropic half spaces, the lower being the unconsolidated sediment consisting of mineral grains and interstitial seawater and the upper a conventional fluid, in this case seawater. An impulse of momentum (e.g., a hammer blow) is applied at a point on the seabed. The resultant wave

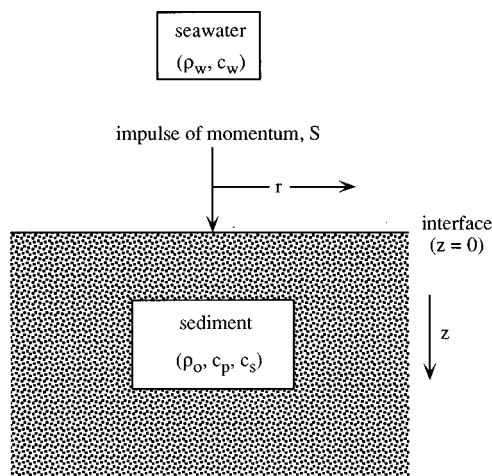


FIG. 1. Schematic showing the geometry of the fluid–sediment interface problem.

fields in both media are to be determined subject to the requirements that the stress and the normal component of particle velocity should be continuous across the interface.

The equation to be solved for the compressional wave field in the (assumed lossless) fluid above the boundary is

$$\nabla^2 \Psi_w - \frac{1}{c_w^2} \frac{\partial^2 \Psi_w}{\partial t^2} = 0, \quad (1)$$

where  $\nabla^2$  is the Laplacian,  $\Psi_w$  is the velocity potential, and  $c_w$  is the phase speed in the fluid (seawater) medium. The equations for the compressional and shear wave in the unconsolidated sediment have been derived by Buckingham<sup>4</sup> by considering the stress tensor characterizing the stress relaxation arising from a microscopic, grain-to-grain sliding process in which the interaction becomes ‘‘harder’’ as the sliding progresses. For the compressional wave, the resultant equation is

$$\nabla^2 \psi - \frac{1}{c_0^2} \frac{\partial^2 \psi}{\partial t^2} + \frac{\lambda_h}{\rho_0 c_0^2} \frac{\partial}{\partial t} \nabla^2 [h(t) \otimes \psi] + \frac{(4/3) \eta_h}{\rho_0 c_0^2} \frac{\partial}{\partial t} \nabla^2 [h_s(t) \otimes \psi] = 0, \quad (2)$$

and for the shear wave,

$$\frac{\eta_h}{\rho_0} \nabla^2 [h_s(t) \otimes \psi_s] - \frac{\partial \psi_s}{\partial t} = 0, \quad (3)$$

where  $\psi$ ,  $\psi_s$  are the velocity potentials of the compressional and shear wave, respectively,  $c_0$  is the compressional wave speed in the limit of low frequency,  $\rho_0$  is the bulk density of the sediment, and the symbol  $\otimes$  denotes a temporal convolution. The coefficients  $\lambda_h$  and  $\eta_h$  are analogs of the bulk and shear viscosities, respectively, of a conventional viscous fluid. In the case of the granular sediment, these coefficients represent the effect of the slippage that occurs when grains either slide against one another ( $\eta_h$ ) or are compressed together ( $\lambda_h$ ). The convolution operations in Eqs. (2) and (3) allow for the fact that the stress and the strain rate associated with grain-to-grain sliding are, in general, out of phase (unlike the case of a viscous fluid, where, according to Newton’s law of viscous flow for streamline motion, the stress is proportional to the strain rate). The material response functions for the compressional and shear wave, respectively, are

$$h(t) = u(t) \frac{t_0^{n-1}}{t^n}, \quad 0 < n < 1, \quad (4)$$

and

$$h_s(t) = u(t) \frac{t_1^{m-1}}{t^m}, \quad 0 < m < 1, \quad (5)$$

where the unit step function,  $u(t)$ , ensures that response of the medium is causal. These fractional-power functions were originally used by Nutting<sup>11</sup> to represent empirically the stress relaxation observed in plastic materials. Equations (4) and (5) have recently been derived by Buckingham<sup>4</sup> on the basis of a slip mechanism in which the grain-to-grain sliding becomes harder as the process progresses (i.e., the coefficient

of the equivalent dashpot representing the grain-to-grain interaction increases with increasing time).

By Fourier transforming Eqs. (1) to (3) with respect to time, the corresponding reduced wave equations are obtained:

$$\nabla^2 \Psi_w + \frac{\omega^2}{c_w^2} \Psi_w = 0, \quad (6)$$

$$\nabla^2 \Psi + \frac{\omega^2}{c_0^2} \Psi + j\omega \frac{1}{\rho_0 c_0^2} \left\{ \lambda_h H(j\omega) + \frac{4}{3} \eta_h H_s(j\omega) \right\} \nabla^2 \Psi = 0, \quad (7)$$

and

$$\frac{\eta_h}{\rho_0} H_s(j\omega) \nabla^2 \Psi_s - j\omega \Psi_s = 0, \quad (8)$$

where  $\omega$  is angular frequency and the upper-case symbols,  $\Psi_w$ ,  $\Psi$ , and  $\Psi_s$ , are the Fourier transforms of the corresponding lower-case velocity potentials. The Fourier-transformed material response functions are

$$H(j\omega) = \frac{\Gamma(1-n)}{(j\omega t_0)^{1-n}}, \quad (9)$$

and

$$H_s(j\omega) = \frac{\Gamma(1-m)}{(j\omega t_1)^{1-m}}. \quad (10)$$

As discussed by Buckingham,<sup>2</sup> when Eqs. (9) and (10) are substituted into Eqs. (7) and (8), the reduced wave equations for the sediment become

$$\nabla^2 \Psi + \frac{\omega^2}{c_0^2 q^2} \Psi = 0, \quad (11)$$

and

$$\nabla^2 \Psi_s + \frac{\omega^2}{c_0^2 q_s^2} \Psi_s = 0, \quad (12)$$

where

$$q = \left[ 1 + \frac{j\omega}{\rho_0 c_0^2} \left\{ \lambda_h H(j\omega) + \frac{4}{3} \eta_h H_s(j\omega) \right\} \right]^{1/2} = \left[ 1 + \frac{\mu_c}{\rho_0 c_0^2} (j\omega t_0)^n + \frac{4\mu_s}{3\rho_0 c_0^2} (j\omega t_1)^m \right]^{1/2} \quad (13a)$$

and

$$q_s = \left[ \frac{j\omega}{\rho_0 c_0^2} \eta_h H_s(j\omega) \right]^{1/2} = \left[ \frac{\mu_s}{\rho_0 c_0^2} (j\omega t_1)^m \right]^{1/2}. \quad (13b)$$

In these expressions,

$$\mu_c = \lambda_h \frac{\Gamma(1-n)}{t_0} \quad (14)$$

and

$$\mu_s = \eta_h \frac{\Gamma(1-m)}{t_1}. \quad (15)$$

In effect,  $\mu_c$  and  $\mu_s$  are stress-relaxation rigidity moduli for the sediment. They represent the stiffness introduced by the

intergranular sliding that occurs when grains are either compressed together ( $\mu_c$ ) or translated relative to one another ( $\mu_s$ ). These two sliding processes, that is, compressional and translational, have been described in detail by Buckingham.<sup>4</sup>

## II. BOUNDARY CONDITIONS

Across the boundary between the sediment and the overlying seawater, the normal and tangential components of stress and the normal component of particle velocity must all be continuous. Since the upper medium is a fluid, which is incapable of supporting shear, the tangential component of the stress at the boundary is zero. To derive the boundary conditions, we consider first the stress tensor for the sediment.

It is implicit in the formulation of Eqs. (2) and (3) that the elements of the stress tensor for the sediment can be expressed in the form<sup>4</sup>

$$\begin{aligned} \tau_{ij} = & \{p - \lambda_h[h(t) \otimes \underline{v}] + \frac{2}{3} \eta_h \operatorname{div}[h_s(t) \otimes \underline{v}]\} \\ & - \eta_h \left[ h_s(t) \otimes \left\{ \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right\} \right], \end{aligned} \quad (16)$$

where  $\delta_{ij}$  is the Kronecker delta and  $p$  is the acoustic pressure fluctuation. This expression is the same as that for a viscous, compressible fluid,<sup>12</sup> except for the presence of the convolution operations involving the material response functions  $h(t)$  and  $h_s(t)$ . Obviously, if these two material response functions were delta functions, Eq. (16) would be identical to the expression for the stress in a viscous fluid (since a function convolved with a delta function is the function itself). The normal stress and the tangential stress, as given by Eq. (16), must both be continuous across the sediment–seawater boundary.

Proceeding in standard fashion, the vector field,  $\underline{v}$ , is expressed from Helmholtz's theorem<sup>13</sup> as the sum of the gradient of a scalar (velocity) potential  $\Psi$  and the curl of a zero-divergence vector potential  $\underline{A}$  as follows:

$$\underline{v} = \operatorname{grad} \psi + \operatorname{curl} \underline{A}, \quad \text{where } \operatorname{div} \underline{A} = 0. \quad (17)$$

From the cylindrical symmetry of the problem, the vector potential may be expressed in cylindrical coordinates as

$$\underline{A} = -\frac{\partial \Psi_s}{\partial r} \underline{e}_\phi, \quad (18)$$

where  $\underline{e}_\phi$  is the unit azimuthal vector. By combining Eqs. (16) to (18), the normal component of the stress at the boundary, in the frequency domain and in cylindrical coordinates, is found to be

$$\begin{aligned} T_{zz}|_{z=0} = & -j\omega\rho_0\Psi + 2\eta_h H_s \\ & \times \left[ \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \Psi}{\partial r} \right) + \frac{1}{r} \frac{\partial^2}{\partial r \partial z} \left( r \frac{\partial \Psi_s}{\partial r} \right) \right] \\ = & -S \frac{\delta(r)}{2\pi r} - j\omega\rho_w \Psi_w, \end{aligned} \quad (19)$$

where  $S$  in the first term on the far right represents the impulse of momentum applied at the surface of the sediment and the second term is the pressure in the water column

immediately above the bottom. Similarly, the tangential stress at the interface is

$$T_{rz}|_{z=0} = -\eta_h H_s \frac{\partial}{\partial r} \left[ 2 \frac{\partial \Psi}{\partial z} + 2 \frac{\partial^2 \Psi_s}{\partial z^2} + k_s^2 \Psi_s \right] = 0, \quad (20)$$

where

$$k_s = \frac{\omega}{c_0 q_s} \quad (21)$$

is the complex wave number of the shear wave.

In deriving Eqs. (19) and (20), the following expressions for the Fourier-transformed radial ( $V_r$ ) and normal ( $V_z$ ) particle velocities in the sediment have been used:

$$V_r = \frac{\partial \Psi}{\partial r} + \frac{\partial^2 \Psi_s}{\partial r \partial z} \quad (22)$$

and

$$V_z = \frac{\partial \Psi}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \Psi_s}{\partial r} \right). \quad (23)$$

Equations (22) and (23) are familiar forms,<sup>14</sup> which are derived directly from Eqs. (17) and (18). It follows from Eq. (23) that continuity of the normal component of velocity across the bottom interface requires that

$$\frac{\partial \Psi}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \Psi_s}{\partial r} \right) = \frac{\partial \Psi_w}{\partial z} \quad \text{at } z=0. \quad (24)$$

The reduced wave equations in the preceding section, Eqs. (6), (11), and (12), are now to be solved subject to the boundary conditions expressed in Eqs. (19), (20), and (24).

## III. THE CHARACTERISTIC EQUATION

When a Hankel transform of order zero is applied over horizontal range,  $r$ , to the three wave equations [Eqs. (6), (11), and (12)], each reduces to an ordinary differential equation in which the independent variable is the depth coordinate,  $z$

$$\frac{\partial^2 \Psi_{wp}}{\partial z^2} + \eta_w^2 \Psi_{1p} = 0, \quad (25)$$

$$\frac{\partial^2 \Psi_p}{\partial z^2} + \eta^2 \Psi_p = 0, \quad (26)$$

and

$$\frac{\partial^2 \Psi_{sp}}{\partial z^2} + \eta_s^2 \Psi_{sp} = 0, \quad (27)$$

where

$$\eta_w = \sqrt{k_w^2 - p^2}, \quad \operatorname{imag}(\eta_w) > 0, \quad (28)$$

$$\eta = \sqrt{k^2 - p^2}, \quad \operatorname{imag}(\eta) > 0, \quad (29)$$

and

$$\eta_s = \sqrt{k_s^2 - p^2}, \quad \operatorname{imag}(\eta_s) > 0. \quad (30)$$

The wave numbers in these expressions are  $k_w = \omega/c_w$ ,  $k = k_0/q$ , and  $k_s = k_0/q_s$ , where  $k_0 = \omega/c_0$ , the Hankel transform variable,  $p$ , is the horizontal wave number, and the



subscript  $p$  indicates that the associated wave function is a Hankel transform. For  $z$  increasing downwards, the solutions of Eqs. (25)–(27) are

$$\Psi_{wp} = B_w \exp(-j \eta_w z), \quad (31)$$

$$\Psi_p = B \exp(j \eta z), \quad (32)$$

and

$$\Psi_{sp} = B_s \exp(j \eta_s z), \quad (33)$$

where the sign of each exponent has been chosen to ensure that the associated field goes to zero infinitely far from the interface and the three  $B$  coefficients are constants of integration which are to be determined from the boundary conditions.

When the condition in Eq. (19) on the normal stress at the boundary is Hankel transformed over range, it reduces to

$$(j \omega \rho_0 + 2 \eta_h p^2 H_s) B + 2 j \eta_s \eta_h p^2 H_s B_s = \frac{S}{4 \pi} + j \omega \rho_w B_w, \quad (34)$$

where the expressions in Eqs. (31) to (33) have been substituted for the various field terms. Similarly, on Hankel transforming the condition on the tangential stress in Eq. (20), the result

$$2 j \eta B - 2 \eta_s^2 B_s + k_s^2 B_s = 0 \quad (35)$$

is obtained. In fact, to derive this expression a Hankel transform of first order was applied to Eq. (20), since the first-order transform of the derivative of a function is equivalent to the zero-order transform of the function itself. Finally, the third of the boundary conditions, on the normal component of velocity in Eq. (24), yields, after a zero-order Hankel transform,

$$j \eta B + p^2 B_s = -j \eta_w B_w. \quad (36)$$

Equations (34)–(36) are a set of simultaneous equations that may be solved for the coefficients  $B$ ,  $B_w$ , and  $B_s$ . The results are as follows:

$$B = \frac{S}{4 \pi} \frac{(2 \eta_s^2 - k_s^2)}{D}, \quad (37)$$

$$B_s = \frac{S}{4 \pi} \frac{2 j \eta}{D}, \quad (38)$$

and

$$B_w = -\frac{S}{4 \pi} \frac{\eta}{\eta_w D} [2 \eta_s^2 - k_s^2 + 2 p^2]. \quad (39)$$

The term  $D$  appearing in the denominator of each of these expressions is

$$D = \left( j \omega \rho_0 + 2 \eta_h p^2 H_s + j \omega \rho_w \frac{\eta}{\eta_w} \right) (2 \eta_s^2 - k_s^2) - 4 \eta \eta_s \eta_h p^2 H_s + 2 j \omega \rho_w p^2 \frac{\eta}{\eta_w}. \quad (40)$$

Thus,  $D$  is a function of  $p$ , the integration variable of the inverse Hankel transform that must be applied to Eqs. (31)–(33) in order to obtain the frequency-domain velocity poten-

tial in the water column and the compressional and shear fields in the sediment. A zero of  $D$  corresponds to a pole in the complex  $p$ -plane and one of these poles represents an interface wave propagating along the water–sediment boundary. Notice that  $D$  does not depend on the bulk material response function,  $H$ .

To investigate the pole representing the interface wave, it is necessary to find the corresponding (complex) root of  $D$ , that is, to solve the characteristic equation

$$D = 0. \quad (41)$$

It is convenient to make the substitution

$$u^2 = -\frac{j \omega \rho_0}{\eta_h H_s} = \frac{\omega^2 \rho_0 t_0}{\eta_h \Gamma(1-m)(j \omega t_1)^m}, \quad (42)$$

in which case the characteristic equation can be expressed as

$$(2 p^2 - u^2)(k_s^2 - 2 p^2) + \frac{j \omega \rho_w \sqrt{k^2 - p^2}}{\eta_h H_s \sqrt{k_w^2 - p^2}} k_s^2 - 4 p^2 \sqrt{k^2 - p^2} \sqrt{k_s^2 - p^2} = 0, \quad (43)$$

where  $b$  is the density ratio

$$b = \frac{\rho_w}{\rho_0} < 1. \quad (44)$$

Equation (43) may be further simplified by letting

$$x = \frac{k^2}{k_s^2}, \quad (45a)$$

$$y = \frac{u^2}{k_s^2} = 1, \quad (45b)$$

and

$$w = \frac{k_w^2}{k_s^2}, \quad (45c)$$

to obtain

$$4 \sqrt{1-x} \sqrt{1-s} - (s-2)(sy-2) - bs^2 y \sqrt{\frac{1-xs}{1-ws}} = 0, \quad (46)$$

which, since  $y$  is identically equal to unity, is equivalent to

$$4 \sqrt{1-x} \sqrt{1-s} - (s-2)^2 - bs^2 \sqrt{\frac{1-xs}{1-ws}} = 0. \quad (47)$$

The variable  $s$ , as defined in Eq. (45a), is a measure of the (complex) interface wave speed relative to the (complex) shear wave speed in the sediment. Before solving for  $s$ , it is worth commenting that the characteristic equation in Eq. (47) is identical in form to that for an elastic solid overlain by a fluid.<sup>15</sup> In other words, the characteristic equation for the pseudo-Scholte wave in Eq. (47) is exactly the same as the corresponding equation for a true Scholte wave,<sup>16</sup> even though the underlying physical mechanisms giving rise to the two types of interface wave are completely different. Notice that for the special case when  $b=0$ , representing the situation where the fluid medium is a vacuum, Eq. (47) re-

duces identically to the characteristic equation for a Rayleigh wave.

#### IV. COMPLEX-ROOT FINDING

In the absence of attenuation, all the parameters in Eq. (47) are real, a real solution for  $s$  exists which is positive and less than unity and which may be found using an iterative Newton–Raphson procedure. This real root yields the phase speed of the interface wave relative to the shear-wave speed. When attenuation in the sediment is present, however, the required root becomes complex, in which case the Newton–Raphson algorithm cannot be applied directly, indicating that some alternative method must be employed.

Rather than turn to a complex-root-finding algorithm, an approximate approach is introduced here, which is based on the observation that, for unconsolidated marine sediments, the imaginary parts of all the complex variables in Eq. (47) are very much smaller than the corresponding real parts. Indicating the real and imaginary parts by single and double primes, respectively, we have

$$s = s' + js'', \quad s'' \ll s', \quad (48a)$$

$$x = x' + jx'', \quad x'' \ll x', \quad (48b)$$

and

$$w = w' + jw'', \quad w'' \ll w'. \quad (48c)$$

Similarly, the characteristic equation may be written as

$$D = D' + jD'' = 0, \quad (49)$$

which is satisfied only if the real ( $D'$ ) and imaginary ( $D''$ ) parts are individually equal to zero.

When the expressions in Eqs. (48) are substituted into Eq. (47), we find that  $D'$  and  $D''$  are, to first order in the small imaginary terms, as follows:

$$D' = 4\sqrt{1-x's'}\sqrt{1-s'} - (s'-2)^2 - bs'^2\sqrt{\frac{1-x's'}{1-w's'}} = 0 \quad (50a)$$

and

$$D'' = -As'' - B = 0, \quad (50b)$$

where

$$A = 2x'\frac{\sqrt{1-s'}}{\sqrt{1-x's'}} + 2\frac{\sqrt{1-x's'}}{\sqrt{1-s'}} + 2(s'-2)^2 + 2bs'\sqrt{\frac{1-x's'}{1-w's'}} - \frac{bs'^2x'}{2\sqrt{1-w's'}\sqrt{1-x's'}} + \frac{bs'^2w'\sqrt{1-x's'}}{2(1-w's')^{3/2}} \quad (50c)$$

and

$$B = 2x''s'\frac{\sqrt{1-s'}}{\sqrt{1-x's'}} - \frac{bs'^3x''}{2\sqrt{1-w's'}\sqrt{1-x's'}} + \frac{bs'^3w''\sqrt{1-x's'}}{2(1-w's')^{3/2}}. \quad (50d)$$

Obviously, from Eq. (50b) the solution for the imaginary part of  $s$  is

$$s'' = -\frac{B}{A}. \quad (51)$$

Equation (50a) is simply the characteristic equation that holds in the absence of attenuation. It may be solved numerically for the real root,  $s'$ , which lies between zero and unity, using a standard, iterative Newton–Raphson procedure. Once  $s'$  has been determined in this way, the imaginary part of  $s$  may be evaluated from Eq. (51), since all the terms in  $A$  and  $B$  are then known. Thus, Eqs. (50a) and (51) provide a means of solving for  $s'$  and  $s''$ , from which the interface wave speed and attenuation may be determined.

#### V. INTERFACE WAVE SPEED AND ATTENUATION

From Eq. (45a), the (complex) wave number of the interface wave is

$$p = \frac{k_s}{\sqrt{s}} = \frac{\omega}{c_o q_s \sqrt{s}}, \quad (52)$$

where  $q_s$  is given by Eq. (13b). Now,  $p$  can be expressed in the form

$$p = \frac{\omega}{c_i} (1 - j\beta_i), \quad (53)$$

where  $c_i$  and  $\beta_i$  are, respectively, the phase speed and the loss tangent of the interface wave. On comparing Eqs. (52) and (53), it follows that

$$\frac{1}{c_i} = \frac{1}{c_o} \operatorname{Re} \left[ \frac{1}{q_s \sqrt{s}} \right] = \operatorname{Re} \sqrt{\frac{\rho_o}{\mu_s s (j\omega t_1)^m}} \quad (54)$$

and

$$\beta_i = -\frac{c_i}{c_o} \operatorname{Im} \left[ \frac{1}{q_s \sqrt{s}} \right] = -c_i \operatorname{Im} \left[ \frac{\rho_o}{\mu_s s (j\omega t_1)^m} \right]^{1/2}. \quad (55)$$

As  $m \ll 1$  and  $s'' \ll s'$ , Eq. (54) may be approximated as follows:

$$c_i \approx \sqrt{\frac{\mu_s}{\rho_o}} s' \approx c_s \sqrt{s'}, \quad (56)$$

where  $c_s$  is the speed of the shear wave in the granular medium

$$\frac{1}{c_s} = \frac{1}{c_o} \operatorname{Re} \left[ \frac{1}{q_s} \right] \approx \sqrt{\frac{\rho_o}{\mu_s}}. \quad (57)$$

Similarly, the loss tangent in Eq. (55) approximates to

$$\beta_i \approx \frac{s''}{2s'} + \frac{m\pi}{4}, \quad (58)$$

and obviously the attenuation coefficient,  $\alpha_i$ , of the interface wave can be calculated from Eqs. (57) and (58), since

$$\alpha_i = \frac{|\omega| \beta_i}{c_i}. \quad (59)$$

## VI. GRAIN-SIZE DEPENDENCIES

It is evident from Eqs. (56)–(59) that the wave properties of the interface wave depend on  $\mu_c$  and  $\mu_s$ , both of which vary with the grain size. The functional forms of these grain-size dependencies have been developed by Buckingham<sup>1,2</sup> on the basis of the Hertz theory of elastic spheres pressed together in contact.<sup>17</sup> The properties of the interface wave also depend on  $c_0$ , the wave speed in the sediment in the limit of low frequency, and on the sediment density  $\rho_0$ . Both  $c_0$  and  $\rho_0$  are functions of the porosity,  $N$ , which itself depends on the grain size, as discussed by Buckingham on the basis of a random-packing model of rough spheres in contact. Since these grain-size dependencies have been established in the cited papers, they are not redeveloped here, instead, the results are simply stated.

The stress-relaxation rigidity moduli increase with grain size, as given by the following fractional-power relationships

$$\mu_c = \mu_0 \left( \frac{u_g}{u_0} \right)^{1/3} \quad (60)$$

and

$$\mu_s = \mu_1 \left( \frac{u_g}{u_0} \right)^{2/3}, \quad (61)$$

where  $\mu_0 = 2 \times 10^9$  Pa,  $\mu_1 = 5.1 \times 10^7$  Pa,  $u_0 = 1000 \mu\text{m}$ , and  $u_g$  is the mean grain diameter in micrometers. For coarse sand, with a grain diameter  $u_g \approx 700 \mu\text{m}$ , and for all finer marine sediments,  $\mu_c \gg \mu_s$ . As an example, a medium sand with  $u_g \approx 400 \mu\text{m}$  shows  $\mu_c \approx 1.5 \times 10^9$  Pa and  $\mu_s \approx 2.8 \times 10^7$  Pa, a difference of almost two orders of magnitude.

The porosity of the sediment, as given by the random-packing model of rough, spherical particles in contact is

$$N = 1 - P \left\{ \frac{u_g + 2\Delta}{u_g + 4\Delta} \right\}^3, \quad (62)$$

where  $P = 0.63$  is the packing factor of a random arrangement of smooth spheres,<sup>18</sup> and  $\Delta$  is the rms roughness of the grains, as measured about the mean surface of a particle. A value of  $\Delta = 3 \mu\text{m}$ , independent of grain size, provides a reasonable description of surficial, unimodal, sandy sediments.

The density,  $\rho_0$ , and bulk modulus,  $\kappa_0$ , of the sediment are given by the following weighted means:

$$\rho_0 = N\rho_w + (1-N)\rho_g \quad (63)$$

and

$$\frac{1}{\kappa_0} = N \frac{1}{\kappa_w} + (1-N) \frac{1}{\kappa_g}, \quad (64)$$

where the densities and bulk moduli of the constitutive mineral grains and seawater are:  $\rho_g = 2700 \text{ kg/m}^3$ ,  $\kappa_g = 1.47 \times 10^{10}$  Pa,  $\rho_w = 1024 \text{ kg/m}^3$ , and  $\kappa_w = 2.25 \times 10^9$  Pa. The sound speed in the absence of intergranular stress relaxation is therefore

$$c_0 = \sqrt{\frac{\kappa_0}{\rho_0}} = \sqrt{\frac{\kappa_w \kappa_g}{[N\rho_w + (1-N)\rho_g][N\kappa_g + (1-N)\kappa_w]}} \quad (65)$$

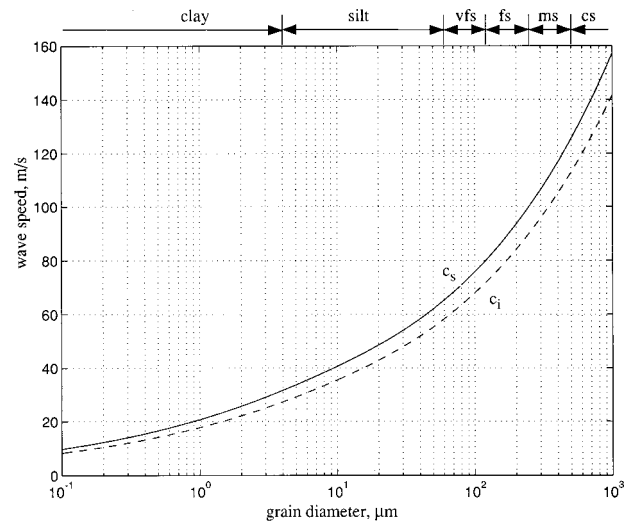


FIG. 2. Wave speeds as a function of grain size. The symbols at the top of the graph denote very fine sand (vfs), fine sand (fs), medium sand (ms), and coarse sand (cs).

which is just Wood's equation<sup>19</sup> for the sound speed in a two-phase medium. In effect,  $c_0$  is the sound speed in the sediment that would be observed if the medium were a suspension, with no interaction between the grains.

Based on Eqs. (56) and (57), in conjunction with the expressions given above, the grain-size dependencies of  $c_i$  and  $c_s$  are shown in Fig. 2, and the ratio of  $c_i$  to  $c_s$  as a function of the grain diameter is plotted in Fig. 3. It is evident from Fig. 3 that the ratio  $c_i/c_s$  depends only weakly on grain size, increasing by about 5%, from 85% to 90%, over the three decades of grain diameter between 1 and 1000 micrometers.

From the attenuation of shear waves in medium sand (with grain diameters between 350 and 380  $\mu\text{m}$ ), as measured under laboratory conditions by Brunson and Johnson<sup>7</sup> and Brunson,<sup>8</sup> Buckingham<sup>2</sup> has estimated that  $m \approx 0.05$ .

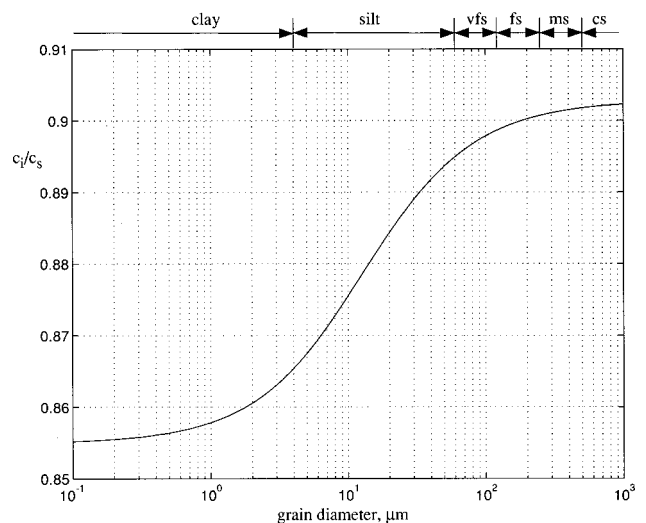


FIG. 3. Interface wave speed relative to shear wave speed as a function of grain size. The variation in the curve arises entirely from the density ratio,  $b$ , between the water column and the sediment, which varies with grain size. The key to the symbols at the top of the graph is given in the legend to Fig. 2.

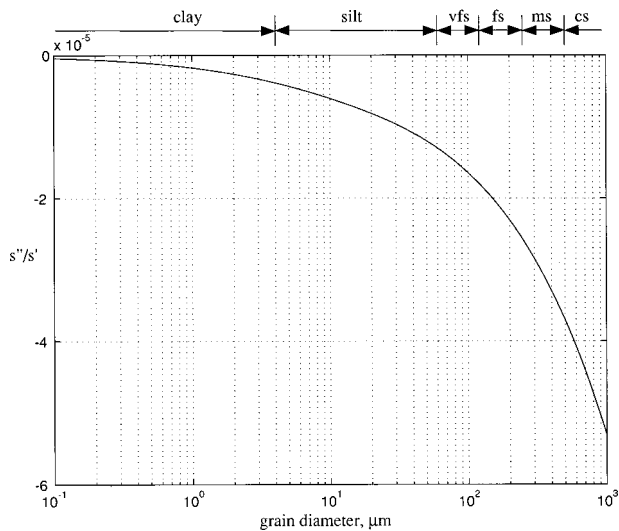


FIG. 4. Ratio of the imaginary to real part of  $s$  as a function of grain size. The key to the symbols at the top of the graph is given in the legend to Fig. 2.

With this value for the shear material response index  $m$ , it turns out, as shown in Fig. 4, that the ratio  $s''/2s'$  is negligible compared with the last term in Eq. (57); that is,

$$\beta_i \approx \frac{m\pi}{4}. \quad (66)$$

Evidently, to this degree of approximation, the loss tangent of the interface wave is equal to the loss tangent of the shear wave.<sup>2</sup> Both loss tangents scale with  $m$ , the material response index for shear, but are essentially independent of the remaining sediment parameters (e.g., the grain size, density, and porosity). With  $m=0.05$ , the value of the loss tangent for the pseudo-Scholte wave from Eq. (66) is  $\beta_i \approx 0.04$ . Incidentally, the value of the material response index for compressional waves is  $n \approx 0.05$ , as estimated by Buckingham<sup>1</sup> from comparison with attenuation data for unconsolidated marine sediments reported by Hamilton.<sup>5</sup> It would seem from these estimates that, tentatively, we have  $m \approx n$ .

The attenuation coefficient,  $\alpha_i$ , of the interface wave may now be computed directly from Eq. (59). Since  $\beta_i$  and  $c_i$  are independent of frequency, at least to the level of approximation in Eqs. (56) and (66), the attenuation coefficient of the interface wave is directly proportional to frequency. This is the same type of frequency dependence exhibited by the attenuation coefficient of the compressional wave and the shear wave. Note that  $\alpha_i$  depends on the grain size primarily through the phase speed  $c_i$ , which is a function of the stress-relaxation modulus  $\mu_s$ , as given in Eq. (61).

Figure 5 shows the attenuation coefficient,  $\alpha_i$ , as a function of grain size for a frequency of 1 Hz, computed using the expressions in Eqs. (56), (57), (59), and (66). The material response index has been taken as  $m=0.05$ , since this value is consistent with measurements of the shear attenuation in sands,<sup>1,2</sup> as discussed above. Note that the attenuation of the pseudo-Scholte wave in Fig. 5 is highest in the fine-grained materials and decays monotonically with increasing grain size, a dependence that arises almost entirely from the increase of phase speed,  $c_i$ , with grain size. This

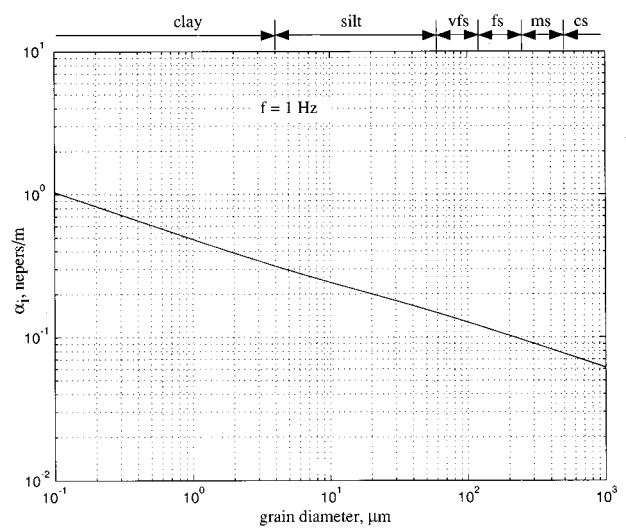


FIG. 5. Attenuation coefficient,  $\alpha_i$ , as a function of grain size. The key to the symbols at the top of the graph is given in the legend to Fig. 2.

increase in phase speed,  $c_i$ , and the reduction in attenuation,  $\alpha_i$ , with increasing grain size are indicative of the higher stiffness exhibited by the coarser-grained materials.

## VII. COMPARISON WITH INTERFACE-WAVE ATTENUATION DATA

A number of experiments on the interface wave propagating along the seawater–sediment boundary have been reported in the literature.<sup>20–34</sup> In many of these experiments, the frequency of the interface wave was well below 10 Hz, implying that the evanescent tail of the wave penetrated several tens of meters into the sediment, where the strong shear-speed gradient introduced significant dispersion. The presence of this type of dispersion makes it difficult to compare these measurements of wave speed and attenuation with the predictions of the theory [Eqs. (56) and (59)], since the latter apply to the case of a homogeneous, isotropic sediment in which the compressional speed and shear speed show no dispersion associated with wave speed profiles.

Two groups of investigators,<sup>20,29</sup> however, have reported observations of the phase speed and attenuation of interface waves with a frequency of 20 Hz or greater, where the effects of profile-related dispersion are rather less pronounced (see Fig. 4 in Ref. 29). Both groups used a line of seismometers on a sandy seabed to detect the signals generated by an explosive or vibrating source placed close to the end of the line. Propagation ranges were of the order of several meters. Table I lists some of the details of these experiments, including the frequency of the interface wave, the observed wave speed, and the measured attenuation coefficient, which has been converted to the equivalent loss tangent. Notice that the three measured loss tangents all agree, within the limits of experimental error, with the theoretical value,  $\beta_i \approx 0.04$ , derived above for the pseudo-Scholte wave.

In similar experiments conducted by Hamilton *et al.*,<sup>23</sup> broadband sources (electrical detonators) were used to generate interface waves in various types of marine sediment. No attenuation data for the interface waves were reported



TABLE I. Measured phase speed and attenuation of the interface wave in sandy sediments.

Investigators	Sediment type	Frequency (Hz)	Phase speed (m/s)	Attenuation (dB/m)	Loss tangent ( $\beta$ )
Bucker <i>et al.</i> (Ref. 20)	sand	20	90	$0.33 \pm 0.16$	$0.027 \pm 0.013$
Bucker <i>et al.</i> (Ref. 20)	sand	25	172	$0.2 \pm 0.13$	$0.025 \pm 0.017$
Holt <i>et al.</i> (Ref. 29)	sand	35	120	$0.6 \pm 0.2$	$0.038 \pm 0.012$

directly, but it was observed by these authors that “Most of the spectral energy of the Stoneley waves was at frequencies less than 100 Hz...” over propagation path lengths of the order of several meters. This limited bandwidth may be interpreted in terms of the expression for the attenuation of the pseudo-Scholte wave in Eq. (59), as discussed below.

Defining the cutoff in the detected signal by the  $e$ -folding frequency,  $f_e$ , at which the product of the attenuation coefficient,  $\alpha_i$ , and the range,  $r$ , is equal to unity, it follows from Eq. (59) that

$$f_c = \frac{c_i}{2\pi r \beta_i}. \quad (67)$$

At the Tower site used by Hamilton *et al.*,<sup>23</sup> the range between the source and the farthest receiver was  $r = 14.3$  m, and the sediment was medium sand with a measured interface wave speed of  $c_i = 170$  m/s. Thus, with  $\beta_i \approx 0.04$ , the  $e$ -folding frequency from Eq. (67) is  $f_e \approx 47.3$  Hz, which is consistent with the restricted bandwidths observed in the experiment. Similarly, for the other five sites investigated by Hamilton *et al.*,<sup>23</sup> where the maximum source–receiver range was 6.3 m, the  $e$ -folding frequencies are  $f_e = 56.2(89)$ ,  $59.4(94)$ ,  $48(76)$ ,  $84(133)$ , and  $77.7(123)$  Hz. The numbers in parentheses are the reported interface wave speeds in m/s. Again, these  $e$ -folding frequencies are all less than 100 Hz, suggesting that attenuation in the sediment was largely responsible for the restricted bandwidths observed in these experiments.

## VIII. CONCLUDING REMARKS

The characteristic equation for the interface (pseudo-Scholte) wave at the seawater–sediment boundary has been derived on the assumption that stress relaxation in the sediment arises at grain boundary contacts. The particular type of stress relaxation considered is characterized by material response functions of the form  $h(t) \propto t^{-n}$ , where  $n$  is positive but small compared with unity. Since the sediment is a dissipative medium, the required root of the characteristic equation is complex, and this root yields the phase speed and the attenuation of the interface wave. To solve for this complex root, an approximate technique is introduced in which the characteristic equation is split into two separate equations, one of which can be solved numerically for the real part of the root. Once the real part has been obtained, it may be substituted into the second equation, which then yields the imaginary part directly. An essential requirement of this technique is that the imaginary parts of all the complex variables in the characteristic equation must be very much smaller than the corresponding real parts, a condition that is easily satisfied by unconsolidated marine sediments.

The phase speed of the interface wave turns out to be weakly dependent on grain size, taking a value between 85% and 90% of the shear wave speed. In fact, the interface wave speed relative to the shear wave speed is the same as if the sediment were an elastic solid of the same density. In other words, Eq. (56) is identical to the relationship between the shear wave speed and the interface wave speed in an elastic medium. This is an important observation, because a number of authors, including Hamilton *et al.*,<sup>23</sup> have performed measurements of the interface wave speed from which they computed the shear wave speed on the assumption that the seabed was an elastic solid. Their results would have remained unchanged if they had used the expression in Eq. (56), derived from the stress-relaxation theory, to compute the shear wave speed. This is not to say that a marine sediment actually acts as an elastic solid. The elastic description fails on a number of counts, including the fact that it does not yield the grain-size dependencies exhibited by the compressional and shear wave speeds.<sup>4</sup>

The attenuation coefficient of the pseudo-Scholte wave that emerges from the stress-relaxation analysis is proportional to the first power of frequency. The corresponding loss tangent is thus independent of frequency, and its value turns out to be essentially the same as that predicted for the shear wave in the sediment. Little information on the attenuation of interface waves in marine sediments is available in the literature, at least for frequencies where dispersion due to a shear speed profile is negligible, but, within the limits of experimental error, the data that have been reported are in agreement with the theoretically derived attenuation coefficient. Moreover, the spectral content of interface wave arrivals on the seabed, as observed by Hamilton *et al.*,<sup>23</sup> is consistent with the predictions of the stress-relaxation theory.

<sup>1</sup>M. J. Buckingham, “Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments,” *J. Acoust. Soc. Am.* **102**, 2579–2596 (1997).

<sup>2</sup>M. J. Buckingham, “Theory of compressional and shear waves in fluid-like marine sediments,” *J. Acoust. Soc. Am.* **103**, 288–299 (1998).

<sup>3</sup>M. J. Buckingham, “Theory of compressional and transverse wave propagation in consolidated porous media,” *J. Acoust. Soc. Am.* **106**, 575–581 (1999).

<sup>4</sup>M. J. Buckingham, “Wave propagation, stress relaxation, and grain-boundary interactions in saturated, unconsolidated marine sediments,” *J. Acoust. Soc. Am.* (submitted).

<sup>5</sup>E. L. Hamilton, “Compressional-wave attenuation in marine sediments,” *Geophysics* **37**, 620–646 (1972).

<sup>6</sup>E. L. Hamilton, “Acoustic properties of sediments,” in *Acoustics and the Ocean Bottom* (Consejo Superior de Investigaciones Científicas, Madrid, (1987), pp. 3–58.

<sup>7</sup>B. A. Brunson and R. K. Johnson, “Laboratory measurements of shear wave attenuation in saturated sand,” *J. Acoust. Soc. Am.* **68**, 1371–1375 (1980).

<sup>8</sup>B. A. Brunson, “Shear wave attenuation in unconsolidated laboratory

- sediments," in *Shear Waves in Marine Sediments* (Kluwer, Dordrecht, 1991), pp. 141–147.
- <sup>9</sup>M. N. Toksöz and D. H. Johnston, *Seismic Wave Attenuation* (Society of Exploration Geophysicists, Tulsa, OK, 1981), p. 459.
- <sup>10</sup>W. L. Roever, T. F. Vining, and E. Strick, "Propagation of elastic wave motion from an impulsive source along a fluid/solid interface. I. Experimental pressure response. II. Theoretical pressure response. III. The pseudo-Rayleigh wave," *Philos. Trans. R. Soc. London, Ser. A* **251**, 455–523 (1959).
- <sup>11</sup>P. Nutting, "A study of elastic viscous deformation," *Proceedings of the American Society for Testing Materials XXI* (1921), pp. 1162–1171.
- <sup>12</sup>P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), p. 927.
- <sup>13</sup>P. M. Morse and H. Feshbach, *Methods of Theoretical Physics: Part I* (McGraw-Hill, New York, 1953), p. 997.
- <sup>14</sup>K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1975), p. 649.
- <sup>15</sup>L. M. Brekhovskikh, *Waves in Layered Media* (Academic, San Diego, 1980), p. 503.
- <sup>16</sup>J. G. Scholte, "Over het verband tussen zeegolven en microseismen, I and II," *Verslag. Ned. Akad. Wet.* **52**, 669–683 (1943).
- <sup>17</sup>S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity* (McGraw-Hill, New York, 1970), p. 567.
- <sup>18</sup>O. K. Rice, "On the statistical mechanics of liquids, and the gas of hard elastic spheres," *J. Chem. Phys.* **12**, 1–18 (1944).
- <sup>19</sup>A. B. Wood, *A Textbook of Sound* (G. Bell and Sons Ltd., London, 1964), p. 610.
- <sup>20</sup>H. P. Buckler, J. A. Whitney, and D. L. Keir, "Use of Stoneley waves to determine the shear velocity in ocean sediments," *J. Acoust. Soc. Am.* **36**, 1595–1596 (1964).
- <sup>21</sup>D. Davies, "Dispersed Stoneley waves on the ocean bottom," *Bull. Seismol. Soc. Am.* **55**, 903–918 (1965).
- <sup>22</sup>E. M. Herron, J. Dorman, and C. L. Drake, "Seismic study of the sediments in the Hudson River," *J. Geophys. Res.* **73**, 4701–4709 (1968).
- <sup>23</sup>E. L. Hamilton, H. P. Buckler, D. L. Keir, and J. A. Whitney, "Velocities of compressional and shear waves in marine sediments determined in situ from a research submersible," *J. Geophys. Res.* **75**, 4039–4049 (1970).
- <sup>24</sup>D. Rauch, "Experimental and theoretical studies of seismic interface waves in coastal waters," in *Bottom-Interacting Ocean Acoustics* (Plenum, New York, 1980), pp. 307–327.
- <sup>25</sup>F. Schirmer, "Experimental determination of properties of the Scholte wave in the bottom of the North Sea," in *Bottom-Interacting Ocean Acoustics* (Plenum, New York, 1980), pp. 285–298.
- <sup>26</sup>H. H. Essen, H. Janle, F. Schirmer, and J. Siebert, "Propagation of surface waves in marine sediments," *J. Geophys.* **49**, 115–122 (1981).
- <sup>27</sup>J. D. Tuthill, B. R. Lewis, and J. D. Garmany, "Stoneley waves, Lopez Island noise, and deep sea noise from 1 to 5 Hz," *Mar. Geophys. Res.* **5**, 95–108 (1981).
- <sup>28</sup>T. M. Brocher, B. T. Iwatake, and D. A. Lindwall, "Experimental studies of low-frequency waterborne and sediment-borne acoustic wave propagation on a continental shelf," *J. Acoust. Soc. Am.* **74**, 960–972 (1983).
- <sup>29</sup>R. M. Holt, J. M. Hovem, and J. Syrstad, "Shear modulus profiling of near bottom sediments using boundary waves," in *Acoustics and the Sea-Bed* (Bath University Press, Bath, 1983), pp. 317–325.
- <sup>30</sup>D. Rauch and B. Schmalfeldt, "Ocean-bottom interface waves of the Stoneley/Scholte type: properties, observations and possible use," in *Acoustics and the Sea-Bed* (Bath University Press, Bath, 1983), pp. 307–316.
- <sup>31</sup>B. Schmalfeldt and D. Rauch, *Explosion-Generated Seismic Interface Waves in Shallow Water: Experimental Results*, SACLANT ASW Research Centre, La Spezia, Report SR-71 (1983).
- <sup>32</sup>D. Rauch, "On the role of bottom interface waves in ocean seismo-acoustics: a review," in *Ocean Seismo-Acoustics* (Plenum, New York, 1986), pp. 623–641.
- <sup>33</sup>D. M. F. Chapman and P. R. Staal, "A summary of DREA observations of interface waves at the seabed," in *Shear Waves in Marine Sediments* (Kluwer, Dordrecht, 1991), pp. 177–184.
- <sup>34</sup>G. Nolet and L. M. Dorman, "Waveform analysis of Scholte modes in ocean sediment layers," *Geophys. J. Int.* **125**, 385–396 (1996).

# Applications of waveguide-type base functions for the eigenproblems of two-dimensional cavities

Yoon Young Kim and Dong Kyun Kim

Department of Mechanical Design, Seoul National University, Shinlim-Dong, Kwanak-Gu, Seoul 151-742, Korea

(Received 13 July 1998; accepted for publication 21 May 1999)

An acoustic modal analysis technique useful for two-dimensional cavities of general shapes is developed as an extension of the technique developed earlier by Kim and Kang [J. Acoust. Soc. Am. **99**, 2938–2946 (1996)] for the modal analysis of convex membranes. Both Neumann and Dirichlet problems can be solved with the present choice of cosinusoidally varying waveguide-type functions instead of sinusoidally varying waveguide-type functions adopted earlier [J. Acoust. Soc. Am. **99**, 2938–2946 (1996)]. A cavity subdivision technique for general nonconvex cavities overcomes the limitation of the earlier technique useful only for convex cavities. Successful applications of the present technique are demonstrated in some practical two-dimensional cavity problems. The present technique is potentially a very efficient technique in handling higher eigenmodes. © 1999 Acoustical Society of America. [S0001-4966(99)01109-1]

PACS numbers: 43.20.Ks [AN]

## INTRODUCTION

A number of methods including direct numerical methods, such as the finite difference method<sup>1,2</sup> and the finite element method,<sup>3</sup> have been developed for the modal analysis of an acoustic cavity. Other methods have also been proposed, such as the collocation,<sup>4,5</sup> point-matching,<sup>6</sup> Rayleigh–Ritz,<sup>7</sup> and Galerkin<sup>8</sup> methods. More references on related subjects may be found in Kim and Kang.<sup>9</sup> Although most of these methods can be applied for the analysis of general shape cavities, some difficulties in terms of numerical efficiency and solution accuracy usually arise in the medium to high frequency range.

One of the present authors has recently proposed a new approach<sup>9</sup> in an effort to develop a method that may be potentially more efficient and accurate in the prediction of up to considerably higher eigenfrequencies. This approach, developed for the modal analysis of simply-supported convex polygonal membranes, uses waveguide-type base functions that satisfy the governing differential equation exactly. The waveguide-type base functions are such that they vary sinusoidally along each side of the membrane and propagate into its interior domain. The limitation of this method is that only the membranes of convex geometry with the Neumann-type boundary conditions can be analyzed.

The present work is an extension of the earlier work by Kim and Kang,<sup>9</sup> but two major contributions are: (1) the analysis of general nonconvex, or multiply-connected cavities, and (2) the analysis of both the Neumann- and Dirichlet-type boundary value problems.

In order to handle general boundary conditions, the formulation based on cosinusoidally varying base functions is suggested instead of sinusoidally varying base functions that are mainly useful for the Dirichlet-type problems. In the present application of the cosine-type base functions, the corresponding zeroth harmonic components do not cause any of the dependence problems observed in some static problems (Kim and Yoon<sup>10</sup>); thus both the Neumann- and Dirichlet-

type problems can be equally handled by the present cosine-type base function formulation. Because of this property, a matrix representing acoustic impedance can be formed, which plays an essential role in the analysis of general nonconvex or multiply-connected cavities.

For the analysis of nonconvex or multiply-connected cavities, the waveguide-type base function approach developed for convex cavities cannot be directly used because of the unidirectional nature of the waveguide-type base functions. Subsequently, the cavity domain is divided into a few large convex domains. The interface continuity conditions are imposed utilizing a matrix associated with acoustic impedance. This continuity matching process results in an assembly procedure similar to that used in the finite element method.

Several numerical examples are considered to demonstrate the validity of the present technique. Although direct computation time comparison is difficult at the present development stage, the fact that a relatively small number of base functions are needed to obtain satisfactory results is addressed. Interesting and important characteristics of the matrix associated with acoustic impedance are also re-

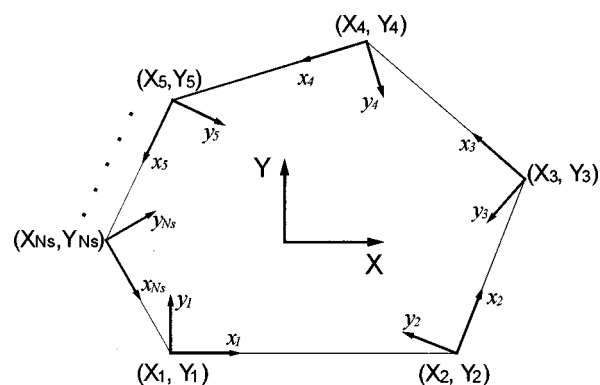


FIG. 1. A convex polygonal acoustic cavity. The local coordinates  $(x_j, y_j)$  have the origin at the vertex whose global coordinates are  $(X_j, Y_j)$ .

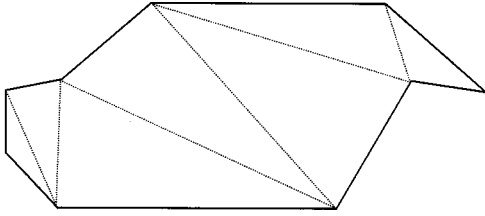


FIG. 2. A typical subdivision of a nonconvex cavity into a few convex subcavities.

marked, which can be utilized in the future research for the analysis of medium to high eigenfrequencies.

## I. WAVEGUIDE-TYPE BASE FUNCTION APPROACH

### A. Convex polygonal cavity

In this section, we first discuss a waveguide-type approach for two-dimensional convex cavities. Then we extend it to the analysis of general nonconvex cavities. The underlying analysis for convex cavities is similar to that given by Kim and Kang.<sup>9</sup> However, the following formulation is based on cosine-type base functions, and may require some care in dealing with the terms related to the zeroth harmonic of cosinusoidally varying base functions.

The governing equation for acoustic pressure  $p(x, y, t)$  in a two-dimensional cavity is

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} = \frac{1}{c^2} \cdot \frac{\partial^2 p}{\partial t^2}. \quad (1)$$

In Eq. (1),  $x$  and  $y$  are Cartesian coordinates and  $c$  is the speed of sound. If the acoustic pressure  $p$  and the particle velocity  $v$  are time-harmonically varying at the angular frequency  $\omega$  as

$$p(x, y, t) = P(x, y) \cdot e^{-i\omega t}, \quad (2a)$$

$$v(x, y, t) = V(x, y) \cdot e^{-i\omega t}, \quad (2b)$$

the well-known Helmholtz equation is obtained:

$$\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + k^2 P = 0. \quad (3)$$

The wave number  $k$  used in Eq. (3) is related to  $\omega$  as  $k = \omega/c$ .

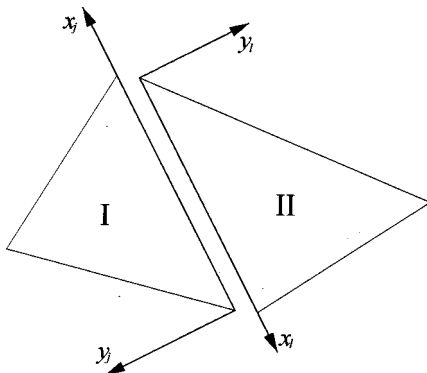


FIG. 3. The interface of two adjacent subcavities is shown (note the different orientations of the local coordinate systems of the subcavities).

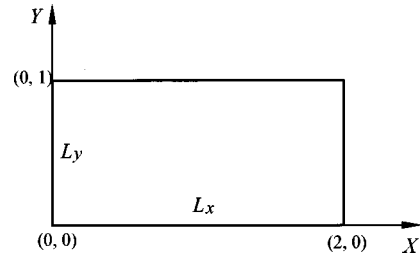


FIG. 4. The geometry of a rectangular acoustic cavity.

Since both the Dirichlet-type ( $p$  prescribed) and the Neumann-type ( $\partial p/\partial n$  or  $v$  prescribed) problems in a convex cavity of Fig. 1 must be solved, the following cosine series expansion of the boundary conditions is introduced:

$$\bar{P}^{(j)}(x_j, y_j=0) \cdot L_j = \frac{\bar{P}_0^{(j)}}{2} + \sum_{m=1}^{N_h} \bar{P}_m^{(j)} \cos \frac{m\pi x_j}{L_j} \quad (j=1, 2, \dots, N_s), \quad (4)$$

$$\bar{V}^{(j)}(x_j, y_j=0) = \sum_{m=0}^{N_h} \bar{V}_m^{(j)} \cos \frac{m\pi x_j}{L_j} \quad (j=1, 2, \dots, N_s), \quad (5)$$

where  $\bar{P}^{(j)}(x_j)$  and  $\bar{V}^{(j)}(x_j)$ , which exclude the harmonic-time dependence, are the pressure and velocity distributions that can be prescribed along each side of the convex polygonal cavity. The particle velocity  $\bar{V}^{(j)}(x_j, y_j=0)$  can be found from  $P$  as,

$$\bar{V}^{(j)}|_{y_j=0} = \frac{1}{i\rho_0\omega} \left. \frac{\partial P}{\partial y_j} \right|_{y_j=0}. \quad (6)$$

The relation between the local coordinates  $(x_j, y_j)$  and the global coordinates  $(X, Y)$  is straightforward. The length of each edge is denoted by  $L_j$  and the highest term in the truncated cosine series is denoted by  $N_h$ . The cosine series coefficients of the acoustic pressure and the particle velocity along each edge are defined as  $\bar{P}_m^{(j)}$  and  $\bar{V}_m^{(j)}$ , respectively. Note that in the definition of  $\bar{P}_m^{(j)}$ , the edge length  $L_j$  is multiplied so that the left-hand side term in Eq. (4) actually

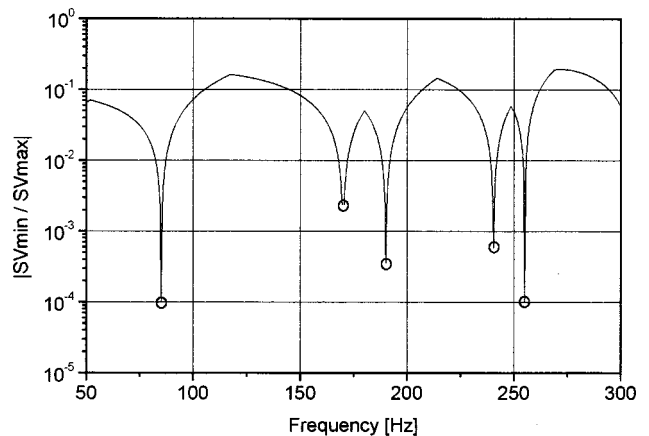


FIG. 5. The plot for the ratio of the minimum singular value to the maximum singular value as a function of frequency.



TABLE I. The comparison of the present results with the ANSYS finite element and the exact solutions for the eigenfrequencies (unit: Hz). ( $N_h$ : the highest harmonic term in each subcavity,  $N_{el}$ : the number of finite elements used in ANSYS.)

Mode No.	Present			ANSYS			Exact
	$N_h=2$	$N_h=4$	$N_h=8$	$N_{el}=50$	$N_{el}=200$	$N_{el}=800$	
1st	85.00	85.00	85.00	85.35	85.09	85.02	85.00
2nd	170.00	170.00	170.00	172.81	170.70	170.17	170.00
3rd	170.00	170.00	170.00	172.81	170.70	170.17	170.00
4th	190.07	190.07	190.07	192.74	190.73	190.23	190.07
5th	240.42	240.42	240.42	244.39	241.41	240.66	240.42
6th	255.00	255.00	255.00	264.51	257.37	255.59	255.00

represents the total force acting on each edge.

Once the boundary conditions are given in the form of Eqs. (4), (5), general solutions to such a form of the boundary conditions may be constructed in the following manner. First, we treat the convex cavity having  $N_s$  sides (shown in Fig. 1) as the union of the  $N_s$  semi-infinite cavities whose boundaries are parallel to each of the polygonal cavity edges. Namely, the total solution  $P(x,y)$  of the polygonal cavity may be constructed from the superposition of  $P^{(l)}(x_l, y_l)$  associated with the  $l$ th semi-infinite cavity:

$$P(X, Y) = \sum_{l=1}^{N_s} P^{(l)}(x_l, y_l), \quad (7)$$

where

$$P^{(l)}(x_l, y_l) = \sum_{n=1}^{N_h} A_n^{(l)} \cos \frac{n\pi x_l}{L_l} \exp(ik_n^{(l)} y_l). \quad (8)$$

Although Eq. (8) should be replaced by an infinite series, the truncated series will be used in numerical calculation. The definition of  $k_n^{(l)}$  in Eq. (8) is

$$k_n^{(l)} = \sqrt{(\omega/c)^2 - (n\pi/L_l)^2}. \quad (9)$$

Note that  $P^{(l)}(x_l, y_l)$ , which consists of  $N_h$  base functions, represents unidirectional waveguide solutions that satisfy the governing differential equation (3) exactly. The base function represents either a propagating wave for  $\omega > \omega_c$  or an evanescent wave if otherwise. The cutoff frequency  $\omega_c$  is given by  $\omega_c = cn\pi/L_l$  ( $n=1, 2, \dots$ ).

Following the procedure described in Kim and Kang,<sup>9</sup> the relation between the participation factors  $A_n^{(l)}$  and  $(\bar{P}_m^{(j)}, \bar{V}_m^{(j)})$  can be written as:

$$\bar{\mathbf{P}} = \mathbf{PM} \cdot \mathbf{A}, \quad (10a)$$

$$\bar{\mathbf{V}} = \mathbf{VM} \cdot \mathbf{A}, \quad (10b)$$

where

$$\bar{\mathbf{P}} = \begin{Bmatrix} \bar{\mathbf{P}}^{(1)} \\ \bar{\mathbf{P}}^{(2)} \\ \vdots \\ \bar{\mathbf{P}}^{(N_s)} \end{Bmatrix}; \quad (11a)$$

$$\bar{\mathbf{V}} = \begin{Bmatrix} \bar{\mathbf{V}}^{(1)} \\ \bar{\mathbf{V}}^{(2)} \\ \vdots \\ \bar{\mathbf{V}}^{(N_s)} \end{Bmatrix}; \quad (11b)$$

$$\mathbf{A} = \begin{Bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(N_s)} \end{Bmatrix} \quad (11c)$$

and

$$\bar{\mathbf{P}}^{(j)} = \begin{Bmatrix} \bar{P}_0^{(j)} \\ \bar{P}_1^{(j)} \\ \vdots \\ \bar{P}_{N_h}^{(j)} \end{Bmatrix}; \quad (12a)$$

$$\bar{\mathbf{V}}^{(j)} = \begin{Bmatrix} \bar{V}_0^{(j)} \\ \bar{V}_1^{(j)} \\ \vdots \\ \bar{V}_{N_h}^{(j)} \end{Bmatrix}; \quad (12b)$$

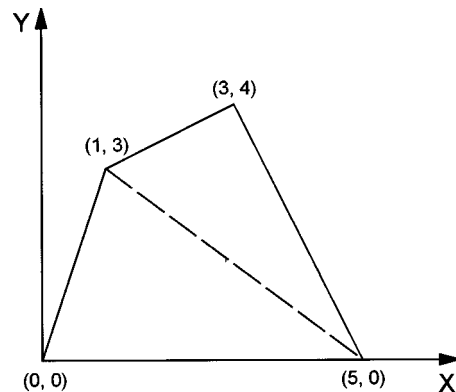


FIG. 6. The geometry of a quadrilateral acoustic cavity. The vertex locations are in parentheses.

TABLE II. The first five eigenfrequencies (unit: Hz) for the cavity shown in Fig. 6 with the condition  $p=0$  along the boundary.

Mode No.	Present			ANSYS			
	$N_h=2$	$N_h=4$	$N_h=8$	$N_{el}=25$	$N_{el}=100$	$N_{el}=400$	$N_{el}=1600$
1st	69.94	69.99	69.99	71.27	70.32	70.13	70.01
2nd	107.50	107.75	107.76	114.14	109.39	108.31	107.86
3rd	109.39	109.96	109.99	116.42	111.61	110.45	110.09
4th	N/A	142.94	143.01	154.80	146.22	144.11	143.23
5th	N/A	146.57	146.70	165.21	151.51	147.99	147.00

$$\mathbf{A}^{(j)} = \begin{Bmatrix} A_0^{(j)} \\ A_1^{(j)} \\ \vdots \\ A_{N_h}^{(j)} \end{Bmatrix}. \quad (12c)$$

In Eq. (12),  $\bar{\mathbf{P}}^{(j)}$ ,  $\bar{\mathbf{V}}^{(j)}$ , and  $\mathbf{A}^{(j)}$  denote the arrays composed of the Fourier coefficients of the pressure, the particle velocity and the participation factor, respectively. The coefficients  $PM_{(m,n)}^{(j,l)}$  and  $VM_{(m,n)}^{(j,l)}$  can be obtained in closed forms from the following formula:

$$PM_{(m,n)}^{(j,l)} = \frac{a}{L_j} \int_0^{L_j} \left[ \cos \frac{n\pi x_l}{L_l} \exp(ik_n^{(l)} y_l) \right]_{y_j=0} \cdot \cos \frac{m\pi x_j}{L_j} dx_j$$

( $a=1$  for  $n=0$ ;  $a=2$  for  $n \neq 0$ ),

(13)

$$VM_{(m,n)}^{(j,l)} = \frac{a}{i\rho_0\omega L_j} \int_0^{L_j} \frac{\partial}{\partial y_j} \left[ \cos \frac{n\pi x_l}{L_l} \exp(ik_n^{(l)} y_l) \right]_{y_j=0} \cdot \cos \frac{m\pi x_j}{L_j} dx_j$$

( $a=1$  for  $n=0$ ;  $a=2$  for  $n \neq 0$ ).

(14)

If the eigenfrequencies of the convex cavity with zero pressure along the boundary ( $\bar{\mathbf{P}}=0$ ) are to be found, the solution  $\omega$  of the following equation needs to be solved:

$$\text{Det}[\mathbf{PM}(\omega)] = 0. \quad (15a)$$

Similarly for hard wall conditions, one uses

$$\text{Det}[\mathbf{VM}(\omega)] = 0. \quad (15b)$$

Note that if the sine series based formulation were utilized, only the Dirichlet-type problems could be handled. Kim and Yoon<sup>10</sup> have addressed an issue that the cosine-based formulation may cause some functional dependence problems in the case of static problems. However, no such problems arise in the present application as nonvanishing independent plane waves can be generated for the base functions associated with the zeroth harmonic term. In addition to this capability of handling both the Dirichlet- and Neumann-type problems, the use of the cosine-based formulation allows handling of general nonconvex, or multiply-connected cavities.

## B. Nonconvex cavity

The method described in the preceding section is applicable only to a simple convex acoustic cavity. This is because the selected base functions are unidirectional so that they propagate only into the corresponding semi-infinite cavity domain. To handle nonconvex or multiply-connected cavities, we propose to divide the cavity into a number of convex subcavities as indicated in Fig. 2; for the analysis of each subcavity, the technique discussed in the previous section is applied.

To use the result obtained in the previous section, the following relation between  $\bar{\mathbf{V}}$  and  $\bar{\mathbf{P}}$  defined along the boundary of each convex subcavity may be constructed. The relation will be utilized in order to satisfy the continuity relation along the subcavity boundaries as shall be seen later:

$$\bar{\mathbf{V}} = \mathbf{AM} \cdot \bar{\mathbf{P}}. \quad (16)$$

In Eq. (16), the matrix  $\mathbf{AM}$  is simply

$$\mathbf{AM} = \mathbf{VM} \cdot \mathbf{PM}^{-1}. \quad (17)$$

Physically this matrix  $\mathbf{AM}$  is associated with a (sub-)cavity impedance relation.

TABLE III. The first five eigenfrequencies (unit: Hz) for the cavity shown in Fig. 6 with the condition  $v=0$  along the boundary.

Mode No.	Present			ANSYS			
	$N_h=2$	$N_h=4$	$N_h=8$	$N_{el}=25$	$N_{el}=100$	$N_{el}=400$	$N_{el}=1600$
1st	43.182	43.180	43.179	43.738	43.323	43.215	43.188
2nd	47.313	47.295	47.291	48.034	47.478	47.337	47.302
3rd	77.419	77.308	77.304	79.540	77.885	77.451	77.340
4th	80.410	80.401	80.400	84.479	81.443	80.662	80.465
5th	92.481	91.666	91.654	97.608	93.145	92.025	91.746

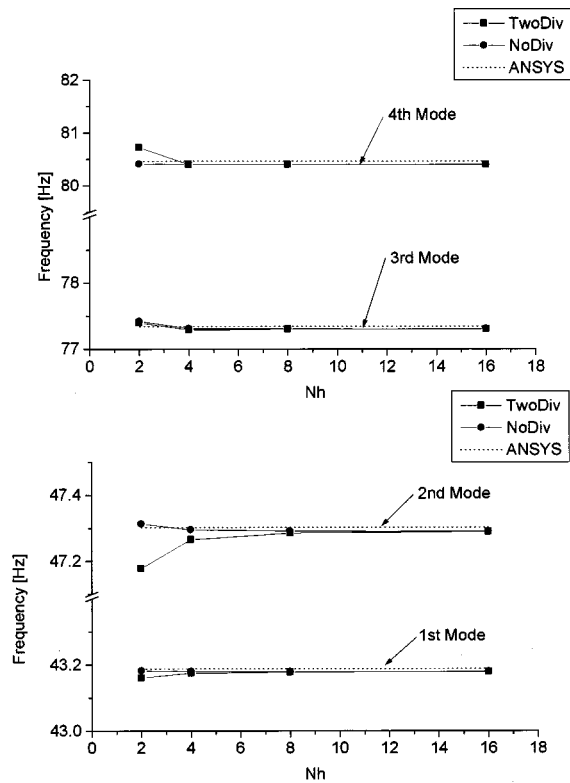


FIG. 7. The convergence rate of the present method for the cavity shown in Fig. 6 with a  $v=0$  boundary condition.

Once a cavity is divided into subcavities, the continuity conditions for normal velocity and pressure along the sub-cavity interface boundaries must be imposed:

$${}^I P = {}^{II} P, \quad {}^I V = {}^{II} V, \quad (18)$$

where superscripts  $I$  and  $II$  refer to the two adjacent subcavities. Since the present formulation is based on Fourier series-type functions, the corresponding Fourier coefficients along the boundary of two adjacent subcavities (see Fig. 3) may be defined in the following series expansions:

$$\begin{aligned} {}^I P(x_i) &= \sum_{m=0}^{N_h} {}^I P_m \cos \frac{m \pi x_i}{L} \quad (0 \leq x_i \leq L), \\ {}^{II} P(x_k) &= \sum_{n=0}^{N_h} {}^{II} P_n \cos \frac{m \pi x_k}{L} \quad (0 \leq x_k \leq L), \\ {}^I V(x_i) &= \sum_{m=0}^{N_h} {}^I V_m \cos \frac{m \pi x_i}{L} \quad (0 \leq x_i \leq L), \end{aligned} \quad (19)$$

TABLE IV. The results obtained from different methods for the first five eigenfrequencies (unit: Hz) in the cavity with  $v=0$  along the boundary.

Mode No.	Two division				No division				ANSYS $N_{el}=1600$
	$N_h=2$	$N_h=4$	$N_h=8$	$N_h=16$	$N_h=2$	$N_h=4$	$N_h=8$	$N_h=16$	
1st	43.161	43.175	43.178	43.179	43.182	43.180	43.179	43.179	43.188
2nd	47.178	47.265	47.285	47.289	47.313	47.295	47.291	47.290	47.302
3rd	77.395	77.289	77.300	77.303	77.419	77.308	77.304	77.303	77.340
4th	80.721	80.402	80.400	80.400	80.410	80.401	80.400	80.400	80.465
5th	90.645	91.617	91.647	91.652	92.481	91.666	91.654	91.653	91.746

$${}^{II} V(x_k) = \sum_{n=0}^{N_h} {}^{II} V_n \cos \frac{m \pi x_k}{L} \quad (0 \leq x_k \leq L).$$

In terms of the Fourier cosine series coefficients, the interface continuity conditions are written as:

$${}^I P_m = (-1)^m \cdot {}^{II} P_m \quad (m=0,1,2,\dots), \quad (20)$$

$${}^I V_m = (-1)^{m+1} \cdot {}^{II} V_m \quad (m=0,1,2,\dots). \quad (21)$$

It is convenient to introduce the symbol “ $e$ ” to designate quantities associated with each subcavity for subsequent analysis. With this notation, Eq. (16) is rewritten as

$$\bar{V}_e = \mathbf{A} \mathbf{M}_e \cdot \bar{P}_e. \quad (22)$$

Then the final system of equations for the assembled total cavity may be put in matrix form, which takes into account the interface continuity

$$\bar{V}_T = \mathbf{A} \mathbf{M}_T \cdot \bar{P}_T, \quad (23)$$

where

$$\bar{V}_T = \sum_{e=1}^{N_{el}} \bar{V}_e; \quad \bar{P}_T = \sum_{e=1}^{N_{el}} \bar{P}_e; \quad \mathbf{A} \mathbf{M}_T = \sum_{e=1}^{N_{el}} \mathbf{A} \tilde{\mathbf{M}}_e. \quad (24)$$

In Eqs. (23), (24), the subscript  $T$  stands for the total cavity and  $N_{el}$  is the number of convex subcavities. The assembling procedure used in Eq. (24) is very similar to that used in the finite element method so that detailed procedures are omitted here. However, the matrix  $\mathbf{A} \tilde{\mathbf{M}}_e$  is a modified matrix from the matrix  $\mathbf{A} \mathbf{M}_e$  in order to take into account the sign change stated in Eqs. (20), (21).

In the case of the cavity with a hard wall, the eigenfrequencies can be calculated from

$$\text{Det}[\mathbf{A} \mathbf{M}_T] = 0. \quad (25)$$

For the numerical computation in search of the eigenfrequencies, the singular value decomposition technique is used. For other types of boundary conditions, a similar procedure can be used.

## II. NUMERICAL EXAMPLES

### A. Rectangular cavity

First, we consider a rectangular cavity with a hard wall boundary shown in Fig. 4. This problem has been selected to confirm that the present technique gives the correct answers for this simple problem. The well-known exact eigenfrequencies are

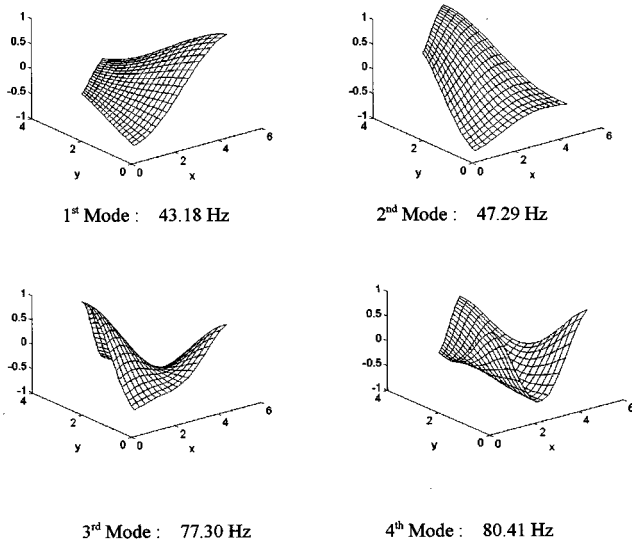


FIG. 8. The first four eigenmodes for the cavity shown in Fig. 6 with a  $v = 0$  boundary condition.

$$f = \frac{c}{2} \sqrt{\left(\frac{m}{L_x}\right)^2 + \left(\frac{n}{L_y}\right)^2} \quad (m, n = 0, 1, 2, \dots).$$

For all the calculations throughout the present work, the sound speed is taken to be  $c = 340$  m/s. The ratio ( $|SV_{\min}/SV_{\max}|$ ) of the minimum singular value to the maximum singular value of the matrix  $\mathbf{VM}$ , which is obtained by the present method with  $N_h = 4$ , is plotted in Fig. 5 as a function of frequency. The eigenfrequencies that make the ratio  $|SV_{\min}/SV_{\max}|$  zero or very small are marked by circles in Fig. 5. For comparison, Table I lists the exact, present and ANSYS finite element results. It is apparent that the present technique, even with a very few number of terms, yields very accurate and rapidly converging results.

### B. Quadrilateral cavity

As the next example, a quadrilateral cavity shown in Fig. 6 is considered. In this example for which no exact solution is possible, the modal analysis of the cavity with and without the cavity subdivision will be carried out. Although this cavity is convex and there is no need to subdivide the cavity into subcavities, it is useful to examine the validity of the present subdivision technique.

TABLE V. MAC values for the first five modes obtained from the present ( $N_h = 4$ ) and ANSYS results.

		ANSYS				
		1st	2nd	3rd	4th	5th
Present	1st	1.000	6.4090E-4	0.0160	0.0018	8.1379E-5
	2nd	5.349E-4	1.0000	0.0026	0.0011	0.0295
	3rd	0.0149	0.0028	0.9989	0.0078	0.0025
	4th	0.0017	0.0011	0.0127	0.9997	0.0087
	5th	7.2978E-5	0.0297	0.0033	0.0089	0.9998

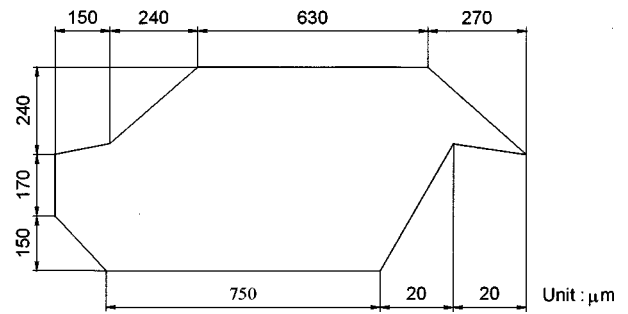


FIG. 9. A half-scaled vehicle model. The cavity inside the model is assumed to be two-dimensional (unit: mm).

### 1. Without cavity division

First, the present results without any divisions are compared with the ANSYS finite element results in Table II for the Dirichlet, and in Table III for the Neumann problems. Note that excellent results are obtained with a very small number of terms.

### 2. With cavity division into two subcavities

Now the same quadrilateral cavity in Fig. 6 is divided into two subcavities in order to demonstrate the effectiveness of the subdivision technique. In this case, two subcavities consist of the triangles connecting the vertices  $[(0,0), (5,0), (1,3)]$  and  $[(1,3), (5,0), (3,4)]$ . Since the results obtained for both the Dirichlet and Neumann problems are almost of the same accuracy, we just list the results for the Neumann problems.

As shown in Table IV, the present results obtained with a two-subcavity division agree well with other results, which demonstrates the validity of the present subdivision technique. The convergence rate of the present method with (denoted by “Two Div”) and without (“No Div”) is illustrated in Fig. 7.

The first four mode shapes obtained by the present technique are shown in Fig. 8. To see how good the present mode shapes are, the modal assurance criterion (MAC)<sup>11</sup> values between the present and ANSYS results are listed in Table V. Recalling the well-known definition of the MAC value,

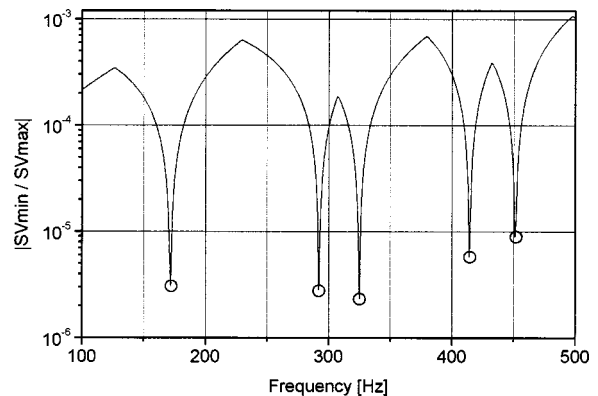


FIG. 10. The magnitude of the minimum singular value to the maximum singular value for the matrix  $\mathbf{AM}_T$ .



TABLE VI. The eigenfrequencies (unit: Hz) obtained by the present method for a half-scaled cavity. The present results are compared with the experiment and finite element results.

Mode No.	Present			ANSYS			Experiment (Ref. 12)
	$N_h=2$	$N_h=4$	$N_h=8$	$N_{el}=94$	$N_{el}=362$	$N_{el}=1152$	
1st (1,0)	171.9	171.7	171.6	172.6	172.2	172.2	180.5
2nd (2,0)	291.9	292.2	292.3	296.3	294.4	294.0	301.1
3rd (0,1)	324.8	324.8	324.8	327.9	325.6	325.1	321.5
4th (3,0)	414.3	414.2	414.4	422.3	417.2	416.2	425.4
5th (2,1)	450.4	451.4	451.5	458.2	453.2	452.3	457.8

$$\text{MAC}(p,A) = \frac{|\sum_{m=1}^{N_x} \sum_{n=1}^{N_y} (P_{\text{present}})_{(x_m, y_n)} (P_{\text{present}})_{(x_m, y_n)}|^2}{(\sum_{m=1}^{N_x} \sum_{n=1}^{N_y} (P_{\text{present}})_{(x_m, y_n)}^2) (\sum_{m=1}^{N_x} \sum_{n=1}^{N_y} (P_{\text{ANSYS}})_{(x_m, y_n)}^2)}$$

one can see an excellent coherence between the present and ANSYS mode shapes.

### C. Half-scaled vehicle model

As a practical application of the present approach, the eigenfrequencies of the cavity of a half-scaled vehicle model<sup>12</sup> are calculated. The geometry of the half-scaled vehicle model is shown in Fig. 9 and the present numerical analysis was conducted with seven subcavities suggested in Fig. 2. The ratio  $|SV_{\min}/SV_{\max}|$  of the assembled matrix  $\mathbf{AM}_T$  as a function of  $\omega$  (with  $N_h=4$ ) is plotted in Fig. 10. The eigenfrequencies are marked by circles in this figure.

The present results with different numbers of terms ( $N_h=2, 4,$  and  $8$ ) are compared against the ANSYS finite element and experimental results:<sup>12</sup> see Table VI. The experiment was conducted for the three-dimensional half-scaled cavity model. The depth of the model, which is not shown in Fig. 8, is 680 mm. The experiment results of Table VI include only the eigenfrequencies corresponding to the

two-dimensional cavity acoustic modes. Again, the present results agree well with the finite element results.

### III. REMARKS

One interesting and useful observation is that the form of the matrix  $\mathbf{AM}$  becomes diagonally dominant as the frequency  $f$  becomes larger. To illustrate this, we plot the magnitude of  $\mathbf{AM}$  in a gray scale in Fig. 11 for various frequencies. We simply select  $\mathbf{AM}$  associated with the upper triangular subdomain in Fig. 6 for the results shown in Fig. 11.

As the frequency becomes larger, dominant terms are limited only for lower indices  $i, j$  and higher diagonal elements. A similar behavior was also reported in the study of elastic wave propagation problems of solid circular cylinders.<sup>13,14</sup> This characteristic has not been incorporated in the present study, but can be potentially very useful in extending the present approach in dealing with eigenproblems in the medium to high frequency range.

It is noted that for the subdivision method, only triangular subcavities were used: see, for instance, Fig. 2. Although other convex shapes such as quadrilaterals may be used, the most useful subcavity shapes are triangular. This is partly because of its convenience in subdivision. Another reason is that the problem of the functional dependence of base functions can be easily overcome as long as the triangles are not isosceles. (See Kim and Kang<sup>9</sup> for more detailed accounts.)

### IV. CONCLUSIONS

A general procedure based on waveguide-type base functions was proposed to handle the eigenproblems of non-convex, or multiply connected two-dimensional cavities. The advantage of this approach is that satisfactory results were obtained with only a few number of base functions. The property that the system matrix strongly shows the diagonal dominance as more terms are used will be potentially very useful in dealing with very high eigenmodes. Subsequent work in this direction may result in the development of a new, very efficient solution technique for the analysis of general two- or three-dimensional cavities.

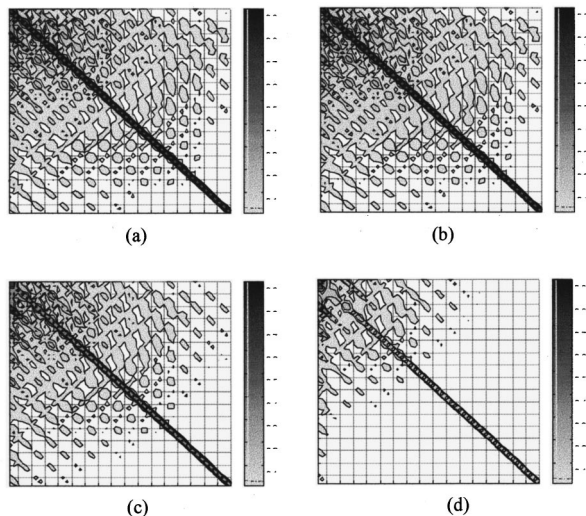


FIG. 11. The magnitude of  $|\mathbf{AM}_{ij}|/\sqrt{|\mathbf{AM}_{ii}||\mathbf{AM}_{jj}|}$  (no sum on  $i, j$ ) for (a)  $f=10$  Hz, (b)  $f=20$  Hz, (c)  $f=60$  Hz and (d)  $f=80$  Hz.

- <sup>1</sup>J. B. Davies and C. A. Muilwyk, "Numerical solutions of uniform hollow waveguides with boundaries of arbitrary shape," Proc. Inst. Electr. Eng. **113**, 277–284 (1966).
- <sup>2</sup>C. W. Steele, "Numerical computation of electric and magnetic field in a uniform waveguide of arbitrary cross-section," J. Comp. Phys. **3**, 48–153 (1968).
- <sup>3</sup>P. L. Arlett, A. K. Bahrani, and O. C. Zienkiewicz, "Application of finite elements to the solution of Helmholtz's equation," Proc. Inst. Electr. Eng. **115**, 1762–1766 (1968).
- <sup>4</sup>H. Y. Yee, "On determination of cutoff frequencies of waveguides with arbitrary cross-section," Proc. IEEE **54**, 64 (1965).
- <sup>5</sup>H. Y. Yee and M. F. Audeh, "Cutoff frequencies of eccentric waveguides," IEEE Trans. Microwave Theory Tech. **MTT-14**, 487–493 (1966).
- <sup>6</sup>R. H. T. Bates, "The theory of point-matching method for perfectly conducting wavelengths and transmission lines," IEEE Trans. Microwave Theory Tech. **MTT-17**, 294–301 (1969).
- <sup>7</sup>R. M. Bulley and J. B. Davies, "Computation of approximate polynomial solutions to TE modes in an arbitrary shaped waveguide," IEEE Trans. Microwave Theory Tech. **MTT-17**, 440–446 (1969).
- <sup>8</sup>P. A. A. Laura, "A simple method for calculation of cut-off frequencies of waveguides with arbitrary cross-section," Proc. IEEE **54**, 1495–1497 (1966).
- <sup>9</sup>Y. Y. Kim and J. H. Kang, "Free vibration analysis of membranes using waveguide-type base functions," J. Acoust. Soc. Am. **99**, 2938–2946 (1996).
- <sup>10</sup>Y. Y. Kim and M. S. Yoon, "A modified Fourier series method for the torsion analysis of bars with multiply-connected cross sections," Int. J. Solids Struct. **34**, 4327–4337 (1997).
- <sup>11</sup>D. J. Ewins, *Modal Testing: Theory and Practice* (Research Studies Press, Great Britain, 1984).
- <sup>12</sup>J. M. Lee, S. W. Kang, and S. H. Kim, "The Acoustic-structural Coupling Analysis of the Passenger Cavity Considering the Characteristic of a Roof/Airgap/Trim," Proceedings of the Korean Society for Noise Vibration Engineering, **6**, 84–90 (1996).
- <sup>13</sup>Y. Y. Kim and C. R. Steele, "End effects and time-harmonic longitudinal wave propagation in a semi-infinite solid cylinder," ASME J. Appl. Mech. **56**, 334–346 (1989).
- <sup>14</sup>Y. Y. Kim and C. R. Steele, "An analytic-asymptotic approach for time-harmonic nonsymmetric wave propagation in a cylinder," Int. J. Solids Struct. **26**, 1143–1157 (1990).

# Nonlinear Stoneley and Scholte waves

G. D. Meegan, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya

Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063

(Received 25 November 1998; accepted for publication 31 March 1999)

Spectral evolution equations are derived for plane, progressive, finite-amplitude Stoneley and Scholte waves that propagate along plane interfaces formed by two semi-infinite, isotropic media in contact. The evolution equations have mathematical forms identical to those obtained previously for Rayleigh waves [E. A. Zabolotskaya, *J. Acoust. Soc. Am.* **91**, 2569–2575 (1992)], and they are expressed explicitly in terms of the second- and third-order elastic constants of the media. Calculations were performed to simulate nonlinear surface wave propagation in several pairs of real media. Harmonic generation and shock formation associated with the Stoneley and Scholte modes are compared with the corresponding processes in Rayleigh waves. Waveform distortion is shown to be very similar for the three types of surface waves when the propagation distance is normalized by an appropriate shock formation distance. © 1999 Acoustical Society of America. [S0001-4966(99)02507-2]

PACS numbers: 43.25.Lj, 43.25.Ba [HEB]

## INTRODUCTION

The three fundamental types of surface acoustic waves in isotropic media are Rayleigh waves<sup>1</sup> (solid–vacuum interfaces), Stoneley waves<sup>2</sup> (solid–solid interfaces), and Scholte waves<sup>3</sup> (solid–fluid interfaces). Because they are localized at the interfaces, surface waves spread cylindrically, and therefore decay less rapidly with distance, than spherically spreading bulk compressional and shear waves. Moreover, surface waves are more easily observed and detected than bulk waves in the interior of a solid. At sufficiently large amplitudes, surface waves distort and develop shocks, as demonstrated recently with laser-generated Rayleigh waves in fused quartz.<sup>4,5</sup> The waveform distortion resembles, but possesses features that distinguish it from, the distortion of finite-amplitude sound in fluids.<sup>6</sup> A review of theoretical models for nonlinear Rayleigh waves is provided by Knight *et al.*<sup>7</sup>

The Hamiltonian formalism used in the present investigation to model nonlinear Stoneley and Scholte waves was developed originally by Zabolotskaya for nonlinear Rayleigh waves.<sup>8</sup> Her spectral evolution equations provide an accurate description of the finite-amplitude waveform distortion and shock formation observed in experiments.<sup>5</sup> The Hamiltonian formalism has been extended to other types of surface waves, e.g., in crystals<sup>9</sup> (for which experimental confirmation of the theory has also been obtained<sup>10</sup>) and in piezoelectric materials.<sup>11</sup> For the case of Rayleigh wave propagation in isotropic solids, the model has been used to describe harmonic generation in plane and cylindrical waves<sup>12</sup> and in diffracting beams,<sup>13</sup> for the investigation of stationary waves,<sup>14</sup> local and nonlocal nonlinearity,<sup>15</sup> the development of time domain evolution equations,<sup>16</sup> and for pulsed nonlinear Rayleigh waves.<sup>17</sup> The model equations developed in the present article for nonlinear Stoneley and Scholte waves are found to have the same mathematical properties (i.e., symmetries of the nonlinearity matrix) as those for nonlinear Rayleigh waves.

## I. LINEAR THEORY

Here we review equations describing the propagation of small-signal elastic surface waves along a planar interface between two half-spaces. The half-spaces consist of two homogeneous, isotropic media in contact at  $z=0$ , as shown in Fig. 1, where  $\rho_i$  is the mass density in the  $i$ th medium ( $i=1,2$ ),  $c_{li}$  is the corresponding longitudinal (compressional) wave speed, and  $c_{ti}$  the transverse (shear) wave speed. Medium 1 is taken to be a solid. The surface wave is known as a Rayleigh wave when medium 2 is a vacuum, a Stoneley wave<sup>2</sup> when it is a solid, and a Scholte wave<sup>3</sup> when it is a fluid.

Plane wave propagation in the  $+x$  direction is considered, with the  $z$  axis taken positive upward, into half-space 2. Small-signal effects such as thermoviscous absorption and geometrical spreading are not taken into account, as they can be included *ad hoc* in a relatively straightforward way. The particle displacement vector  $u_j^{(i)}$  ( $j=x,z$ ) in the  $i$ th medium satisfies the linear wave equation

$$\rho_i \frac{\partial^2 u_j^{(i)}}{\partial t^2} = c_{jklm}^{(i)} \frac{\partial^2 u_l^{(i)}}{\partial x_k \partial x_m}, \quad (1)$$

where  $c_{jklm}^{(i)}$  are the stiffness constants, only two of which are independent for an isotropic solid, and only one for a fluid. The solutions of Eq. (1) that are presented below must also satisfy the boundary conditions on displacement and stress appropriate for the given surface wave, and the condition that the field vanish as  $|z| \rightarrow \infty$ . The stress tensor  $\sigma_{jk}^{(i)}$  is related to the particle displacement by Hooke's law:

$$\sigma_{jk}^{(i)} = c_{jklm}^{(i)} \frac{\partial u_l^{(i)}}{\partial x_m}. \quad (2)$$

The notation and solutions presented below provide the basis for the nonlinear theory developed in Sec. II.

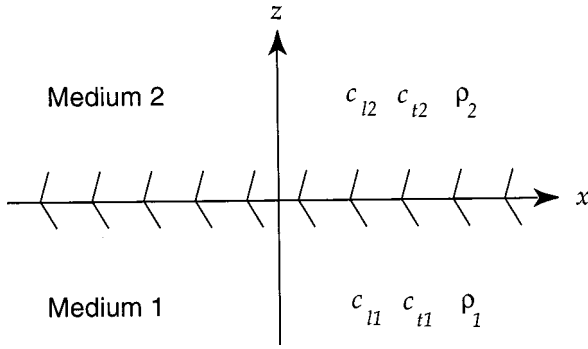


FIG. 1. Two half-spaces in contact at  $z=0$ .

### A. Stoneley waves

The following four boundary conditions are required for continuity of displacement and stress at the interface:

$$u_j^{(1)}|_{z=0} = u_j^{(2)}|_{z=0}, \quad j = x, z, \quad (3)$$

$$\sigma_{zk}^{(1)}|_{z=0} = \sigma_{zk}^{(2)}|_{z=0}, \quad k = x, z. \quad (4)$$

The surface wave solution of Eqs. (1)–(4) at angular frequency  $\omega$  may be expressed in the form

$$\left. \begin{aligned} u_x^{(1)} &= i(\xi_{t1} e^{\xi_{t1} kz} + \eta_1 e^{\xi_{l1} kz}) u_0 e^{i(kx - \omega t)} \\ u_z^{(1)} &= (e^{\xi_{t1} kz} + \eta_1 \xi_{l1} e^{\xi_{l1} kz}) u_0 e^{i(kx - \omega t)} \end{aligned} \right\} z < 0, \quad (5)$$

$$\left. \begin{aligned} u_x^{(2)} &= i(\xi_{t2} e^{-\xi_{t2} kz} + \eta_2 e^{-\xi_{l2} kz}) \theta u_0 e^{i(kx - \omega t)} \\ u_z^{(2)} &= -(e^{-\xi_{t2} kz} + \eta_2 \xi_{l2} e^{-\xi_{l2} kz}) \theta u_0 e^{i(kx - \omega t)} \end{aligned} \right\} z > 0, \quad (6)$$

where the coefficients  $\eta_1$ ,  $\eta_2$ , and  $\theta$  are defined in Appendix A,  $k = \omega/c$  is the wave number,  $c$  is the phase speed of the surface wave, and the depth penetration coefficients are

$$\xi_{li} = \sqrt{1 - c^2/c_{li}^2}, \quad \xi_{ti} = \sqrt{1 - c^2/c_{ti}^2}. \quad (7)$$

The characteristic equation that determines  $c$  is<sup>2</sup>

$$\begin{aligned} &\xi_1^2 \xi_2^2 [(m-1)^2 - (m\xi_{l1} + \xi_{l2})(m\xi_{t1} + \xi_{t2})] \\ &+ 4(m\xi_1^2 - \xi_2^2)(m\xi_{l1}\xi_{t1} - \xi_{l2}\xi_{t2} - m + 1) \\ &+ 4(m\xi_1^2 - \xi_2^2)^2(\xi_{l1}\xi_{t1} - 1)(\xi_{l2}\xi_{t2} - 1) = 0, \end{aligned} \quad (8)$$

where  $\xi_i = c/c_{ti}$ , and  $m = \rho_2/\rho_1$  is the density ratio. The amplitude  $u_0$  is the sole arbitrary coefficient in Eqs. (5) and (6).

The characteristic equation for the phase speed  $c$  of Stoneley waves is independent of  $\omega$  and therefore Stoneley waves, like Rayleigh waves, are nondispersive. Unlike Rayleigh waves, which exist in any isotropic solid, Stoneley waves are not supported by all material pairs. For any given six material parameters that define the interface (namely, the

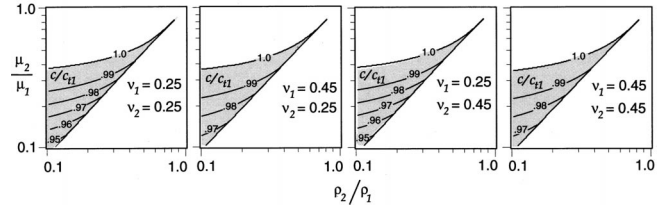


FIG. 2. Parameter spaces in which Stoneley waves exist.

density and two elastic constants for each solid), a real solution of Eq. (8) does not necessarily exist whose value is less than the bulk wave speeds in both media (as required for the wave to be trapped at the interface). Several researchers have studied both analytically and numerically the existence of real roots of Eq. (8).<sup>3,18–22</sup> For example, Owen<sup>21</sup> considered over 900 pairs of metals, out of which he identified only 30 that support Stoneley waves.

In order to identify conditions for the existence of Stoneley waves, real roots of the characteristic equation were calculated numerically for a range of material parameters. The results are shown in Fig. 2, where  $\nu_1$  and  $\nu_2$  are the Poisson ratios of the lower and upper solids, respectively. We have followed the convention, commonly used in the literature,<sup>22</sup> of showing where Stoneley waves exist in the parameter space defined by the density ratio  $\rho_2/\rho_1$  (horizontal axes) and the shear modulus ratio  $\mu_2/\mu_1$  (vertical axes). Stoneley waves exist only for parameters associated with the shaded regions, with the corresponding normalized propagation speed  $c/c_{t1}$  (relative to the propagation speed of a shear wave in the lower medium) given by the contours. A list of the specific elastic solids considered is given in Table I, where representative material properties are collected from the literature.<sup>23–27</sup> All combinations of the materials in Table I that support Stoneley waves are listed in Table II. Note that the Stoneley wave speed is always slightly less than the smaller of the shear wave speeds in the two solids.

A schematic representation of a Stoneley wave is provided in Fig. 3. The particle trajectory in the lower solid depicted here is elliptical and changes direction just below the interface. The particle trajectory in the upper solid is also elliptical but does not change direction at any depth. The trajectory changes direction in only one of the solids.

### B. Scholte waves

Here we assume the upper half-space is an ideal fluid. We thus require continuity of only the normal components of

TABLE I. Representative properties of several elastic solids.

Medium	$\rho$ (kg/m <sup>3</sup> )	$c_l$ (m/s)	$c_t$ (m/s)	$\mathcal{A}$ (GPa)	$\mathcal{B}$ (GPa)	$\mathcal{C}$ (GPa)
Aluminum	2700	5919	2981	–320	–200	–190
Copper	8900	4627	2223	–1600	1400	–1600
Nickel	8800	5746	3015	–730	–230	180
Steel	7700	5939	3283	–760	–250	–90
Glass	2300	5637	3297	420	–120	130
Sandstone	2270	2780	1770	–620	–2400	–1000



TABLE II. Several elastic solid pairs that support Stoneley waves.

Medium 1	Medium 2	$c$ (m/s)
Nickel	Aluminum	2959
Nickel	Glass	2983
Steel	Glass	3224

the displacement and stress at the interface. The shear stress in the solid must vanish at the interface, and we have

$$u_z^{(1)}|_{z=0} = u_z^{(2)}|_{z=0}, \quad (9)$$

$$\sigma_{zz}^{(1)}|_{z=0} = \sigma_{zz}^{(2)}|_{z=0}, \quad \sigma_{zx}^{(1)}|_{z=0} = 0. \quad (10)$$

The solution in this case may be written

$$\left. \begin{aligned} u_x^{(1)} &= i(\xi_{t1} e^{\xi_{t1} kz} + \eta_1 e^{\xi_{l1} kz}) u_0 e^{i(kx - \omega t)} \\ u_z^{(1)} &= (e^{\xi_{t1} kz} + \eta_1 \xi_{l1} e^{\xi_{l1} kz}) u_0 e^{i(kx - \omega t)} \end{aligned} \right\} z < 0, \quad (11)$$

$$\left. \begin{aligned} u_x^{(2)} &= i \eta_2 e^{-\xi_{l2} kz} u_0 e^{i(kx - \omega t)} \\ u_z^{(2)} &= -\eta_2 \xi_{l2} e^{-\xi_{l2} kz} u_0 e^{i(kx - \omega t)} \end{aligned} \right\} z > 0, \quad (12)$$

where  $\eta_1$  and  $\eta_2$  are defined in Appendix B,  $\xi_{li}$  and  $\xi_{ti}$  are defined by Eqs. (7), and the small-signal Scholte wave speed  $c = \omega/k$  is determined by the characteristic equation<sup>3</sup>

$$4 \xi_{l1} \xi_{t1} - (1 + \xi_{t1}^2)^2 = m(1 - \xi_{t1}^2)^2 \xi_{l1} / \xi_{l2}, \quad (13)$$

where again  $m = \rho_2 / \rho_1$ .

Scholte waves, like Rayleigh and Stoneley waves, are also nondispersive. Numerical solutions for the real roots of Eq. (13) are presented in Fig. 4, where  $\nu_1 = 0.25$  is the Poisson ratio of the solid. A real root corresponding to a Scholte wave always exists. The dependence of the propagation speed on the Poisson ratio is very weak, varying only a few percent for  $0 < \nu_1 < \frac{1}{2}$ . Solutions of the characteristic equation were also obtained for interfaces formed by water ( $\rho_2 = 1000 \text{ kg/m}^3$ ,  $c_{l2} = 1481 \text{ m/s}$ ) above each of the solids listed in Table I. These results are summarized in Table III, where except for the Berea sandstone, the Scholte wave speed is within 1% of the speed of sound in the water.

Schematic representation of the Scholte wave is provided in Fig. 5. The amplitude of the particle motion drops off exponentially with distance in both directions away from the interface. Note that the horizontal ( $x$ ) component of the

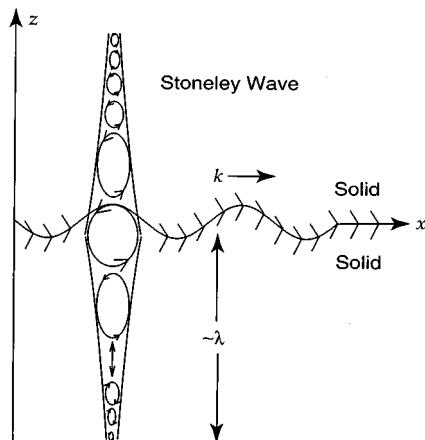


FIG. 3. Schematic representation of particle trajectories in a Stoneley wave.

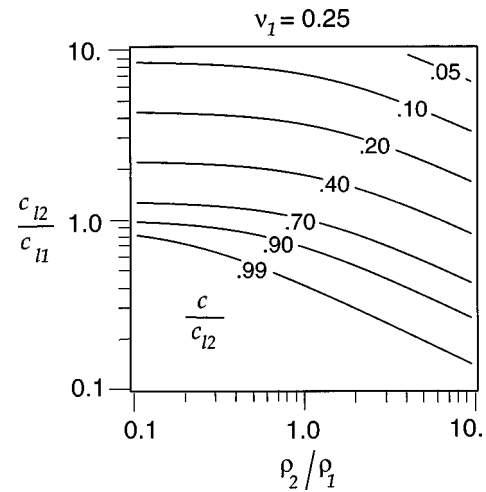


FIG. 4. Parameter space in which Scholte waves exist for  $\nu_1 = 0.25$ .

particle displacement is discontinuous across the traction-free interface. Specifically, Eqs. (11) and (12) yield

$$\left. \frac{u_x^{(2)}}{u_x^{(1)}} \right|_{z=0} = \frac{\eta_2}{\xi_{t1} + \eta_1} = \frac{(1 - \xi_{t1}^2) \xi_{l1}}{(1 + \xi_{t1}^2 - 2 \xi_{t1} \xi_{l1}) \xi_{l2}}. \quad (14)$$

From the first of Eqs. (7) we see that  $\xi_{l2} \rightarrow 0$  as  $c \rightarrow c_{l2}$ , in which case the ratio in Eq. (14) becomes very large. For an aluminum–water interface, the value of this ratio is 42. We shall return to this point in our discussion of nonlinear effects in Sec. III B.

### C. Rayleigh waves

The Scholte wave solution applies to Rayleigh waves in the limit where the upper fluid becomes a vacuum,  $m \rightarrow 0$ . All stresses in the solid must vanish at the interface:

$$\sigma_{zk}^{(1)}|_{z=0} = 0, \quad k = x, z. \quad (15)$$

The particle displacements are given by

$$\left. \begin{aligned} u_x^{(1)} &= i(\xi_{t1} e^{\xi_{t1} kz} + \eta_1 e^{\xi_{l1} kz}) u_0 e^{i(kx - \omega t)} \\ u_z^{(1)} &= (e^{\xi_{t1} kz} + \eta_1 \xi_{l1} e^{\xi_{l1} kz}) u_0 e^{i(kx - \omega t)} \end{aligned} \right\} z < 0, \quad (16)$$

where  $\eta_1$  is defined in Appendix B, with  $\xi_{l1}$  and  $\xi_{t1}$  given again by Eqs. (7). Equation (13) reduces for  $m=0$  to the characteristic equation for the propagation speed of Rayleigh waves:

$$4 \xi_{l1} \xi_{t1} = (1 + \xi_{t1}^2)^2, \quad (17)$$

the square of which yields the familiar result<sup>1</sup>

$$\xi_1^6 - 8 \xi_1^4 + (24 - 16q^2) \xi_1^2 - 16(1 - q^2) = 0, \quad (18)$$

where  $\xi_1 = c/c_{t1}$  and  $q = c_{t1}/c_{l1}$ .

TABLE III. Scholte wave speeds for water over several solids.

Medium	$c$ (m/s)
Aluminum	1475
Nickel	1478
Steel	1480.6
Glass (Pyrex)	1472
Dry Berea Sandstone	1318

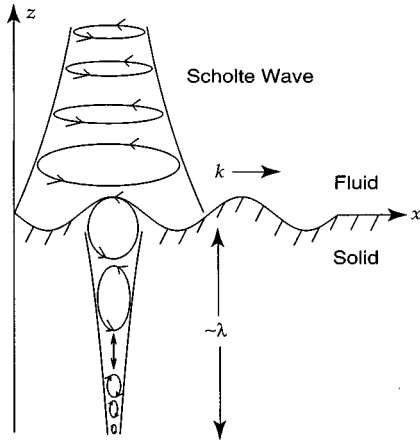


FIG. 5. Schematic representation of particle trajectories in a Scholte wave.

## II. NONLINEAR THEORY

The approach we follow to develop mathematical models for nonlinear Stoneley and Scholte waves is a direct extension of the method used previously to investigate nonlinear Rayleigh waves.<sup>8</sup> In this approach, the solution of the nonlinear surface wave problem is assumed to be similar to the linear solution insofar as the depth dependence of each frequency component of the wave is the same as in the linear approximation. Hamiltonian formalism is then used to obtain a set of equations describing the spectral evolution of nonlinear surface waves.

We assume that the nonlinear evolution of a surface wave may be described by a harmonic summation of linear surface wave modes, but augmented with amplitudes that vary arbitrarily in time:

$$u_j^{(i)} = \sum_{n=-\infty}^{\infty} a_n(t) u_{jn}^{(i)}(z) e^{inkx}, \quad j=x,z. \quad (19)$$

The depth functions  $u_{jn}^{(i)}(z)$ , following multiplication by  $e^{in(kx-\omega t)}$ , are obtained as solutions of Eq. (1), or equivalently

$$c_{jklm}^{(i)} \frac{\partial^2 (u_{ln}^{(i)} e^{inkx})}{\partial x_k \partial x_m} = -n^2 \omega^2 \rho_i u_{jn}^{(i)} e^{inkx}, \quad (20)$$

that satisfy the relevant linear boundary conditions at  $z=0$ : Eqs. (3) and (4) for Stoneley waves, Eqs. (9) and (10) for Scholte waves, or Eq. (15) for Rayleigh waves. The time dependent functions  $a_n(t)$ , which are as yet unknown, must therefore reduce to the corresponding expressions in linear theory when the small-signal limit is taken. In particular, for progressive wave propagation in the  $+x$  direction we must have  $a_n(t)$  proportional to  $e^{-in\omega t}$  in the small-signal limit. Any deviation from this dependence results exclusively from nonlinearity.

The solutions derived below are not expected to satisfy completely the second-order boundary conditions for Stoneley, Scholte, or Rayleigh waves, nor are they even expected to satisfy completely the equations of motion at second order. For any solution to satisfy completely all boundary conditions and equations of motion at second order, the interaction of surface waves with bulk compressional and shear

wave modes of propagation must be taken into account explicitly. Provided that bulk waves constitute a negligible component of the field structure, the solutions derived below on the basis of Eq. (19) constitute the principal second-order contribution to the surface wave mode of propagation. Further discussion of this matter may be found in Appendix C.

### A. Energy densities

The nonlinear evolution equations follow from the Hamiltonian

$$H = T + V, \quad (21)$$

where  $T$  is the kinetic and  $V$  the potential energy density of the wave. Expressions for  $T$  and  $V$  are derived separately for Stoneley and Scholte waves. Lagrangian coordinates are used in the formulation.

#### 1. Stoneley waves

The expressions for  $u_{jn}^{(i)}(z)$  in Eqs. (19) are taken here to be

$$\left. \begin{aligned} u_{xn}^{(1)}(z) &= i \frac{n}{|n|} \left( \xi_{11} e^{|n|k\xi_{11}z} + \eta_1 e^{|n|k\xi_{11}z} \right) \\ u_{zn}^{(1)}(z) &= e^{|n|k\xi_{11}z} + \eta_1 \xi_{11} e^{|n|k\xi_{11}z} \end{aligned} \right\} z < 0, \quad (22)$$

$$\left. \begin{aligned} u_{xn}^{(2)}(z) &= i \frac{n}{|n|} \theta \left( \xi_{12} e^{-|n|k\xi_{12}z} + \eta_2 e^{-|n|k\xi_{12}z} \right) \\ u_{zn}^{(2)}(z) &= -\theta \left( e^{-|n|k\xi_{12}z} + \eta_2 \xi_{12} e^{-|n|k\xi_{12}z} \right) \end{aligned} \right\} z > 0. \quad (23)$$

The dependence on  $z$  is thus the same as in the linear solutions given by Eqs. (5) and (6), but modified to describe a wave having arbitrary harmonic number  $n$  with respect to a reference (fundamental) frequency  $\omega$ . We introduced the factor  $n/|n| = \text{sgn } n$  in Eqs. (22) and (23) so that the particle motion at the interface is retrograde for both positive and negative harmonic number  $n$ . In the limit of small signals, in which case  $a_n(t)$  is time harmonic, Eqs. (22) and (23) may be substituted into Eqs. (19) to obtain a linear superposition of harmonically related solutions that are equivalent to Eqs. (5) and (6).

The kinetic energy (per unit area) is

$$T = \frac{\rho_1}{2\lambda} \int_0^\lambda dx \int_{-\infty}^0 [(\dot{u}_x^{(1)})^2 + (\dot{u}_z^{(1)})^2] dz + \frac{\rho_2}{2\lambda} \int_0^\lambda dx \int_0^\infty [(\dot{u}_x^{(2)})^2 + (\dot{u}_z^{(2)})^2] dz, \quad (24)$$

where  $\lambda = 2\pi/k$  is the wavelength at the fundamental frequency in the Fourier expansion, and the dots indicate time derivatives. Substitution of Eqs. (19), (22), and (23) yields

$$T = \frac{\rho_1 \zeta}{4k} \sum_n \frac{\dot{a}_n \dot{a}_{-n}}{|n|}, \quad (25)$$

where  $\zeta$  is defined in Appendix A.

The elastic (strain) energy density, through third order in the displacement, may be written

$$\mathcal{E} = \mathcal{E}_2 + \mathcal{E}_3, \quad (26)$$

where  $\mathcal{E}_2$  is quadratic in the displacement and  $\mathcal{E}_3$  is cubic:<sup>28</sup>

$$\mathcal{E}_2 = \frac{\mu}{4} \left( \frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right)^2 + \left( \frac{K}{2} - \frac{\mu}{3} \right) \left( \frac{\partial u_j}{\partial x_j} \right)^2, \quad (27)$$

$$\begin{aligned} \mathcal{E}_3 = & \left( \mu + \frac{\mathcal{A}}{4} \right) \frac{\partial u_i}{\partial x_k} \frac{\partial u_j}{\partial x_i} \frac{\partial u_j}{\partial x_k} + \left( \frac{\mathcal{B}}{2} + \frac{K}{2} - \frac{\mu}{3} \right) \frac{\partial u_j}{\partial x_j} \left( \frac{\partial u_i}{\partial x_k} \right)^2 \\ & + \frac{\mathcal{A}}{12} \frac{\partial u_i}{\partial x_k} \frac{\partial u_k}{\partial x_j} \frac{\partial u_j}{\partial x_i} + \frac{\mathcal{B}}{2} \frac{\partial u_i}{\partial x_k} \frac{\partial u_k}{\partial x_i} \frac{\partial u_j}{\partial x_j} + \frac{\mathcal{C}}{3} \left( \frac{\partial u_j}{\partial x_j} \right)^3. \end{aligned} \quad (28)$$

Here,  $K$  is the bulk modulus,  $\mu$  is again the shear modulus, and  $\mathcal{A}$ ,  $\mathcal{B}$ , and  $\mathcal{C}$  are Landau's third-order elastic constants.

The potential energy density corresponding to the quadratic terms is thus

$$V_2 = \frac{1}{\lambda} \int_0^\lambda dx \int_{-\infty}^0 \mathcal{E}_2^{(1)} dz + \frac{1}{\lambda} \int_0^\lambda dx \int_0^\infty \mathcal{E}_2^{(2)} dz, \quad (29)$$

where  $\mathcal{E}_2^{(i)}$  is the elastic energy in the  $i$ th half-space. This expression is a generalization of Eq. (14) of Ref. 8, and we therefore have

$$\begin{aligned} V_2 = & \frac{1}{2\lambda} u_j^{(1)} \sigma_{jz}^{(1)} \Big|_{-\infty}^0 - \frac{1}{2\lambda} \int_0^\lambda dx \int_{-\infty}^0 u_j^{(1)} \frac{\partial \sigma_{jk}^{(1)}}{\partial x_k} dz \\ & + \frac{1}{2\lambda} u_j^{(2)} \sigma_{jz}^{(2)} \Big|_0^\infty - \frac{1}{2\lambda} \int_0^\lambda dx \int_0^\infty u_j^{(2)} \frac{\partial \sigma_{jk}^{(2)}}{\partial x_k} dz. \end{aligned} \quad (30)$$

First-order relations may be used to evaluate individual terms in Eq. (30). Boundary conditions given by Eqs. (3) and (4), together with the requirement that the disturbance vanish at  $|z| = \infty$ , eliminate the first and third terms. The divergence of the stress tensor may be rewritten at first order, from Eqs. (2) and (20), as

$$\frac{\partial \sigma_{jk}^{(i)}}{\partial x_k} = -\rho_i \omega^2 \sum_{n=-\infty}^{\infty} n^2 a_n(t) u_{jn}^{(i)}(z) e^{inkx}. \quad (31)$$

Equation (31) permits the integrals in Eq. (30) to be evaluated, giving

$$V_2 = \frac{\rho_1 \zeta}{4k} \omega^2 \sum_n |n| a_n a_{-n}. \quad (32)$$

The contribution to the potential energy due to cubic terms is

$$V_3 = \frac{1}{\lambda} \int_0^\lambda dx \int_{-\infty}^0 \mathcal{E}_3^{(1)} dz + \frac{1}{\lambda} \int_0^\lambda dx \int_0^\infty \mathcal{E}_3^{(2)} dz. \quad (33)$$

This calculation is performed analytically via direct substitution of Eq. (28) to obtain

$$V_3 = \mu_1 k^2 \sum_{l+m+n=0} w_{lmn} a_l a_m a_n, \quad (34)$$

where

$$w_{lmn} = w_{lmn}^{(1)} + (\mu_2 / \mu_1) \theta^3 w_{lmn}^{(2)}, \quad (35)$$

and for the  $i$ th solid

$$\begin{aligned} w_{lmn}^{(i)} = & -|l m n| \left( \frac{\alpha'_i}{|l| \xi_{li} + |m| \xi_{li} + |n| \xi_{li}} \right. \\ & + \frac{\beta'_i}{|l| \xi_{li} + |m| \xi_{li} + |n| \xi_{li}} \\ & \left. + \frac{\gamma'_i}{|l| \xi_{li} + |m| \xi_{li} + |n| \xi_{li}} \right). \end{aligned} \quad (36)$$

The constants  $\alpha'_i$ ,  $\beta'_i$ , and  $\gamma'_i$  are identical to those obtained previously for nonlinear Rayleigh waves [see Eq. (19) of Ref. 8], and they are defined in terms of both Landau's and Murnaghan's notation in Appendix A of Ref. 7.

## 2. Scholte waves

The  $z$  dependence in Eqs. (19) is taken here to be

$$\left. \begin{aligned} u_{xn}^{(1)}(z) &= i \frac{n}{|n|} \left( \xi_{l1} e^{|n|k\xi_{l1}z} + \eta_1 e^{|n|k\xi_{l1}z} \right) \\ u_{zn}^{(1)}(z) &= e^{|n|k\xi_{l1}z} + \eta_1 \xi_{l1} e^{|n|k\xi_{l1}z} \end{aligned} \right\} z < 0, \quad (37)$$

$$\left. \begin{aligned} u_{xn}^{(2)}(z) &= i \frac{n}{|n|} \eta_2 e^{-|n|k\xi_{l2}z} \\ u_{zn}^{(2)}(z) &= -\eta_2 \xi_{l2} e^{-|n|k\xi_{l2}z} \end{aligned} \right\} z > 0, \quad (38)$$

which follow from Eqs. (11) and (12). The kinetic energy is given again by Eq. (25),

$$T = \frac{\rho_1 \zeta}{4k} \sum_n \frac{\dot{a}_n \dot{a}_{-n}}{|n|}, \quad (39)$$

but with  $\zeta$  defined in Appendix B.

The elastic energy in the solid, half-space 1, is given again by Eqs. (27) and (28). The same equations may also be used for the fluid, half-space 2, when the following substitutions are made.<sup>29</sup>  $\mu=0$ ,  $\mathcal{A}=0$ ,  $\mathcal{B}=-K$ ,  $\mathcal{C}=\frac{1}{2}K(1-B/A)$ , where  $K=\rho_2 c_{l2}^2$ , and  $B/A$  is the parameter of nonlinearity of the fluid.<sup>30</sup> For the fluid, Eqs. (27) and (28) thus reduce to

$$\frac{\mathcal{E}_2}{\rho c_l^2} = \frac{1}{2} \left( \frac{\partial u_j}{\partial x_j} \right)^2, \quad (40)$$

$$\frac{\mathcal{E}_3}{\rho c_l^2} = -\frac{1}{2} \frac{\partial u_i}{\partial x_k} \frac{\partial u_k}{\partial x_i} \frac{\partial u_j}{\partial x_j} + \frac{1}{6} \left( 1 - \frac{B}{A} \right) \left( \frac{\partial u_j}{\partial x_j} \right)^3. \quad (41)$$

Use of Eq. (29) leads again to Eq. (32),

$$V_2 = \frac{\rho_1 \zeta}{4k} \omega^2 \sum_n |n| a_n a_{-n}, \quad (42)$$

but with the new definition of  $\zeta$ . Likewise, the integral of the cubic terms in the expression for the elastic energy, Eq. (33), leads to Eq. (34),

$$V_3 = \mu_1 k^2 \sum_{l+m+n=0} w_{lmn} a_l a_m a_n, \quad (43)$$

but with

$$w_{lmn} = w_{lmn}^{(1)} + \eta_2^3 w_{lmn}^{(2)}, \quad (44)$$

$w_{lmn}^{(1)}$  given by Eq. (36), and

$$w_{lmn}^{(2)} = \frac{\rho_2 c_{l2}^2}{\mu_1} \frac{\xi_0^6 [\beta + 4(\xi_0^{-4} - \xi_0^{-2})]}{3(1 - \xi_0^2)^{1/2}} \frac{|lmn|}{|l| + |m| + |n|}, \quad (45)$$

where we introduced the velocity ratio  $\xi_0 = c/c_{l2}$  and the standard coefficient of nonlinearity for the fluid,

$$\beta = 1 + B/2A. \quad (46)$$

The interaction matrix for the fluid [ $w_{lmn}^{(2)}$ , Eq. (45)] is less complicated than for the solid [ $w_{lmn}^{(1)}$ , Eq. (36)] because no shear forces exist in the fluid, and therefore only terms involving  $\xi_{l2}$  contribute to the cubic potential energy.

## B. Evolution equations

The dynamical equations are obtained from the Hamiltonian through cubic order,

$$H = T + V_2 + V_3, \quad (47)$$

where the following expressions were obtained in Sec. II A for both Stoneley and Scholte waves:

$$T = \frac{\rho_1 \zeta}{4k} \sum_n \frac{\dot{a}_n \dot{a}_{-n}}{|n|}, \quad (48)$$

$$V_2 = \frac{\rho_1 \zeta}{4k} \omega^2 \sum_n |n| a_n a_{-n}, \quad (49)$$

$$V_3 = \mu_1 k^2 \sum_{l+m+n=0} w_{lmn} a_l a_m a_n. \quad (50)$$

These expressions differ only in the definitions of the parameters  $\zeta$  and  $w_{lmn}$ , as described above. Moreover, equations of exactly the same form are obtained also for Rayleigh waves, and therefore the algebraic manipulations used previously<sup>7,8</sup> need not be repeated here.

Specifically, generalized coordinates are identified, Hamilton's canonical equations are introduced, dynamical equations describing the evolution of the waveform in time are obtained, approximations based on the assumption of progressive waves are made, and finally, the results are transformed to describe the evolution of a progressive wave in space. In terms of velocities rather than displacements, the final result for the  $i$ th medium is

$$v_j^{(i)}(x, z, t) = \frac{1}{2} \sum_{n=1}^{\infty} v_n(x) u_{jn}^{(i)}(z) e^{in(kx - \omega t)} + \text{c.c.}, \quad (51)$$

$$j = x, z,$$

where  $u_{xn}^{(i)}$  and  $u_{zn}^{(i)}$  are given by Eqs. (22) and (23) for Stoneley waves and by Eqs. (37) and (38) for Scholte waves, and the notation c.c. designates the complex conjugate of the preceding summation. The slowly varying functions  $v_n$  are determined by the following system of coupled spectral equations:

$$\frac{dv_n}{dx} = -\frac{\mu_1 n^2 k}{2\rho_1 c^3 \zeta} \sum_{l+m=n} \frac{|lm|}{lm} R_{lm} v_l v_m, \quad (52)$$

where for Stoneley waves

$$R_{lm} = R_{lm}^{(1)} + (\mu_2/\mu_1) \theta^3 R_{lm}^{(2)}, \quad (53)$$

for Scholte waves (with  $\xi_0 = c/c_{l2}$ )

$$R_{lm} = R_{lm}^{(1)} + \frac{1}{8} \frac{\rho_2 c_{l2}^8}{\rho_1 c_{l1}^8} \frac{\xi_0^{12}}{(1 - \xi_0^2)^2} \frac{\beta + 4(\xi_0^{-4} - \xi_0^{-2})}{|l| + |m| + |l+m|}, \quad (54)$$

and for Rayleigh waves [obtained by setting  $\rho_2 = 0$  in Eq. (54)]

$$R_{lm} = R_{lm}^{(1)}. \quad (55)$$

The nonlinearity matrices  $R_{lm}^{(i)}$  are defined by

$$R_{lm}^{(i)} = \frac{\alpha'_i}{|l|\xi_{li} + |m|\xi_{li} + |l+m|\xi_{li}} + \frac{\alpha'_i}{|l|\xi_{li} + |m|\xi_{li} + |l+m|\xi_{li}}$$

$$+ \frac{\alpha'_i}{|l|\xi_{li} + |m|\xi_{li} + |l+m|\xi_{li}} + \frac{\beta'_i}{|l|\xi_{li} + |m|\xi_{li} + |l+m|\xi_{li}}$$

$$+ \frac{\beta'_i}{|l|\xi_{li} + |m|\xi_{li} + |l+m|\xi_{li}} + \frac{\beta'_i}{|l|\xi_{li} + |m|\xi_{li} + |l+m|\xi_{li}}$$

$$+ \frac{3\gamma'_i}{|l|\xi_{li} + |m|\xi_{li} + |l+m|\xi_{li}}, \quad (56)$$

where the constants  $\alpha'_i$ ,  $\beta'_i$ , and  $\gamma'_i$  are defined explicitly in terms of the second- and third-order elastic constants of the media, and they are provided in terms of both Landau's and Murnaghan's notation in Appendix A of Ref. 7. The result for Rayleigh waves, based on Eq. (55), is the same as that obtained previously.<sup>7,8</sup>

Because the symmetries of  $R_{lm}$  are common to all three types of surface waves, the nonlinear evolution of the three wave types is expected to be similar. If we define the function<sup>17</sup>

$$F(l, m) = |l+m| R_{lm}, \quad (57)$$

the symmetries can be expressed as follows (with  $n = l+m$ ):<sup>15-17</sup>

$$R_{lmn} = R_{mnl} = R_{nlm}, \quad (58)$$

$$F(l, -m) = F(m, -l), \quad (59)$$

$$F(l, m) = F(m, l), \quad (60)$$

$$F(Nl, Nm) = F(l, m). \quad (61)$$

As a result, all relations developed previously for Rayleigh waves<sup>7,8,12-17</sup> on the basis of these symmetries can be applied to Stoneley and Scholte waves. This is significant because it means that a large body of research devoted specifically to Rayleigh waves can be applied to other types of surface waves. For example, the results which describe the diffraction of nonlinear Rayleigh wave beams,<sup>13</sup> local and nonlocal nonlinearity in Rayleigh waves,<sup>15</sup> and the development of time domain equations for nonlinear Rayleigh waves<sup>16</sup> can be applied to nonlinear Stoneley and Scholte waves in a reasonably straightforward way. An especially useful application of the permutation symmetry in Eq. (58) is in the derivation of the following relation for energy conservation.<sup>14</sup>

$$\sum_{n=1}^{\infty} \frac{|v_n|^2}{n} = \text{const.} \quad (62)$$



### III. ANALYSIS OF THE EVOLUTION EQUATION

For both analytical and numerical studies of spectral interactions, it is often more convenient to rewrite Eq. (52) in the form

$$\frac{dv_n}{dx} = \frac{\mu_1 n^2 k}{2\rho_1 c^3 \zeta} \left( 2 \sum_{m=n+1}^{\infty} R_{m,n-m} v_m v_{m-n}^* - \sum_{m=1}^{n-1} R_{m,n-m} v_m v_{n-m} \right), \quad (63)$$

where the first summation accounts for all spectral interactions generating difference frequencies, and the second corresponds to sum frequency generation.

#### A. Approximate shock formation distance

Shull *et al.*<sup>12</sup> proposed a simple approximate expression for the shock formation distance for Rayleigh waves by comparing the rate of second-harmonic generation with the corresponding process that occurs for sound waves in fluids. As shown by Knight *et al.*<sup>31</sup> via comparison with the shock formation distance calculated numerically for a variety of real materials, the expression derived by Shull *et al.* is accurate to within about 20%. We thus anticipate that the procedure followed by Shull *et al.* can be used to obtain a reasonable estimate of the shock formation distance for Stoneley and Scholte waves. A more accurate approximate analytical expression for the shock formation distance was obtained by Knight *et al.*,<sup>31</sup> but for our purposes here the earlier expression is adequate.

Thus consider a source at  $x=0$  which radiates at a single frequency such that

$$v_n(0) = \begin{cases} v_0 & n=1 \\ 0 & n>1 \end{cases} \quad (64)$$

where  $v_0$  is the relative velocity amplitude (assumed positive) at  $x=0$ . The boundary condition may be expressed in terms of the particle velocity components in medium 1, evaluated at the interface, via Eqs. (51):

$$v_x^{(1)}(0,0,t) = (\xi_{t1} + \eta_1) v_0 \sin \omega t, \quad (65)$$

$$v_z^{(1)}(0,0,t) = (1 + \eta_1 \xi_{11}) v_0 \cos \omega t. \quad (66)$$

Now solve Eq. (63) via successive approximations to obtain an approximate solution for second-harmonic generation in a lossless medium. The first two equations are

$$\frac{dv_1}{dx} = 0, \quad (67)$$

$$\frac{dv_2}{dx} = -\frac{2\mu_1 k}{\rho_1 c^3 \zeta} R_{11} v_1^2, \quad (68)$$

the solutions of which, subject to Eqs. (64), are

$$v_1 = v_0, \quad (69)$$

$$v_2 = -\frac{1}{2} \frac{R_{11}}{|R_{11}|} \frac{x}{\bar{x}} v_0^2, \quad (70)$$

where

$$\bar{x} = \frac{\rho_1 c^3 \zeta}{4\mu_1 |R_{11}| v_0 k}. \quad (71)$$

Equation (71) is the approximate shock formation distance, deduced by noting that the magnitude of Eq. (70) is  $|v_2|/v_0 = x/2\bar{x}$ , i.e., the same result as for second-harmonic generation in a plane sound wave in a fluid when  $\bar{x}$  is the shock formation distance for that sound wave.<sup>6</sup>

By further extension of the analogy with fluids, we define a coefficient of nonlinearity  $\beta$  as follows:

$$\bar{x} = \frac{1}{|\beta| \epsilon k}. \quad (72)$$

For a fluid,  $\epsilon$  is the peak acoustic Mach number at the source, i.e., particle velocity divided by sound speed. For a progressive wave,  $\epsilon$  is also equivalent to peak strain at the source. In the case of surface waves, one must decide whether the vertical or horizontal component of  $\epsilon$  should be the fundamental quantity determining the source amplitude, and when that choice is made, whether the motion in the lower or upper medium is more convenient for reference.

We begin by writing the shock formation distance in terms of the horizontal and vertical components of strain in the lower, elastic solid medium (recall that the upper medium is a vacuum for Rayleigh waves, fluid for Scholte waves, and solid for Stoneley waves):

$$\bar{x} = \frac{1}{|\beta_{x1}| \epsilon_{x1} k} = \frac{1}{|\beta_{z1}| \epsilon_{z1} k}, \quad (73)$$

where  $\epsilon_{j1} = |v_j^{(1)}(0,0,t)|/c$  is the strain, with  $|v_j^{(1)}(0,0,t)|$  denoting the magnitude of the  $j$ th ( $j=x,z$ ) Cartesian component of the particle velocity at the interface [recall Eqs. (65) and (66)], at the source, in medium 1. From a theoretical point of view, it is the horizontal component that bears the most similarity to the nonlinear distortion observed in sound waves.<sup>15,16</sup> From an experimental point of view, however, the vertical component is easiest to measure, in which case the second definition is more convenient. We therefore present both definitions. Inversion of the definitions to solve for the coefficients of nonlinearity gives

$$\beta_{x1} = \frac{4\mu_1 R_{11}}{\rho_1 c^2 \zeta |\xi_{t1} + \eta_1|}, \quad (74)$$

$$\beta_{z1} = \frac{4\mu_1 R_{11}}{\rho_1 c^2 \zeta |1 + \eta_1 \xi_{11}|}. \quad (75)$$

Equations (74) and (75) apply to all three types of surface waves. For Scholte waves, the pressure in the overlaying fluid is the quantity that is typically measured in an experiment. In this case, the approximate shock formation distance can be written as

$$\bar{x} = \frac{1}{|\beta_f| \epsilon_f k}, \quad (76)$$

where  $\epsilon_f = p_0/\rho_2 c^2$ ,  $p_0$  is the peak sound pressure in the fluid at the interface,  $\rho_2$  is the density of the fluid, and  $c$  is the Scholte wave speed. We thus obtain

TABLE IV. Coefficients of nonlinearity for several surface waves.

Surface wave	Medium	$\beta_{x1}$	$\beta_{z1}$	$\beta_f$
Rayleigh	Steel	0.72	0.48	
Rayleigh	Nickel	0.44	0.28	
Rayleigh	Glass	-2.1	-1.4	
Rayleigh	Aluminum	1.8	1.2	
Rayleigh	Copper	-3.2	-2.0	
Rayleigh	Sandstone	290	210	
Scholte	Steel-Water	436	150	3.49
Scholte	Nickel-Water	168	54.3	3.46
Scholte	Glass-Water	81.0	30.8	3.39
Scholte	Aluminum-Water	126	38.2	3.44
Scholte	Copper-Water	160	54.2	3.45
Stoneley	Steel-Glass	0.29	0.14	
Stoneley	Nickel-Glass	0.23	0.12	

$$\beta_f = \frac{4\mu_1 R_{11}}{\rho_1 c^2 \zeta |\eta_2|}, \tag{77}$$

where  $R_{11}$  here is given by Eq. (54).

Coefficients of nonlinearity were calculated for several types of Stoneley, Scholte, and Rayleigh waves. For the solids that were considered we used the constants given in Table I, and for water we used  $\rho=1000 \text{ kg/m}^3$ ,  $c_l = 1481 \text{ m/s}$ , and  $\beta=3.5$ . The results are listed in Table IV. Positive values of  $\beta_{x1}$  indicate waveform distortion similar to that in fluids, in the following sense: In the horizontal velocity waveform (for which the particle motion is along the propagation direction) positive portions advance in space and negative portions recede, leading to shock formation in compression phases of waveforms. For negative values of  $\beta_{x1}$  the opposite occurs, and shocks are produced in the expansion phases.

Consider first the nonlinearity coefficient  $\beta_{x1}$  for Rayleigh waves. For the first five materials (all but sandstone) the magnitudes are of order one, and the values for two of the materials, glass and copper, are negative. Similar results are obtained with predictions for compressional waves. The large value for sandstone is consistent with coefficients of nonlinearity with magnitudes of order  $10^3$  to  $10^4$  that have been reported for compressional waves in sandstone.<sup>32,33</sup> The large nonlinearity is due to the granular microstructure. More extensive calculations and discussions of nonlinearity coefficients for Rayleigh waves are provided by Knight *et al.*<sup>31</sup>

The values of  $\beta_{x1}$  for Stoneley waves in the two pairs of solids considered are consistent in magnitude with the values for Rayleigh waves.

### B. Considerations for Scholte waves

Consider now the values of  $\beta_{x1}$  in Table IV for Scholte waves, which correspond to the first five materials listed for Rayleigh waves, with water assumed for the upper (fluid) medium. A general observation may be made regarding the introduction of water, replacing the vacuum, in the upper half-space. The magnitude of  $\beta_{x1}$  increases by factors of order 10 to 100. Recall that  $\beta_{x1}$  is associated with the strain at the source in the lower (solid) medium. For example, in the case of aluminum, let a constant source amplitude be maintained while a vacuum in the upper medium is replaced

with water. As  $\beta_{x1}$  changes from 1.8 to 126 when the water is introduced, the shock formation distance is correspondingly reduced by a factor of 70.

The large values of  $\beta_{x1}$  for Scholte waves follow from Eq. (54). As seen in Table III,  $c$  is often very close to  $c_{l2}$ , therefore  $\xi_0 \approx 1$ , and the second term in Eq. (54) can be very large in comparison with the first. Finite-amplitude effects are then determined completely by the nonlinearity in the fluid, with nonlinearity in the solid no longer playing a significant role.

A physical explanation for the magnitude of  $\beta_{x1}$  for Scholte waves follows from comparison of the particle motions in the two media. The ratio of the horizontal particle displacements (and therefore the ratio of the corresponding particle velocities) at the interface is given for Scholte waves by Eq. (14), and as was pointed out, this ratio becomes very large, in proportion to  $(1 - c^2/c_{l2}^2)^{-1/2}$ , when the Scholte wave speed approaches the speed of sound in the fluid. Insight into how the rate of nonlinear distortion in Scholte waves is affected may be obtained by taking the limit  $c \rightarrow c_{l2}$  in Eq. (54), and then making use of Eq. (14), noting that  $u_x^{(2)}/u_x^{(1)} = v_x^{(2)}/v_x^{(1)}$ :

$$\beta_{x1} \approx \beta \left( \frac{v_x^{(2)}}{v_x^{(1)}} \right)_{x=z=0}, \quad c \rightarrow c_{l2}, \tag{78}$$

where  $\beta$  is the coefficient of nonlinearity for the fluid alone. Similarly, Eq. (77) reveals that

$$\beta_f \approx \beta, \quad c \rightarrow c_{l2}. \tag{79}$$

Inspection of Table IV shows that Eq. (79) is accurate to within 5% for the Scholte waves listed ( $\beta=3.5$ ). This reaffirms that for  $c \approx c_{l2}$ , finite-amplitude effects are determined almost completely by the nonlinearity in the fluid. Furthermore, the shock formation distance for a Scholte wave (for  $c \approx c_{l2}$ ) with initial sound pressure amplitude  $p_0$  (as measured in the fluid at the interface,  $z=0$ ) is approximately the same as the shock formation distance for a plane acoustic wave having the same initial pressure amplitude.

### C. Mode coupling

For Scholte waves with  $c \approx c_{l2}$ , coupling between the surface wave mode and the acoustic mode in the fluid may need to be taken into account, in which case a basic assumption underlying the present theory is violated. The same issue arises with Stoneley waves, when  $c$  approaches the slower of the shear wave speeds in the two elastic solids. The length scale associated with mode coupling is the coherence length

$$L = \frac{\pi}{k - k_b} = \frac{c}{2f(1 - c/c_b)}, \tag{80}$$

where  $f = \omega/2\pi$  is the source frequency,  $k = \omega/c$  is the corresponding wave number for the surface mode, and  $k_b = \omega/c_b$  is the wave number at the same frequency for the bulk mode with the smallest propagation speed in the two media. For Stoneley waves  $c_b$  is the smaller of the two shear wave speeds, for Scholte waves it is the sound speed in the fluid, and for Rayleigh waves it is the single shear wave speed. For propagation distances  $x \ll L$  the surface wave is in

phase with, and may therefore couple with, a bulk wave mode of propagation, and the present nonlinear theory cannot be expected to be valid in this region.

An appropriate length scale over which significant nonlinear effects occur is the shock formation distance  $\bar{x}$ . If this length is substantially less than the coherence length,

$$\bar{x} \ll L, \quad (81)$$

mode coupling may be anticipated. Insofar as  $\bar{x}$  is inversely proportional to the characteristic source amplitude  $v_0$ , Eq. (81) provides an indication of when the source amplitude is sufficiently high that mode coupling may occur.

For example, in previous work<sup>5</sup> the above evolution equation for nonlinear Rayleigh waves [i.e., based on the nonlinearity matrix given by Eq. (55)] was used to model distortion and shock formation in laser-generated surface waves in fused quartz. For this experiment Eq. (80) yields  $L = 0.9$  mm, and the shock formation distance was calculated to be  $\bar{x} = 1.3$  mm. Therefore Eq. (81) does not apply, mode coupling is not expected to be a factor, and excellent agreement between theory and experiment was indeed obtained.

#### D. Numerical results

For numerical calculations, it is useful to introduce the dimensionless quantities

$$V_n = v_n / v_0, \quad X = x / \bar{x}, \quad A_n = \alpha_n \bar{x}, \quad (82)$$

where we have chosen the approximate shock formation distance  $\bar{x}$  to be the length scale, and  $v_0$  is the scale by which we measure particle velocity. The quantity  $\alpha_n$  is the small-signal absorption coefficient corresponding to the  $n$ th-harmonic component. Absorption is introduced *ad hoc* in Eq. (63) to facilitate numerical calculations when shocks exist in the waveforms. Making these substitutions in Eq. (63), we obtain for the evolution equation, with the absorption term included,

$$\frac{dV_n}{dX} + A_n V_n = \frac{n^2}{8|R_{11}|} \left( 2 \sum_{m=n+1}^N R_{m,n-m} V_m V_{m-n}^* - \sum_{m=1}^{n-1} R_{m,n-m} V_m V_{n-m} \right), \quad (83)$$

where the upper limit of the first summation has been replaced with a maximum harmonic number  $N$ , and the harmonic index now spans the interval  $1 \leq n \leq N$ . The energy relation associated with Eq. (83) is<sup>14</sup>

$$\frac{d}{dX} \sum_{n=1}^N \frac{|V_n|^2}{n} = -2 \sum_{n=1}^N A_n \frac{|V_n|^2}{n}, \quad (84)$$

which for  $A_n = 0$  reduces to

$$\sum_{n=1}^N \frac{|V_n|^2}{n} = \text{const.} \quad (85)$$

Equations (83) were integrated numerically using a fourth-order Runge–Kutta algorithm. For thermoviscous dissipation, the absorption coefficient for surface waves is pro-

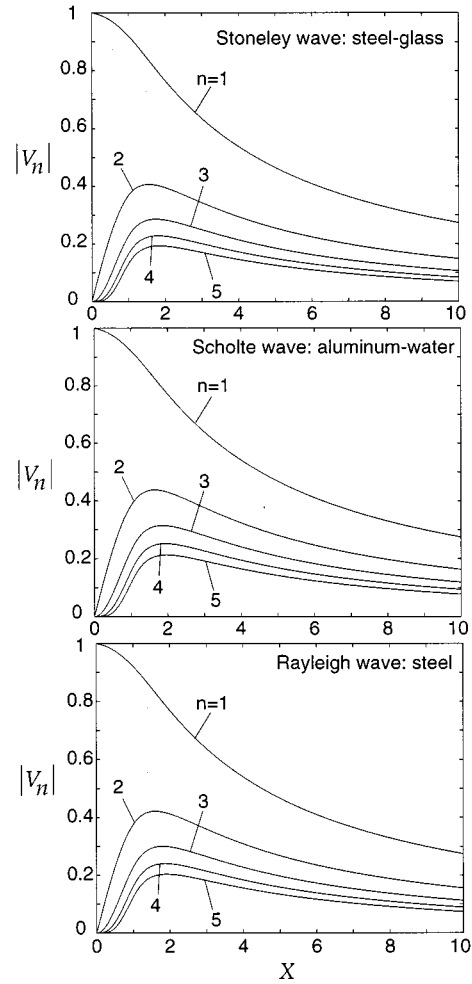


FIG. 6. Harmonic propagation curves for Stoneley, Scholte, and Rayleigh waves.

portional to frequency squared.<sup>34</sup> Therefore, the absorption term in Eq. (83) was taken to be of the form

$$A_n = n^2 A_1. \quad (86)$$

A monofrequency source condition was chosen according to Eqs. (64)–(66), and we used  $A_1 = 0.02$  in the calculations.

A set of harmonic propagation curves obtained from the integration is shown in Fig. 6. Propagation curves for different pairs of media show very similar behavior. The shapes of the propagation curves are nearly identical for different media and different types of surface waves when plotted versus the dimensionless length scale  $X = x / \bar{x}$ , where  $\bar{x}$  is given by Eq. (73).

A set of distorting waveforms calculated at  $z = 0$  is shown in Figs. 7 and 8. Figure 7 displays the horizontal components of particle velocity,  $v_x^{(1)}$ , of the initially sinusoidal surface waves. Figure 8 shows the corresponding vertical components of particle velocity,  $v_z^{(1)}$ . The three types of surface waves have very similar waveforms. The slight differences may be attributed, at least in part, to the use of an approximate shock formation distance. For this reason, the Scholte waveform shown in Fig. 7 has not quite formed a shock (vertical discontinuity in horizontal particle velocity) at  $X = 1$ , whereas the Stoneley wave has formed a shock.

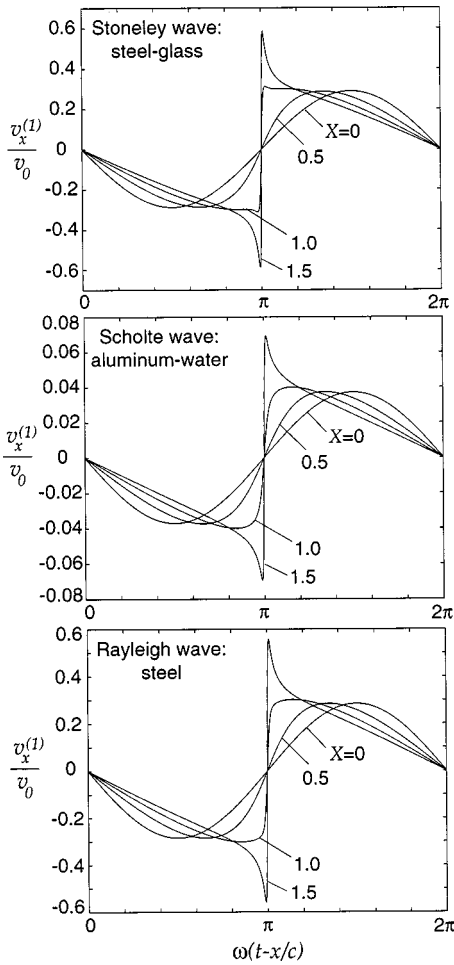


FIG. 7. Horizontal velocity waveforms for Stoneley, Scholte, and Rayleigh waves at  $z=0$ , in medium 1.

The processes of harmonic generation, waveform distortion, and shock formation are thus found to be qualitatively similar for all three types of surface waves.

### E. Comparison of interaction matrices

The interaction matrices  $R_{lm}$  for the three different types of surface waves are now compared. In order to make this comparison, we introduce the nonlinear interaction kernel<sup>31</sup>

$$K_{lm} = |l+m| \frac{R_{lm}}{R_{11}}. \quad (87)$$

Notice that  $K_{lm}$  is proportional to the function  $F(l,m)$ , which is defined in Eq. (57) and was introduced in earlier work.<sup>15</sup> For a fluid the kernel is a constant,  $K_{lm}=2$ . It is convenient to plot  $K_{lm}$  as a continuous function of  $m/l$ , where  $m/l > 0$  corresponds to sum frequency generation,  $m/l < 0$  to difference frequency generation, and  $m/l = -1$  to the generation of a signal at zero frequency. In Fig. 9 we show the nonlinear interaction kernels for a Stoneley wave (steel–glass interface), a Scholte wave (aluminum–water interface), and a Rayleigh wave (steel). Similar plots have been presented previously<sup>17,31</sup> for nonlinear Rayleigh waves. The behavior of  $K_{lm}$  for  $m/l < 0$  is similar for the three types of surface waves, insofar as the kernel vanishes at  $m/l = -1$  and the variation is rapid down to  $m/l \approx -5$ . For  $m/l > 0$ ,

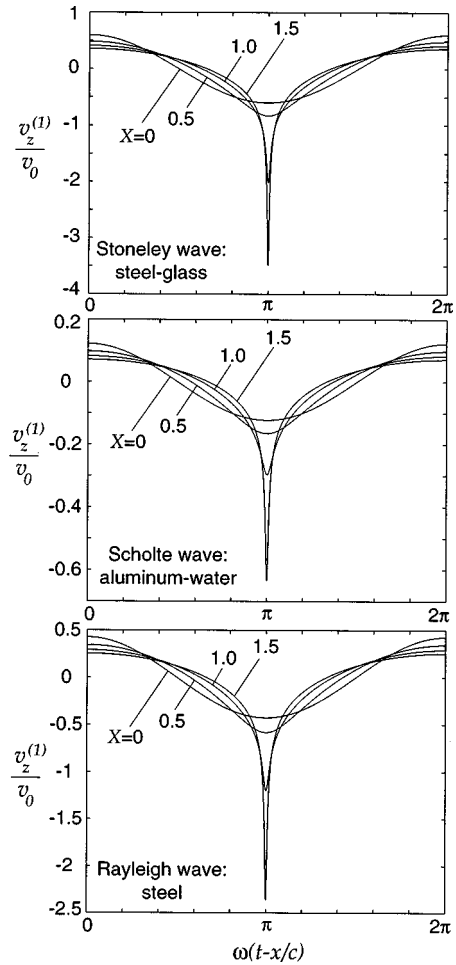


FIG. 8. Vertical velocity waveforms for Stoneley, Scholte, and Rayleigh waves at  $z=0$ , in medium 1.

$K_{lm}$  varies rapidly for Rayleigh and Stoneley waves up to  $m/l \approx 5$ , whereas  $K_{lm}$  is virtually constant and equal to the value for a fluid in the case of Scholte waves. Harmonic generation in Scholte waves thus appears surface wavelike for difference frequencies and bulk wavelike for sum frequencies. Since an initially sinusoidal wave will undergo pri-

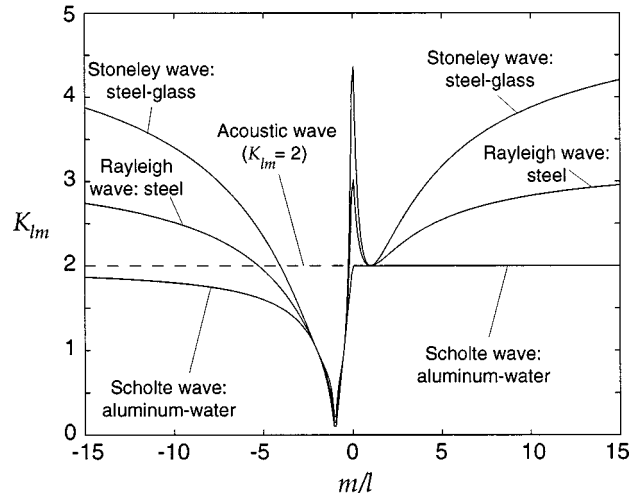


FIG. 9. Comparison of nonlinear interaction kernels.



marily sum frequency generation as it begins to distort, the spectrum of a Scholte wave may be expected to be similar to the spectrum of an acoustic wave for short propagation distances.

#### IV. SUMMARY

A theory was developed for nonlinear Stoneley and Scholte waves, and it reduces to earlier results for nonlinear Rayleigh waves. Nonlinear interaction matrices are presented for all three types of surface waves, and these matrices are found to exhibit the same symmetries. Coefficients of nonlinearity were defined and calculated for surface waves in several media. The evolution equation was integrated numerically, and when propagation distance was normalized by an approximate shock formation distance, we found that the Stoneley, Scholte, and Rayleigh waves behave in a qualitatively similar manner.

#### ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, with additional support provided by Los Alamos National Laboratory.

#### APPENDIX A: STONELEY WAVE RELATIONS

For Stoneley waves, the following definitions apply:

$$\zeta = \zeta_1 + m \theta^2 \zeta_2, \quad (\text{A1})$$

$$\zeta_1 = \xi_{t1} + \frac{1}{\xi_{t1}} + \eta_1^2 \left( \xi_{t1} + \frac{1}{\xi_{t1}} \right) + 4 \eta_1, \quad (\text{A2})$$

$$\zeta_2 = \xi_{t2} + \frac{1}{\xi_{t2}} + \eta_2^2 \left( \xi_{t2} + \frac{1}{\xi_{t2}} \right) + 4 \eta_2, \quad (\text{A3})$$

and

$$\eta_1 = - \frac{(1 + \xi_{t2} \xi_{t1}) + (1 - \xi_{t2} \xi_{t2}) \theta}{\xi_{t2} + \xi_{t1}}, \quad (\text{A4})$$

$$\eta_2 = - \frac{(1 + \xi_{t1} \xi_{t2}) + (1 - \xi_{t1} \xi_{t1}) \theta}{\xi_{t2} + \xi_{t1}}, \quad (\text{A5})$$

$$\theta = \frac{2 \xi_{t2} (1 - \xi_{t1} \xi_{t1}) (\xi_2^2 - m \xi_1^2) - \xi_2^2 \xi_1^2 (\xi_{t2} + \xi_{t1})}{2 \xi_{t1} (1 - \xi_{t2} \xi_{t2}) (\xi_2^2 - m \xi_1^2) + m \xi_2^2 \xi_1^2 (\xi_{t2} + \xi_{t1})}, \quad (\text{A6})$$

where  $\xi_i = c/c_{ti}$ .

#### APPENDIX B: SCHOLTE WAVE RELATIONS

For Scholte waves, the following definitions apply:

$$\zeta = \zeta_1 + m \eta_2^2 \zeta_2, \quad (\text{B1})$$

$$\zeta_1 = \xi_{t1} + \frac{1}{\xi_{t1}} + \eta_1^2 \left( \xi_{t1} + \frac{1}{\xi_{t1}} \right) + 4 \eta_1, \quad (\text{B2})$$

$$\zeta_2 = \xi_{t2} + \frac{1}{\xi_{t2}}, \quad (\text{B3})$$

where

$$\eta_1 = - \left( \frac{1 + \xi_{t1}^2}{2 \xi_{t1}} \right), \quad (\text{B4})$$

$$\eta_2 = - \left( \frac{1 - \xi_{t1}^2}{2 \xi_{t2}} \right). \quad (\text{B5})$$

#### APPENDIX C: DEPTH DEPENDENCE AND THE THEORY OF GUSEV *et al.*

Another theoretical model for nonlinear Scholte waves was presented recently by Gusev and co-workers.<sup>35</sup> These authors considered only nonlinearity in the liquid, not in the solid. Our Scholte wave theory takes into account the nonlinearity of both the liquid and the solid, and this fact alone is sufficient to distinguish our theoretical model from theirs.

There are, however, more fundamental distinctions to be made. In particular, the analytical approach of Gusev and co-workers is the same one they used previously to model nonlinear Rayleigh waves in isotropic solids<sup>36</sup> and, subsequently, surface waves in crystals.<sup>37</sup> In these articles, the authors discuss the depth dependence of nonlinear surface waves and its influence on the nonlinear evolution equations. More to the point, they suggest that the depth dependence we assume in Eq. (19), specifically, the set of functions  $u_{jn}$  from linear theory, is inappropriate for nonlinear theory.<sup>36,37</sup> Here we respond to this comment, and also point out other features that distinguish our theory from theirs.

We begin by considering the depth dependence. Equation (19) is the main part of the asymptotic solution of the nonlinear problem. The small parameter in this problem is the wave amplitude, which may be characterized by a dimensionless quantity  $\epsilon$ . Equation (19) is thus classified as  $O(\epsilon)$ . The full asymptotic solution also includes terms of  $O(\epsilon^2)$  and higher, the form of which is

$$u_j = \sum_{n=-\infty}^{\infty} e^{inkx} \left[ a_n(t) u_{jn}(z) + \int_0^{\infty} b_n(t, k_z) v_{jn}(z, k_z) dk_z \right] + O(\epsilon^3), \quad (\text{C1})$$

where the depth functions  $u_{jn}$  and  $v_{jn}$  are each  $O(1)$ , with  $a_n = O(\epsilon)$  and  $b_n = O(\epsilon^2)$ . The  $O(\epsilon)$  summation over  $a_n u_{jn}$  in Eq. (C1), which is identical to Eq. (19), is a combination of purely surface wave modes. The integral over  $b_n v_{jn}$  is  $O(\epsilon^2)$ , and it includes contributions corresponding to the continuous spectrum of vertical wave numbers  $k_z$  for all bulk wave modes of propagation. Both the  $O(\epsilon)$  and  $O(\epsilon^2)$  expressions in Eq. (C1) are required to satisfy completely the nonlinear equations of motion and boundary conditions at  $O(\epsilon^2)$ . The depth dependence  $v_{jn}$  differs from  $u_{jn}$  and, as a result, the depth dependence of the full asymptotic solution indeed differs from the depth dependence for pure surface waves. Evolution equations at  $O(\epsilon^2)$  for the slowly varying amplitudes of progressive waves [such as Eq. (52)] do not depend on the  $O(\epsilon^2)$  correction in Eq. (C1).

In principle, it is possible to calculate the  $O(\epsilon^2)$  terms in Eq. (C1) and thus determine the lowest-order correction to the depth dependence obtained from linear theory. However, this calculation requires  $O(\epsilon^3)$  terms to be included in evo-

lution equations such as Eq. (52), and therefore  $O(\epsilon^4)$  terms must be retained in elastic energy equations such as Eq. (26). Gusev *et al.*<sup>36,37</sup> do not include any of these higher-order terms.

Gusev *et al.*<sup>36,37</sup> used a method different from ours to derive their evolution equations. They begin with a reasonable form for the solution at  $O(\epsilon)$  and obtained an evolution equation at  $O(\epsilon^2)$  subject to the condition that the solution obey the boundary conditions and equations of motion at  $O(\epsilon^2)$  in a vanishingly small region near the free surface. However, they do not ensure that their solution satisfies the equations of motion throughout the half-space occupied by the solid, nor is their result verified even at a distance on the order of the penetration depth away from the surface. Their evolution equations for Rayleigh waves,<sup>36</sup> Scholte waves,<sup>35</sup> and surface waves in crystals<sup>37</sup> differ from the equations derived by our Hamiltonian formalism and by the multiple scale approach used by Parker.<sup>38</sup> In contrast, our results for Rayleigh waves have been shown<sup>7</sup> to be identical to those obtained by Parker, even though the derivations are entirely different.

Moreover, the results derived by Gusev and co-workers have several unusual features: (1) They do not satisfy conservation of energy; (2) The frequency dependence of the quasilinear solution of their evolution equation for the case of sum and difference frequency generation does not agree with the result derived from the fundamental equations of motion by simple perturbation theory; (3) If an external body force were included in their theory, one that vanishes at the interface but which is resonant with the surface wave mode (and will therefore pump energy in or out of that mode), their evolution equation would predict zero effect. Our results are free from each of these disadvantages.

<sup>1</sup>J. W. S. Rayleigh, "On waves propagated along the plane surface of an elastic solid," *Proc. London Math. Soc.* **17**, 4–11 (1885).

<sup>2</sup>R. Stoneley, "Elastic waves at the surface of separation between two solids," *Proc. R. Soc. London, Ser. A* **106**, 416–428 (1924).

<sup>3</sup>J. G. Scholte, "On the Stoneley wave equation," *Proc. K. Ned. Akad. Wet.* **45**, Pt. 1: 20–25, Pt. 2: 159–164 (1942).

<sup>4</sup>A. Lomonosov and P. Hess, "Laser excitation and propagation of nonlinear surface acoustic wave pulses," in *Nonlinear Acoustics in Perspective*, edited by R. J. Wei (Nanjing University Press, Nanjing, 1996), pp. 106–111.

<sup>5</sup>A. Lomonosov, V. G. Mikhalevich, P. Hess, E. Yu. Knight, M. F. Hamilton, and E. A. Zabolotskaya, "Laser-generated nonlinear Rayleigh waves with shocks," *J. Acoust. Soc. Am.* **105**, 2093–2096 (1999).

<sup>6</sup>D. T. Blackstock, M. F. Hamilton, and A. D. Pierce, "Progressive waves in lossless and lossy fluids," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, Boston, 1998), Chap. 4.

<sup>7</sup>E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "General theory for the spectral evolution of nonlinear Rayleigh waves," *J. Acoust. Soc. Am.* **102**, 1402–1417 (1997).

<sup>8</sup>E. A. Zabolotskaya, "Nonlinear propagation of plane and circular Rayleigh waves in isotropic solids," *J. Acoust. Soc. Am.* **91**, 2569–2575 (1992).

<sup>9</sup>M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Nonlinear surface acoustic waves in crystals," *J. Acoust. Soc. Am.* **105**, 639–651 (1999).

<sup>10</sup>R. E. Kumon, M. F. Hamilton, Yu. A. Il'inskii, E. A. Zabolotskaya, P. Hess, A. Lomonosov, and V. G. Mikhalevich, "Pulsed nonlinear surface acoustic waves in crystals," in *Proceedings of the 16th International Congress on Acoustics*, edited by P. K. Kuhl and L. A. Crum (Acoustical Society of America, New York, 1998), Vol. III, pp. 1557–1558.

<sup>11</sup>M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Nonlinear

surface wave propagation in a piezoelectric material," *J. Acoust. Soc. Am.* **100**, 2567(A) (1996).

<sup>12</sup>D. J. Shull, M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, "Harmonic generation in plane and cylindrical nonlinear Rayleigh waves," *J. Acoust. Soc. Am.* **94**, 418–427 (1993).

<sup>13</sup>D. J. Shull, E. E. Kim, M. F. Hamilton, and E. A. Zabolotskaya, "Diffraction effects in nonlinear Rayleigh wave beams," *J. Acoust. Soc. Am.* **97**, 2126–2137 (1995).

<sup>14</sup>M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, "On the existence of stationary nonlinear Rayleigh waves," *J. Acoust. Soc. Am.* **93**, 3089–3095 (1993).

<sup>15</sup>M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Local and nonlocal nonlinearity in Rayleigh waves," *J. Acoust. Soc. Am.* **97**, 882–890 (1995).

<sup>16</sup>M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Evolution equations for nonlinear Rayleigh waves," *J. Acoust. Soc. Am.* **97**, 891–897 (1995).

<sup>17</sup>E. Yu. Knight, "Generalization of the theory for nonlinear Rayleigh waves to nonplanar and transient waveforms, and investigation of pulse propagation," M. A. Thesis, The University of Texas at Austin (1995).

<sup>18</sup>K. Sezawa and K. Kanai, "The range of possible existence of Stoneley waves and some related problems," *Bull. Earthquake Res. Inst.* **17**, 1–8 (1939).

<sup>19</sup>L. Cagniard, "Réflexion et réfraction des ondes séismiques progressives." *Thèse* (Gauthier-Villars et Cie., Paris, 1939).

<sup>20</sup>J. D. Achenbach, *Wave Propagation in Elastic Solids* (American Elsevier, New York, 1973).

<sup>21</sup>T. E. Owen, "Surface wave phenomena in ultrasonics," *Prog. Appl. Mater. Res.* **6**, 71–87 (1964).

<sup>22</sup>B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Florida, 1990).

<sup>23</sup>L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), pp. 461–462.

<sup>24</sup>R. F. S. Hearmon, "The elastic constants of crystals and other anisotropic materials; The third- and higher-order elastic constants," in *Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology, Group III, Vol. 11*, edited by K. H. Hellwege and A. M. Hellwege (Springer-Verlag, New York, 1979), pp. 245–286.

<sup>25</sup>L. K. Zarembo and V. A. Krasil'nikov, "Nonlinear phenomena in the propagation of elastic waves in solids," *Sov. Phys. Usp.* **13**, 778–797 (1971).

<sup>26</sup>P. A. Johnson and T. J. Shankland, "Nonlinear generation of elastic waves in granite and sandstone: Continuous wave and travel time observations," *J. Geophys. Res.* **94**, 17727–17733 (1989).

<sup>27</sup>K. W. Winkler and X. Liu, "Measurements of third-order elastic constants in rocks," *J. Acoust. Soc. Am.* **100**, 1392–1398 (1996).

<sup>28</sup>L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Pergamon, New York, 1986), p. 107.

<sup>29</sup>S. Kostek, B. K. Sinha, and A. N. Norris, "Third-order elastic constants for an inviscid fluid," *J. Acoust. Soc. Am.* **94**, 3014–3017 (1993).

<sup>30</sup>R. T. Beyer, "Parameter of nonlinearity in fluids," *J. Acoust. Soc. Am.* **32**, 719–721 (1960).

<sup>31</sup>E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "On Rayleigh wave nonlinearity, and analytical approximation of the shock formation distance," *J. Acoust. Soc. Am.* **102**, 2529–2535 (1997).

<sup>32</sup>G. D. Meegan, P. A. Johnson, R. A. Guyer, and K. R. McCall, "Observations of nonlinear elastic wave behavior in sandstone," *J. Acoust. Soc. Am.* **94**, 3187–3191 (1993).

<sup>33</sup>P. A. Johnson and P. N. J. Rasolofosaon, "Nonlinear elasticity and stress-induced anisotropy in rocks," *J. Geophys. Res.* **101**, 3113–3124 (1996).

<sup>34</sup>R. W. Lardner, "Nonlinear Rayleigh waves: Harmonic generation, parametric amplification, and thermoviscous damping," *J. Appl. Phys.* **55**, 3251–3260 (1984).

<sup>35</sup>V. E. Gusev, W. Lauriks, and J. Thoen, "Evolution equation for nonlinear Scholte waves," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45**, 170–178 (1998).

<sup>36</sup>V. E. Gusev, W. Lauriks, and J. Thoen, "Theory for the time evolution of nonlinear Rayleigh waves in an isotropic solid," *Phys. Rev. B* **55**, 9344–9347 (1997).

<sup>37</sup>V. E. Gusev, W. Lauriks, and J. Thoen, "New evolution equations for the nonlinear surface acoustic waves on an elastic solid of general anisotropy," *J. Acoust. Soc. Am.* **103**, 3203–3215 (1998).

<sup>38</sup>D. F. Parker, "Waveform evolution for nonlinear surface acoustic waves," *Int. J. Eng. Sci.* **26**, 59–75 (1988).

# The stabilized self-starter

Michael D. Collins

Naval Research Laboratory, Washington, D.C. 20375

(Received 2 November 1998; revised 4 June 1999; accepted 14 June 1999)

A stability problem is eliminated from the self-starter, which is a parabolic equation technique for generating initial conditions. The stability problem is associated with evanescent modes that correspond to nearly vertical propagation. It causes the solution to break down at certain frequencies for problems involving relatively low attenuation. The error is restricted to relatively short ranges unless the medium is strongly range dependent near the source. The stability problem is eliminated by placing the pole of the smoothing operator in the self-starter away from the eigenvalues of the depth operator. [S0001-4966(99)02510-2]

PACS numbers: 43.30.Bp, 43.30.Ma [DLB]

## INTRODUCTION

Parabolic equation techniques are useful for solving range-dependent propagation problems.<sup>1</sup> The self-starter is a parabolic equation technique for generating initial conditions.<sup>2</sup> This approach has been generalized to problems involving elastic,<sup>3</sup> poro-elastic,<sup>4</sup> and poro-acoustic<sup>5</sup> layers and handles shear waves, slow waves, and interface waves. It can be used as an efficient forward model for geoacoustic inversion problems.<sup>6</sup> In this paper, the self-starter is modified to eliminate a stability problem. The stabilized self-starter is described for the acoustic case but applies directly to the other cases.

The stability problem escaped detection until recently because it (a) only occurs near certain frequencies, (b) is restricted to problems involving relatively weak attenuation, and (c) only corrupts the solution at relatively short ranges unless the medium is strongly range dependent near the source. The stability problem is associated with eigenvalues of the depth operator that nearly vanish and correspond to nearly vertical propagation. Since the field is singular near the source, the self-starter is implemented by applying a smoothing operator to reduce the contribution of the evanescent modes. The stability problem is resolved by moving the pole of the smoothing operator away from the eigenvalues.

The stabilized self-starter is derived in Sec. I. Examples are presented in Sec. II to illustrate the stability problem and the stabilized self-starter.

## I. DERIVATION

In the far field, the normal mode solution of the Helmholtz equation is of the form,

$$p(r, z) = \sum_n \left( \frac{2\pi i}{k_n r} \right)^{1/2} \phi_n(z_0) \phi_n(z) \exp(ik_n r), \quad (1)$$

where  $p$  is the acoustic pressure,  $r$  is the range,  $z$  is the depth, and  $z_0$  is the source depth. The modes  $\phi_n$  and eigenvalues  $k_n^2$  satisfy the equation,

$$k_0^2(1+X)\phi_n = k_n^2\phi_n, \quad (2)$$

$$X \equiv k_0^{-2} \left( \rho \frac{\partial}{\partial z} \frac{1}{\rho} \frac{\partial}{\partial z} + k^2 - k_0^2 \right), \quad (3)$$

where  $\rho$  is the density,  $k$  is the wave number, and  $k_0$  is a representative wave number. We assume that the medium is truncated at a finite depth so that the spectrum is discrete.

By repeatedly applying the operator on the left side of Eq. (2), we obtain

$$(k_0^2(1+X))^m \phi_n = (k_n^2)^m \phi_n, \quad (4)$$

for any integer  $m$ . From Eq. (4) and the Taylor series representation of an arbitrary analytic function  $f$ , we obtain

$$f(k_0^2(1+X))\phi_n = f(k_n^2)\phi_n. \quad (5)$$

This equation can be used to derive various parabolic equation techniques, depending on the choice of  $f$ . From Eqs. (1) and (5) and the representation,

$$\delta(z-z_0) = \sum_n \phi_n(z_0) \phi_n(z), \quad (6)$$

we obtain

$$p(r, z) = \left( \frac{2\pi i}{k_0 r} \right)^{1/2} (1+X)^{-1/4} \exp(ik_0 r(1+X)^{1/2}) \times \delta(z-z_0). \quad (7)$$

Numerical difficulties associated with the singularity of the delta function can be avoided by integration. Multiplying and dividing Eq. (7) by the operator  $(1-\nu X)^2$  and applying a rational approximation, we obtain the self-starter

$$p(r, z) = \left( \frac{2\pi i}{k_0 r} \right)^{1/2} g(X) \xi(z), \quad (8)$$

$$g(X) \equiv \exp(ik_0 r) \prod_{j=1}^N \frac{1 + \alpha_{j,N} X}{1 + \beta_{j,N} X} \equiv (1-\nu X)^2 (1+X)^{-1/4} \exp(ik_0 r(1+X)^{1/2}), \quad (9)$$

$$\xi(z) \equiv (1-\nu X)^{-2} \delta(z-z_0), \quad (10)$$

where the parameter  $\nu$  is selected so that  $1-\nu X$  is invertible. The self-starter was implemented in Ref. 6 for the case  $\nu = -1$ . This choice is natural because the operator in Eq. (7) is a function of  $1+X$ . It is effective for many applications because the operator  $1+X$  is invertible for problems involving attenuation.

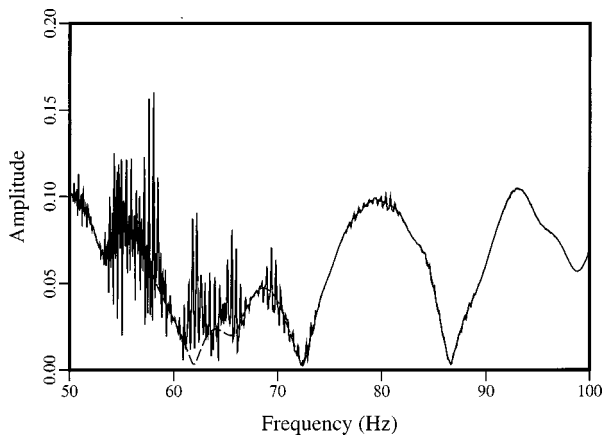


FIG. 1. Results for example A which is range independent. Parabolic equation solutions at  $r=1$  km and  $z=30$  m generated using the self-starter for the cases  $\nu=-1$  (solid curve) and  $\nu=1$  (dashed curve). The  $\nu=-1$  solution is unstable and contains spikes. The  $\nu=1$  solution is smooth. The displayed quantity is the amplitude of the acoustic pressure.

When the attenuation vanishes, the operator  $1+X$  is not invertible at certain frequencies. In this case, one of the eigenvalues vanishes and the solution of the Helmholtz equation is unbounded if the corresponding mode is excited. When the attenuation is weak, an eigenvalue may be sufficiently small to cause numerical difficulties. This problem can be avoided by selecting  $\nu$  appropriately. From Eq. (2), we obtain

$$(1-\nu X)\phi_n = \frac{(1+\nu)k_0^2 - \nu k_n^2}{k_0^2} \phi_n. \quad (11)$$

From Eq. (11), we would expect to encounter numerical problems if the condition,

$$k_n^2 \cong \frac{1+\nu}{\nu} k_0^2, \quad (12)$$

holds for some  $n$ . Many problems can be handled effectively by taking  $\nu=1$  and selecting  $k_0$  large enough so that  $k_0^2 \gg |k_n^2|$  for the eigenvalues that are close to the real line. The

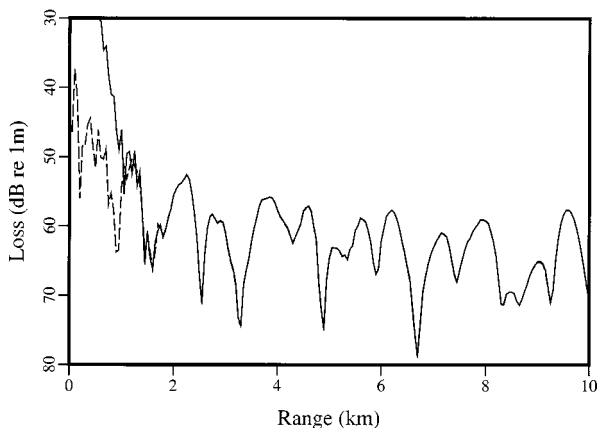


FIG. 2. Results at 57.6 Hz for example B, which is range independent. Parabolic equation solutions at  $z=30$  m generated using the self-starter for the cases  $\nu=-1$  (solid curve) and  $\nu=1$  (dashed curve). The  $\nu=1$  solution was determined to be accurate by convergence tests. The  $\nu=-1$  solution contains a large error at short ranges.

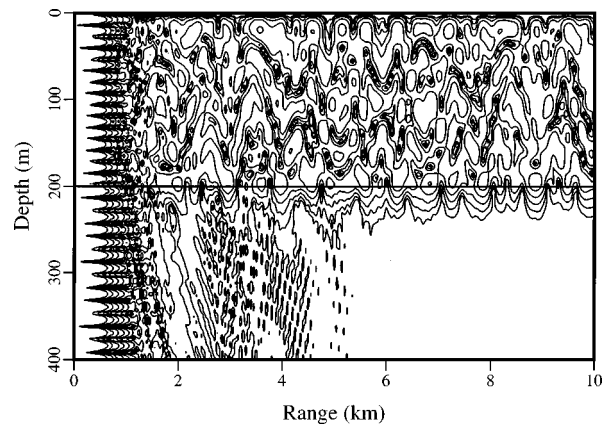


FIG. 3. Results at 57.6 Hz for example B, which is range independent. Parabolic equation solution generated using the self-starter for the case  $\nu=-1$ . The instability manifests itself as a highly excited evanescent mode for  $r < 1$  km.

constraint on  $k_0$  can be relaxed by taking  $\nu$  sufficiently small, or avoided by allowing  $\nu$  to be complex. If the attenuation is relatively weak, the eigenvalues are near the real line and the condition in Eq. (12) can be avoided by taking  $\nu=i$ . Although the fourth root in Eq. (9) is also nearly singular when an eigenvalue nearly vanishes, this factor does not cause numerical difficulties since  $g(X)$  is not required to be accurate near  $X=-1$ .

## II. EXAMPLES

In this section, we illustrate the errors that can be caused by the instability and show that the modification described in Sec. I resolves the problem. Each of the examples involves a source at  $z=100$  m in a shallow-water environment consisting of two homogeneous layers. The sound speed is 1500 m/s in the water column and 1700 m/s in the sediment, where the density is 1.5 g/cm<sup>3</sup> and the attenuation is 0.1 dB/ $\lambda$ . For the calculations, we take  $N=8$ , use a range step of 50 m, use a depth grid spacing of 1 m, truncate the domain with a pressure release boundary at  $z=2$  km, and construct the self-starter at  $r=50$  m.

Example A is a range-independent problem involving a 200-m-deep water column. The frequency dependence of the instability is illustrated in Fig. 1. The solution is smooth for the case  $\nu=1$ . The solution is corrupted by spikes for the case  $\nu=-1$ . The instability is worse at lower frequencies because the attenuation per unit length increases linearly with frequency. The amplitude of the spikes decreases above about 70 Hz, where the  $\nu=1$  and  $\nu=-1$  solutions begin to converge. The amplitude of the spikes can be reduced by increasing the truncation depth and introducing artificial attenuation deep in the sediment so that the offending eigenvalues move away from the origin into the complex plane.

For example B, we consider the problem of example A at 57.6 Hz, which corresponds to one of the spikes in Fig. 1. Results for  $\nu=1$  and  $\nu=-1$  appear in Fig. 2. The  $\nu=1$  solution was determined to be accurate by performing convergence tests on the truncation depth and other numerical parameters. The  $\nu=-1$  solution contains large errors for



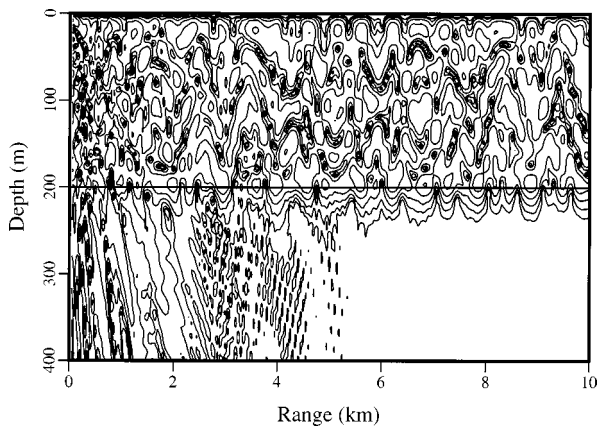


FIG. 4. Results at 57.6 Hz for example B, which is range independent. Parabolic equation solution generated using the self-starter for the case  $\nu = 1$ . The solution is not corrupted by an evanescent mode at short range.

$r < 1$  km but is accurate at long ranges. Contour plots of the solutions appear in Figs. 3 and 4. The  $\nu = -1$  solution is dominated by a single mode for  $r < 1$  km. Since the coefficient of this mode decays rapidly with range, the  $\nu = 1$  and  $\nu = -1$  solutions are nearly identical at long range.

Example B illustrates that the stability problem does not affect the solution at long range for range-independent problems. Example C illustrates that the stability problem can degrade the solution at all ranges when the medium is range dependent near the source. We modify the environment of examples A and B so that the ocean depth is 200 m at the source, increases linearly to 300 m at  $r = 2$  km, and is constant for  $r > 2$  km. Results for  $\nu = 1$  and  $\nu = -1$  appear in Fig. 5. The  $\nu = 1$  solution was determined to be accurate by performing convergence tests. The  $\nu = -1$  solution contains large errors for all ranges because the range dependence near the source causes information in the strongly excited evanescent mode to couple into the propagating modes. Since mode coupling occurs when the matrices in the parabolic equation are updated between range steps, the error in the  $\nu = -1$  solution depends on the range step size.

### III. CONCLUSION

An instability in the self-starter has been resolved by moving the pole of the smoothing operator away from the eigenvalues. The instability only occurs at certain frequencies for problems involving relatively weak attenuation. The instability only affects the solution at short ranges unless the

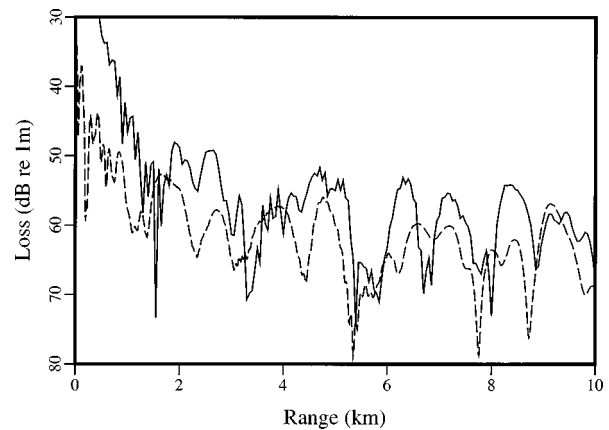


FIG. 5. Results at 57.6 Hz for example C, which involves a sloping ocean bottom near the source. Parabolic equation solutions at  $z = 30$  m generated using the self-starter for the cases  $\nu = -1$  (solid curve) and  $\nu = 1$  (dashed curve). The  $\nu = 1$  solution was determined to be accurate by convergence tests. The  $\nu = -1$  solution contains a large error at all ranges.

medium is strongly range dependent near the source. The stabilized self-starter can also be applied to parabolic equations for elastic, poro-elastic, and other media. We have found that  $\nu = 1$  is an effective choice for a wide range of acoustics problems. The constraint on the representative wave number can be avoided by taking  $\nu = i$ , which is an effective choice for problems involving widely distributed eigenvalues.

### ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The author thanks E. K. Scheer of the Woods Hole Oceanographic Institution for providing results that led to the detection of the stability problem.

- <sup>1</sup>F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994), pp. 343–412.
- <sup>2</sup>M. D. Collins, “A self-starter for the parabolic equation method,” *J. Acoust. Soc. Am.* **92**, 2069–2074 (1992).
- <sup>3</sup>M. D. Collins, “A two-way parabolic equation for elastic media,” *J. Acoust. Soc. Am.* **93**, 1815–1825 (1993).
- <sup>4</sup>M. D. Collins, W. A. Kuperman, and W. L. Siegmann, “A parabolic equation for poro-elastic media,” *J. Acoust. Soc. Am.* **98**, 1645–1656 (1995).
- <sup>5</sup>M. D. Collins, J. F. Lingeitch, and W. L. Siegmann, “Wave propagation in poro-acoustic media,” *Wave Motion* **25**, 265–272 (1997).
- <sup>6</sup>R. J. Cederberg and M. D. Collins, “Application of an improved self-starter to geoacoustic inversion,” *IEEE J. Ocean Eng.* **22**, 102–109 (1997).

# An efficient parabolic equation solution based on the method of undetermined coefficients

Dalcio K. Dacol, Michael D. Collins, and Joseph F. Lingeitch  
*Naval Research Laboratory, Washington, D.C. 20375*

(Received 28 September 1998; revised 4 June 1999; accepted 14 June 1999)

An implementation of the self-starter based on the method of undetermined coefficients is described and tested. For many problems, this parabolic equation technique for solving short range propagation problems provides an efficiency gain of an order of magnitude or more over the finite difference solution. With this forward model, it is possible to solve geoacoustic inverse problems in seconds on the current generation of desktop computers. The approach can be implemented for the inverse problem so that efficiency is essentially independent of the depth of the water column. [S0001-4966(99)02610-7]

PACS numbers: 43.30.Bp, 43.30.Ma, 43.30.Pc [DLB]

## INTRODUCTION

Normal mode and wave number integration techniques<sup>1</sup> for solving ocean acoustics problems are commonly implemented using two approaches. The finite difference solution is based on the discretization of the depth operator in the wave equation. The method of undetermined coefficients is based on expressing the solution in terms of analytic functions in homogenous layers and solving for a set of coefficients. The efficiency of finite difference solutions is relatively insensitive to the complexity of the depth dependence of the properties of the waveguide. The method of undetermined coefficients provides greater efficiency for problems involving a relatively small number of layers.

Parabolic equation techniques<sup>2</sup> are usually implemented using finite differences. In this paper, we describe an implementation of the self-starter<sup>3</sup> that is based on the method of undetermined coefficients. This parabolic equation technique can be used to efficiently solve geoacoustic inverse problems, which often require thousands of forward solves. When parabolic equation techniques were first applied to this problem, run times were on the order of an hour.<sup>4</sup> Run times were reduced to on the order of a minute by using a short range configuration and a finite difference implementation of the self-starter.<sup>5</sup> We further reduce run times to on the order of a second by using the undetermined coefficients implementation of the self-starter.

We describe the undetermined coefficients solution in Sec. I. We test robustness and accuracy in Sec. II. An efficient approach for handling the water column is described in Sec. III. The application to geoacoustic inversion is illustrated in Sec. IV.

## I. THE UNDETERMINED COEFFICIENTS SOLUTION

We consider propagation out to ranges over which the waveguide may be treated as stratified. In the far field, the normal mode solution is<sup>6</sup>

$$p(r, z) = \sum_m \left( \frac{2\pi i}{k_m r} \right)^{1/2} \phi_m(z_0) \phi_m(z) \exp(ik_m r), \quad (1)$$

where  $r$  is the range,  $z$  is the depth, and  $z_0$  is the source depth. The  $m$ th normal mode  $\phi_m$  and eigenvalue  $k_m^2$  satisfy

$$k_0^2(1+X)\phi_m = k_m^2\phi_m, \quad (2)$$

$$X = k_0^{-2} \left( \rho \frac{\partial}{\partial z} \frac{1}{\rho} \frac{\partial}{\partial z} + k^2 - k_0^2 \right), \quad (3)$$

where  $\rho$  is the density,  $k = \omega/c$  is the wave number,  $\omega$  is the circular frequency,  $c$  is the sound speed, and  $k_0$  is a representative wave number. The modes also satisfy the boundary conditions  $\phi_m(0) = \phi_m(z_M) = 0$ , where the truncation depth  $z = z_M$  is taken to be sufficiently large to prevent errors due to artificial reflections.

It follows from Eq. (2) that

$$f(X)\phi_m = f\left(\frac{k_m^2 - k_0^2}{k_0^2}\right)\phi_m, \quad (4)$$

for an arbitrary analytic function  $f(X)$ . Various choices for  $f(X)$  are used to derive parabolic equation techniques. The self-starter is based on the case,

$$f(X) = (1+X)^{-1/4} \exp(ik_0 r_0(1+X)^{1/2}), \quad (5)$$

where  $r_0$  is a fixed range. From Eqs. (1), (4), and (5) and the modal representation of the delta function,

$$\delta(z - z_0) = \sum_m \phi_m(z_0) \phi_m(z), \quad (6)$$

we obtain the self-starter,

$$p(r_0, z) = \left( \frac{2\pi i}{k_0 r_0} \right)^{1/2} f(X) \delta(z - z_0). \quad (7)$$

The solution of Eq. (7) is obtained by approximating  $f(X)$  with a rational function and then applying numerical techniques. We apply a rational approximation of the form,

$$g(X) = \frac{\sum_{j=0}^{n-2} \alpha_j X^j}{1 + \sum_{j=1}^n \beta_j X^j}. \quad (8)$$

The two additional terms in the denominator are included to guarantee that the singularity of the delta function is mollified. The coefficients are selected so that  $g(X) \cong f(X)$  for the

propagating modes and the evanescent modes are annihilated. A least squares approach for obtaining  $\alpha_j$  and  $\beta_j$  is described in Ref. 5. Although it is relatively easy to obtain these coefficients, this expression for  $g(X)$  is not suitable for the undetermined coefficients solution. Factoring the polynomial in the denominator and expanding in partial fractions, we obtain

$$g(X) = \sum_{j=1}^n \frac{\gamma_j}{1 + \delta_j X}. \quad (9)$$

Substituting this approximation for  $f(X)$  into Eq. (7), we obtain

$$p(r_0, z) = \left( \frac{2\pi i}{k_0 r_0} \right)^{1/2} \sum_{j=1}^n \hat{p}_j(z), \quad (10)$$

$$\hat{p}_j(z) = \frac{\gamma_j}{1 + \delta_j X} \delta(z - z_0). \quad (11)$$

Rearranging Eq. (11), we obtain

$$\left( \rho \frac{d}{dz} \frac{1}{\rho} \frac{d}{dz} + k^2 - h_j^2 \right) \hat{p}_j = \sigma_j \delta(z - z_0), \quad (12)$$

$$h_j^2 = k_0^2 (1 - \delta_j^{-1}), \quad (13)$$

$$\sigma_j = k_0^2 \delta_j^{-1} \gamma_j, \quad (14)$$

and conclude that  $\hat{p}_j$  corresponds to the wave number spectrum for the horizontal wave number  $h_j$  and the source amplitude  $\sigma_j$ . The solution of Eq. (12) may be obtained using the method of undetermined coefficients, which is composed of exponential functions in each layer. The coefficients of these functions are obtained by enforcing the source condition, the boundary conditions, and interface conditions for continuity of pressure and particle velocity.

## II. ROBUSTNESS AND ACCURACY

This section contains examples of the undetermined coefficients solution that illustrate robustness and accuracy. All of the examples involve a 35-Hz source located in a homogeneous water column in which  $c = 1500$  m/s.

For example A, we take  $z_0 = 75$  m and consider a homogeneous waveguide of thickness 200 m. Although  $p(r, z)$  is differentiable at  $z = z_0$ , the terms on the right side of Eq. (10) contain cusps at  $z = z_0$ . To illustrate that these cusps cancel reliably in numerical solutions, we consider the function,

$$F(z) = (1 + X)^{-1} (1 - X)^{-1} \delta(z - z_0). \quad (15)$$

Since  $X$  is a second order differential operator, the application of the first inverse operator to the delta function produces a function that is continuous but has a cusp at  $z = z_0$ . The cusp is eliminated by applying the second inverse operator to obtain  $F(z)$ , which has two continuous derivatives at  $z = z_0$ . It is well known that the numerical implementation of this solution is robust unless one of the inverse operators is singular or nearly singular. The undetermined coefficients solution of Eq. (15) is based on expanding in partial fractions to obtain

$$F(z) = \frac{1}{2} F_+(z) + \frac{1}{2} F_-(z), \quad (16)$$

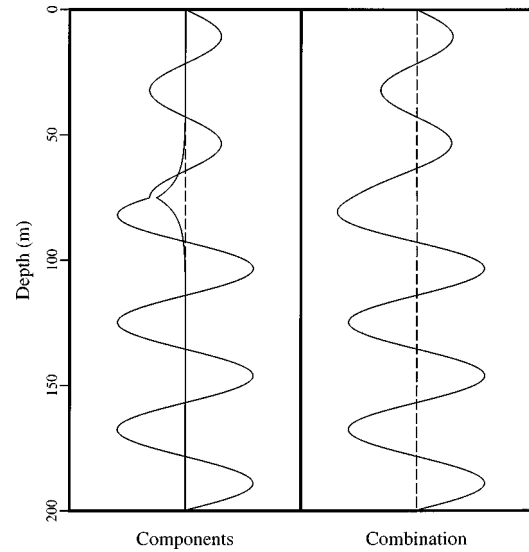


FIG. 1. Results for example A, which tests the reliability of the numerical implementation of the partial fractions expansion. The components  $F_+(z)$  and  $F_-(z)$  contain cusps at the source depth. Their combination  $F(z)$  appears smooth.

$$F_{\pm}(z) = (1 \pm X)^{-1} \delta(z - z_0). \quad (17)$$

It is not obvious that the numerical implementation of this form of the solution is robust since the cusps in  $F_+(z)$  and  $F_-(z)$  must cancel additively. Numerical solutions for  $F_+(z)$  and  $F_-(z)$  and their sum appear in Fig. 1. The components contain cusps at  $z = z_0$ , but their sum appears smooth.

For example B, we take  $r_0 = 500$  m and  $z_0 = 290$  m and consider a 300-m deep water column that overlies a sediment in which  $c = 1700$  m/s,  $\rho = 1.5$  g/cm<sup>3</sup>, and the attenuation is 0.5 dB/λ. This example was used in Ref. 5 to illustrate the accuracy of the finite difference implementation of the self starter. We take  $n = 14$  and use the least squares approach of Ref. 5 to obtain the coefficients of  $g(X)$ . We place 80 accuracy constraints on  $-0.7 \leq X \leq 0$  and one stability constraint at  $X = -1$ . The accuracy of the rational approximation is illustrated in Fig. 2. The real and imaginary parts of  $p(r_0, z)$  appearing in Fig. 3 are in agreement with the reference solution, which was obtained by integrating the wave number spectrum.

## III. PRECOMPUTATION IN THE WATER COLUMN

The undetermined coefficients solution can be implemented by approximating the entire medium in terms of homogeneous layers. In this section, we consider an alternative approach that is more efficient for geoacoustic inverse problems. It is based on the concept of precomputation, which has proven to be effective for other problems in ocean acoustics.<sup>7</sup>

The sound speed profile is assumed to be known in the water column  $z < z_w$ , which is treated as an inhomogeneous layer. Before performing a search for the acoustic parameters in the sediment layers, the finite difference approach is applied to obtain the Dirichlet and Neumann solutions  $D_j$  and  $N_j$ , which satisfy Eq. (12) in the water column and the conditions,

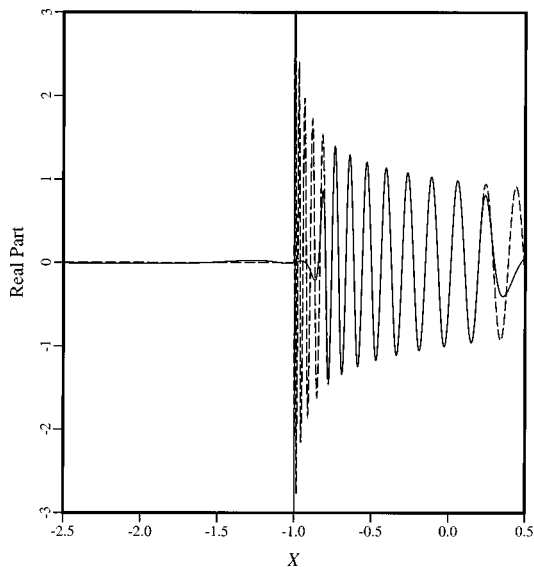


FIG. 2. Results for example B. Comparison of the rational function  $g(X)$  (solid curve) and the function  $f(X)$  that it approximates (dashed curve). The agreement is excellent in the region  $-0.7 \leq X \leq 0$ , where the accuracy constraints were imposed. The magnitude of  $g(X)$  is small in the region  $X < -1$ , which corresponds to evanescent modes.

$$D_j(z_w) = \frac{dN_j}{dz}(z_w) = D_j(0) = N_j(0) = 0. \quad (18)$$

In the water column, the wave number spectrum is of the form

$$\hat{p}_j = A_{1,j}D_j + B_{1,j}N_j, \quad (19)$$

where the coefficients satisfy the source condition  $A_{1,j} + B_{1,j} = 1$ . In the  $i$ th homogeneous layer in the ocean bottom, the wave number spectrum is of the form

$$\hat{p}_j = A_{i,j} \exp(iK_{i,j}z) + B_{i,j} \exp(-iK_{i,j}z), \quad (20)$$

$$K_{i,j} = (k_i^2 - h_j^2)^{1/2}, \quad (21)$$

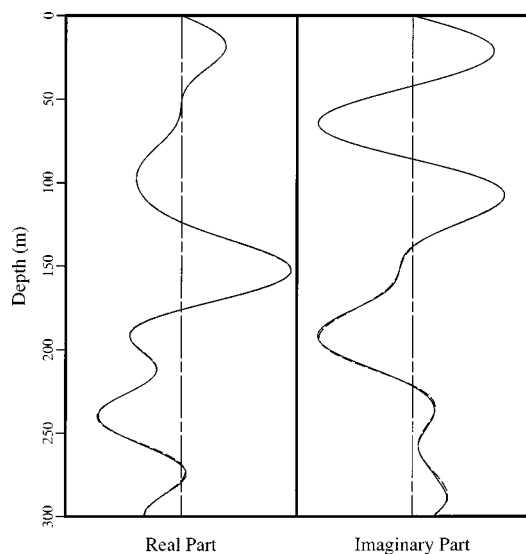


FIG. 3. Results for example B. The real and imaginary parts of  $p(r_0, z)$  in the water column. The self-starter solution (solid curves) is in agreement with the spectral solution (dashed curves).

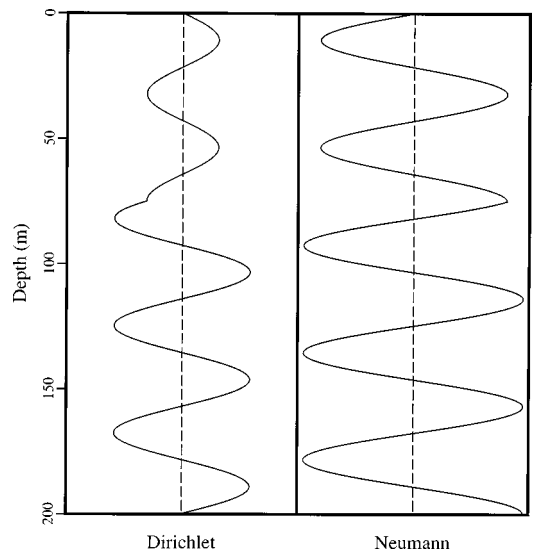


FIG. 4. Results for example C. The Dirichlet and Neumann solutions of Eq. (12) for  $h_j = 0$ . Note the cusps at the source depth  $z = 75$  m.

where  $k = k_i$  in the  $i$ th layer.

The coefficients  $A_{i,j}$  and  $B_{i,j}$  are determined by enforcing the boundary condition  $\hat{p}_j(z_M) = 0$  and continuity of  $p_j$  and  $\rho^{-1} dp_j/dz$  at the interfaces between layers. The boundary condition  $\hat{p}_j(0) = 0$  is automatically satisfied because  $D_j$  and  $N_j$  satisfy this condition. Since it is not necessary to recompute  $D_j$  and  $N_j$  when the sediment parameters are modified, the precomputation approach is efficient for solving geoacoustic inverse problems that involve a relatively small number of sediment layers. One of the advantages of this approach for solving the inverse problem is that efficiency is essentially independent of the depth of the water column. Example C illustrates the Dirichlet and Neumann solutions for a 35-Hz source at  $z = 75$  m in a 200-m deep water column in which  $c = 1500$  m/s. The solutions of Eq. (12) appearing in Fig. 4 were generated using the finite difference method for the case  $h_j = 0$ .

#### IV. APPLICATION TO GEOACOUSTIC INVERSION

In this section, we illustrate the application of the undetermined coefficients solution as the forward model in geoacoustic inversion. The experimental configuration involves a vertical array of receivers placed at a range on the order of ten wavelengths. The entries of the parameter vector  $\mathbf{x}$  describe the sediment. The pressure vector  $\mathbf{p}$  contains the complex pressures measured at the receivers. The goal of the inverse problem is to determine the parameter vector that minimizes the energy,

$$E(\mathbf{x}) = \sqrt{1 - |\hat{\mathbf{p}}^*(\mathbf{x}) \cdot \hat{\mathbf{p}}(\mathbf{x}_0)|^2}, \quad (22)$$

where  $\hat{\mathbf{p}}(\mathbf{x})$  is the normalized pressure vector corresponding to the parameter set  $\mathbf{x}$ , the superscript asterisk denotes complex conjugation,  $\mathbf{x}_0$  corresponds to the desired geoacoustic data, and  $\hat{\mathbf{p}}(\mathbf{x}_0)$  corresponds to the measured acoustic data.

We solve the optimization problem using the simulated annealing algorithm discussed in Ref. 4 and the coordinate rotation discussed in Ref. 8. The purpose of the coordinate rotation is to improve the efficiency and reliability of the



TABLE I. Parameter values and bounds for example D.

$i$	$x_i$	$A_i$	$B_i$	units
1	1625	1550	1750	m/s
2	1.2	1.1	1.5	g/cm <sup>3</sup>
3	0.1	0.05	0.3	dB/λ
4	1700	1600	1850	m/s
5	1.3	1.1	1.5	g/cm <sup>3</sup>
6	0.2	0.05	0.3	dB/λ
7	1800	1650	1850	m/s
8	1.5	1.1	1.5	g/cm <sup>3</sup>
9	0.3	0.1	0.5	dB/λ

parameter search for problems involving long valleys (i.e., coupling between parameters). The parameter space is

$$\Omega = \{\mathbf{x} | A_i < x_i < B_i\}, \quad (23)$$

where  $A_i$  and  $B_i$  are the bounds on the  $i$ th parameter. The basis vectors  $\{\mathbf{v}_i\}$  for the optimal rotation are the eigenvectors of the covariance matrix,

$$K = \int_{\Omega} \nabla E (\nabla E)^t d\Omega. \quad (24)$$

The integral in Eq. (24) may be evaluated efficiently with the Monte Carlo method,<sup>9,10</sup> even when the dimension of  $\mathbf{x}$  is large. Since the integral,

$$I_i = \int_{\Omega} (\mathbf{v}_i \cdot \nabla E)^2 d\Omega, \quad (25)$$

is stationary with respect to rotations of  $\mathbf{v}_i$ , the rotated coordinate axes are optimally aligned with the most prominent valleys of the parameter landscape. The eigenvectors that correspond to the largest eigenvalues indicate the most resolvable underlying parameters.

Example D involves a 35-Hz source at  $z=290$  m in a 300-m deep water column in which  $c=1500$  m/s. The ocean bottom consists of two 10-m-thick sediment layers overlying a homogeneous half-space basement. The parameters in the upper layer are the sound speed  $x_1$ , density  $x_2$ , and attenuation  $x_3$ . The parameters in the lower layer are the sound speed  $x_4$ , density  $x_5$ , and attenuation  $x_6$ . The parameters in the basement are the sound speed  $x_7$ , density  $x_8$ , and attenuation  $x_9$ . The values and ranges of the parameters are given in Table I. For this example, we take  $r_0=500$  m, use the rational approximation of example B, and place the receivers at  $z_j = (-10 + 20j)$  m for  $1 \leq j \leq 15$ . The eigenvectors appear in Fig. 5. The first eigenvector indicates strong coupling between the sound speeds in the layers. The other eigenvectors indicate significant coupling between other parameters.

The parameter search in the original coordinates appears in Fig. 6. Two of the parameters converge. Several parameters misleadingly lock in to the wrong values. The parameter search in rotated coordinates appears in Fig. 7. Five of the parameters converge. There is no false convergence. The parameter searches involved 200 iterations and 1800 forward solves and required 0.64 s on a Silicon Graphics computer with a MIPS R10000 processor. The run time was 3.02 s on a computer that required 37 s for a related problem involving 2200 forward solves using the finite difference implementa-

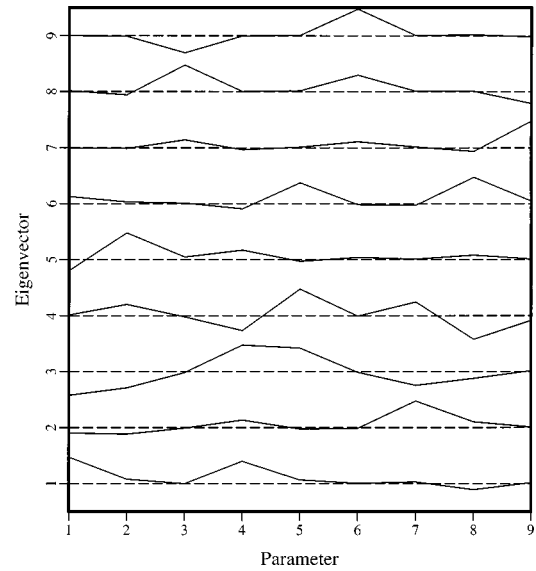


FIG. 5. The eigenvectors of the covariance matrix for example D, which are ordered according to decreasing eigenvalue. The first few eigenvectors correspond to the most resolvable parameters for the chosen source frequency and experimental configuration. Since the eigenvectors consist of combinations of different parameters, the parameter landscape contains long valleys.

tion of the self-starter.<sup>5</sup> For this shallow water problem, a gain in efficiency of about a factor of 10 is achieved by going from finite differences to undetermined coefficients. For problems involving relatively deep water columns, the efficiency gain is greater when the implementation involves the precomputation approach described in Sec. III.

## V. CONCLUSION

A parabolic equation solution based on the method of undetermined coefficients has been developed. This implementation of the self-starter is based on known analytic functions in homogeneous layers. The coefficients are determined by enforcing interface, boundary, and source conditions and

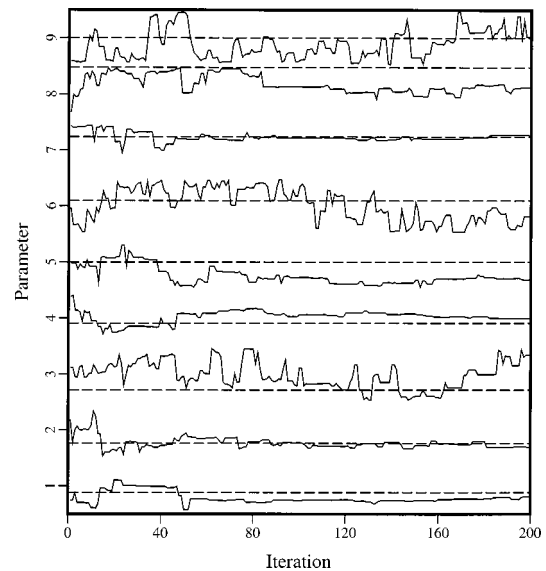


FIG. 6. Geoacoustic inversion results for example D obtained by searching in the original coordinates. The convergence is good for parameters 2 and 7. Parameters 1, 4, 5, and 8 lock in to the wrong values.

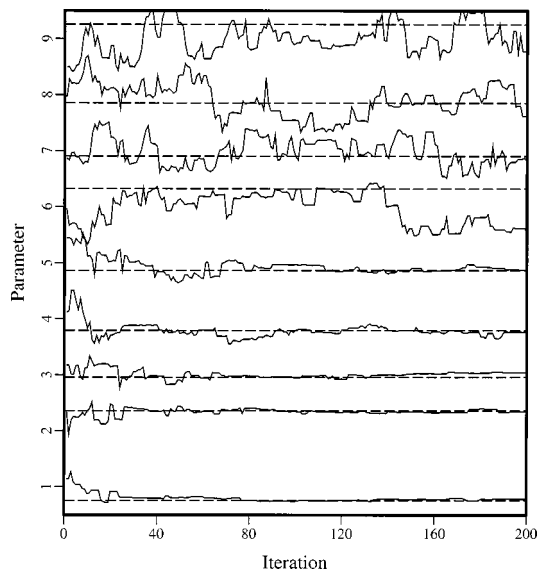


FIG. 7. Geoacoustic inversion results for example D obtained by searching in rotated coordinates. The convergence is good for the first five parameters.

solving a system of equations. The approach is efficient for solving geoacoustic inverse problems, especially with the precomputation approach that can be applied when the parameters of the water column are known. Inverse problems

can be solved in seconds on the current generation of desktop computers with the undetermined coefficients solution as the forward model.

#### ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

- <sup>1</sup>F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994), pp. 203–341.
- <sup>2</sup>*Ibid.*, pp. 343–412.
- <sup>3</sup>M. D. Collins, “A self-starter for the parabolic equation method,” *J. Acoust. Soc. Am.* **92**, 2069–2074 (1992).
- <sup>4</sup>M. D. Collins, W. A. Kuperman, and H. Schmidt, “Nonlinear inversion for ocean-bottom properties,” *J. Acoust. Soc. Am.* **92**, 2770–2783 (1992).
- <sup>5</sup>R. J. Cederberg and M. D. Collins, “Application of an improved self-starter to geoacoustic inversion,” *IEEE J. Ocean Eng.* **22**, 102–109 (1997).
- <sup>6</sup>F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994), pp. 272–275.
- <sup>7</sup>W. A. Kuperman, M. B. Porter, J. S. Perkins, and R. B. Evans, “Rapid computation of acoustic fields in three-dimensional ocean environments,” *J. Acoust. Soc. Am.* **89**, 125–133 (1991).
- <sup>8</sup>M. D. Collins and L. Fishman, “Efficient navigation of parameter landscapes,” *J. Acoust. Soc. Am.* **98**, 1637–1644 (1995).
- <sup>9</sup>P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, New York, 1984), pp. 384–393.
- <sup>10</sup>M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986), pp. 89–116.

# A phase-compensated distorted wave Born approximation representation of the bistatic scattering by weakly scattering objects: Application to zooplankton

Dezhang Chu<sup>a)</sup>

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution,  
Woods Hole, Massachusetts 02543

Zhen Ye

Department of Physics, National Central University, Chung-li, Taiwan, Republic of China

(Received 13 November 1998; revised 26 May 1999; accepted 8 June 1999)

The distorted wave Born approximation (DWBA) method has been successfully used in modeling the acoustic backscattering by weakly scattering zooplankton [Stanton *et al.*, *J. Acoust. Soc. Am.* **94**, 3463–3472 (1993), Wiebe *et al.*, *IEEE J. Ocean. Eng.* **22**(3), 445–464 (1997)]. However, the previously developed DWBA model ignores the imaginary part of the scattering amplitude and thus results in a zero-extinction cross section. As a consequence, the model fails to predict the scattering-induced attenuation which could be important under certain circumstances. In this paper, a phase-compensated DWBA-based approximation is presented. The improved method allows us to compute not only the scattering strength but also the acoustic attenuation. The new scattering model is validated by comparing with the existing exact solution for certain representative finite objects. The results from this study can be applied to bioacoustic applications where the attenuation due to scattering and/or multiple scattering by zooplankton is relevant, and where this information might be used to infer the acoustic properties of live animals. © 1999 Acoustical Society of America. [S0001-4966(99)00410-5]

PACS numbers: 43.30.Ft, 43.30.Xm, 43.30.Gr, 43.20.Fn [DLB]

## INTRODUCTION

Mathematical modeling of the acoustical scattering by individual zooplankton has evolved from the simplest sphere model<sup>1</sup> to the recent distorted wave Born approximation (DWBA),<sup>2,3</sup> which is capable of describing the acoustic scattering by arbitrarily shaped inhomogeneous scatterers. The DWBA model has been successfully applied to both laboratory and shipboard experimental data,<sup>3,4</sup> as well as field data.<sup>5</sup> However, when the scattering-induced attenuation is noticeable and cannot be ignored, such as in the case of scattering by densely aggregated zooplankton, the conventional DWBA fails to describe the scattering characteristics. To illustrate this, consider a plane wave propagating through a random medium as shown in Fig. 1. The attenuation through the medium can be described in terms of the extinction cross section  $\sigma_e(\mathbf{p})$ :<sup>6</sup>

$$I_s(z) = I_0 e^{-z} \int_{\mathbf{p}} \rho_{\mathbf{p}} \sigma_e(\mathbf{p}) d\mathbf{p}, \quad (1)$$

where  $I_0$  is the original acoustic intensity at  $z=0$  while  $I_s$  is the observed intensity at  $z$ .  $\rho(\mathbf{p})$  is the number of scatterers in a unit volume with parameter  $\mathbf{p}$  referring to the properties of the aggregated scatterers. According to the forward-scattering theorem,<sup>7</sup> the extinction cross section  $\sigma_e(\mathbf{p})$  can be expressed in terms of the imaginary part of the scattering amplitude in the forward direction as

$$\sigma_e = \text{Im} \left( \frac{4\pi f_{\mathbf{p}}(\hat{\mathbf{o}})}{k} \right), \quad (2)$$

where  $k$  is the acoustic wave number of the incident wave in the surrounding medium, and  $f_{\mathbf{p}}(\hat{\mathbf{o}})$  is the forward-scattering amplitude of the individual scatterer with parameter  $\mathbf{p}$ . Equation (2) indicates that the extinction cross section  $\sigma_e$  is proportional to the imaginary part of the scattering amplitude in the forward direction. However, because of its inherent deficiency, the DWBA fails to predict the imaginary part of the scattering amplitude even for some regularly shaped targets such as spheres. To overcome this shortcoming, a modified DWBA model is proposed in this paper, which adds a heuristic phase compensation term. The phase-compensated DWBA solution is able to estimate the extinction cross section using the aforementioned forward-scattering theorem. The paper is organized in the following way. Section I briefly reviews the Born and DWBA models, while Sec. II provides the theoretical development of the phase-compensated DWBA model. In Sec. III, the results derived in Sec. II are extended to a more general case and then applied to zooplankton scattering. Simulation results are presented and discussed. Finally, conclusions are drawn in Sec. VI.

## I. BORN APPROXIMATION (BA) AND DISTORTED WAVE BORN APPROXIMATION (DWBA)

Before introducing the Born approximation (BA) and distorted wave Born approximation (DWBA), consider a unit plane wave,  $e^{i\mathbf{k}\cdot\mathbf{r}}$ , impinging on a volumetric foreign body

<sup>a)</sup>Electronic mail: dchu@whoi.edu

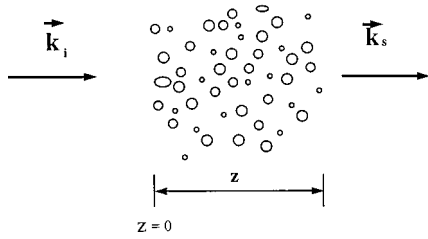


FIG. 1. Scattering by a cloud of randomly distributed fluid scatterers due to a plane incident wave.

(scatterer),  $V$ , whose acoustic properties are different from those of the surrounding medium. The integral equation representation of the scattering by such a volume is given by<sup>6</sup>

$$P_{\text{scat}} = \int_V [k^2 \gamma_k P_{\text{int}}(\mathbf{r}_0) G(\mathbf{r}|\mathbf{r}_0) + \gamma_\rho \nabla P_{\text{int}}(\mathbf{r}_0) \cdot \nabla G(\mathbf{r}|\mathbf{r}_0)] dV_0, \quad (3)$$

where  $P_{\text{int}}(\mathbf{r}_0)$  and  $G(\mathbf{r}|\mathbf{r}_0)$  are internal field and free-space Green's function, respectively, with  $\mathbf{r}_0$  representing the position vector within the volume  $V$ .  $\gamma_k$  and  $\gamma_\rho$  are defined as

$$\gamma_k = \frac{\kappa_1 - \kappa}{\kappa_1} = \frac{1 - gh^2}{gh^2}, \quad \gamma_\rho = \frac{\rho_1 - \rho}{\rho_1} = \frac{g-1}{g}, \quad (4)$$

where  $\kappa = 1/\rho c^2$  is the compressibility and the subscript “1” refers to the inhomogeneous volume,  $\rho$  and  $c$  are respectively density and sound speed in the surrounding medium.  $g = \rho_1/\rho$  and  $h = c_1/c$  are density and sound-speed contrasts of the inhomogeneous volume to the surrounding medium, respectively. Since the internal field  $P_{\text{int}}$  is unknown, the exact scattering field has to be obtained by solving the integral equation. In reality, most scatterers have complex shapes with inhomogeneous acoustic properties within their bodies; it is nearly impossible to obtain exact solutions for scattering by such objects. Approximation schemes have to be resorted to in order to estimate the acoustic scattering response. In particular, if the internal field in Eq. (3) can be represented by an analytical function, the integral equation reduces to an integral that can be evaluated directly.

### A. Born approximation

For weakly scattering  $g \approx 1$  and  $h \approx 1$ , which are the cases for most zooplankton applications, one of the commonly used approximations is the well-known Born approximation (BA).<sup>8</sup> In the BA, the internal field  $P_{\text{int}}(\mathbf{r}_0)$  in Eq. (3) is replaced with the incident field,  $e^{i\mathbf{k}_i \cdot \mathbf{r}_0}$ , since it is very close to the incident field with the absence of scatterers. To evaluate the performance of the BA, we compare it with the exact solutions for backscattering by a fluid sphere. The scattering geometry is shown in Fig. 2, where  $\theta_s = 180^\circ$  for backscattering. Substituting approximate internal field  $P_{\text{int}}(\mathbf{r}_0) = e^{i\mathbf{k}_i \cdot \mathbf{r}_0}$  into Eq. (3) and using the free-space Green's function

$$G(\mathbf{r}|\mathbf{r}_0) = e^{ik|\mathbf{r}-\mathbf{r}_0|}/4\pi|\mathbf{r}-\mathbf{r}_0| \approx e^{ikr} e^{-i\mathbf{k}_s \cdot \mathbf{r}_0}/4\pi r, \quad \text{for } |\mathbf{r}| \gg |\mathbf{r}_0|,$$

where  $|\mathbf{k}_i| = |\mathbf{k}_s| = k$ , we obtain

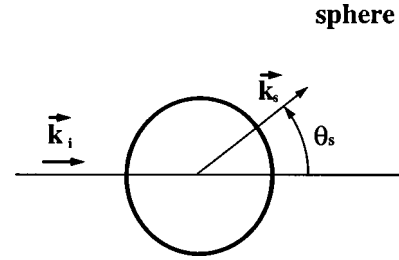


FIG. 2. Geometry of the bistatic scattering by a sphere.  $\theta_s = 0^\circ$  for forward scattering and  $\theta_s = 180^\circ$  for backscattering.

$$P_{\text{scat}}^{\text{BA}} = \frac{e^{ikr}}{r} \frac{k^2}{4\pi} \int_V (\gamma_k + \hat{\mathbf{k}}_s \cdot \hat{\mathbf{k}}_i \gamma_\rho) e^{i(\mathbf{k}_i - \mathbf{k}_s) \cdot \mathbf{r}_0} dV_0 = \frac{e^{ikr}}{r} f_{\text{scat}}^{\text{BA}}, \quad (5)$$

where the symbol “ $\hat{\cdot}$ ” stands for the unit vector,  $f_{\text{scat}}^{\text{BA}}$  is the scattering amplitude based on the BA and is found to be [Appendix, (A22)]

$$f_{\text{scat}}^{\text{BA}}(\theta_s) = k^2 a^3 (\gamma_k + \gamma_\rho \cos \theta_s) \frac{j_1(2ka \sin(\theta_s/2))}{2ka \sin(\theta_s/2)}, \quad (6)$$

where  $j_1(x)$  is the spherical Bessel function of order 1 and the scattering angle  $\theta_s$  is the angle between the incident and scattering directions depicted in Fig. 2. The comparison for the backscattering ( $\theta_s = 180^\circ$ ) between the exact partial wave (modal series) solution<sup>9</sup> and the BA with different density and sound-speed contrasts  $g$  and  $h$  is given in Fig. 3. To make the results more representative, we have plotted the scattering form function  $f_\infty$  (a dimensionless function) instead of the scattering amplitude. Since for a finite object the form function and the scattering amplitude are different only by a scaling constant,  $a/2$  for a sphere, the term “scattering amplitude” will be used for discussing scattering while the form function will be used for plotting. Three observations can be found by inspecting Fig. 3: (1) for  $ka \ll 1$ , BA provides very good fits in all cases. This is because when the acoustic wavelength is much longer than the object dimension, the internal field is essentially the same as that when the scatterer is absent; (2) the phase difference increases as sound-speed contrast  $h$  and  $ka$  increase. This is due to the fact that the phase difference is directly related to the travel time difference determined by the sound-speed contrast and has a larger impact as  $ka$  increases; (3) the amplitude difference increases with  $g$  and  $h$  but maintains a regular oscillatory pattern over the entire  $ka$  range. This phenomenon is expected since larger  $g$  and  $h$  imply stronger scattering, hence, the weak scattering assumption, on which the BA is based, may be violated.

### B. Distorted wave born approximation

Apparently, the BA cannot provide satisfactory results even for a moderate  $ka$ . It can be shown that the amplitude mismatch is due to the exclusion of higher-order internal waves (multiple bounces within the scatterer). Since the characteristics of the higher-order scattering strongly depend on the shape and orientation of the scatterer, as well as the



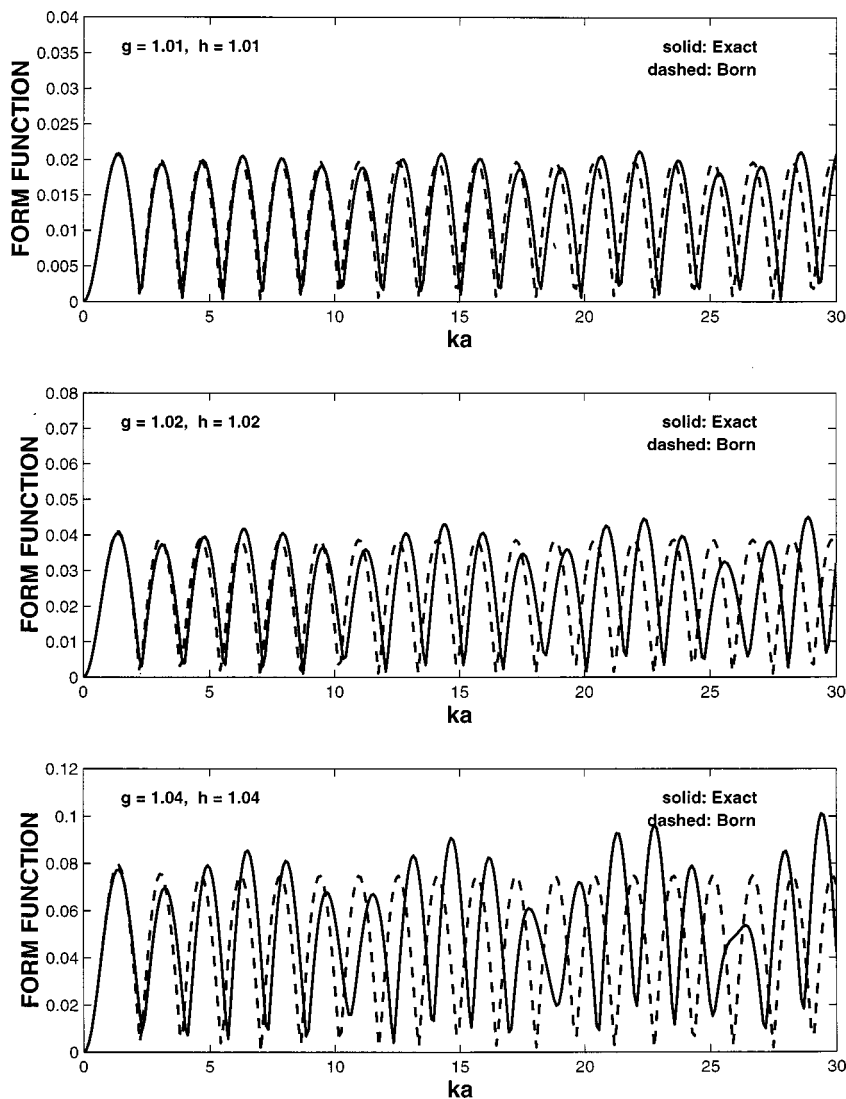


FIG. 3. Comparison of the back-scattering form function ( $\theta_s = 180^\circ$ ) of a fluid sphere between the exact solution and the Born approximation (BA) with different combinations of the density and sound-speed contrasts ( $g$  and  $h$ ). The solid lines are computed from the exact solution (Ref. 9) and the dashed lines are computed from Eq. (6).

scattering geometry, it is very difficult to account for this amplitude mismatch. The phase mismatch, however, is essentially due to the sound-speed contrast and the scatterer's dimension in the direction of the incident wave, and is relatively easy to characterize. The distorted wave Born approximation (DWBA) is introduced to compensate such a phase mismatch. Mathematically, the DWBA replaces the internal field  $P_{\text{in}}(\mathbf{k}_0)$  with the same function form as in the BA, but further replaces the wave vector  $\mathbf{k}_i$  in the phase by  $\mathbf{k}_1$ , representing a distorted wave. In addition, a corresponding modification of the free-space Green's function is included. For such a distorted Green's function, the wave number in its exponential within the inhomogeneous volume is  $|\mathbf{k}_s| = k_1 = k/h$ . With these modifications, the scattering amplitude using the DWBA can be expressed as

$$\begin{aligned}
 f_{\text{scat}}^{\text{DWBA}}(\theta_0) &= \frac{k_1^2}{4\pi} \int_V (h^2 \gamma_k + \hat{k}_i \cdot \hat{k}_s \gamma_\rho) e^{i(\mathbf{k}_i - \mathbf{k}_s) \cdot \mathbf{r}_0} dV \\
 &= k_1^2 a^3 (h^2 \gamma_k + \gamma_\rho \cos \theta_s) \\
 &\quad \times \left[ \frac{j_1(2k_1 a \sin(\theta_s/2))}{2k_1 a \sin(\theta_s/2)} \right]. \quad (7)
 \end{aligned}$$

The only major difference between Eq. (6) and Eq. (7) is that

the wave numbers  $k'$ 's in the square-bracketed term are replaced by  $k_1$ 's. Although  $|k_1| \approx k$  for weakly scattering, a slight change from  $k$  to  $k_1$  in the argument of the spherical Bessel function gives rise to a significant difference in the frequency response of the scattering, especially for large  $ka$ . Figure 4 shows the comparison between the exact solution and the DWBA for backscattering ( $\theta_s = 180^\circ$ ). In Fig. 4, all simulation parameters are kept the same as in Fig. 3. Clearly, the phase difference has been greatly reduced. For the case of the weak scattering,  $g$  and  $h$  are close to unity ( $g = h = 1.01$ ); the agreement between the exact solution and the DWBA is nearly perfect. Because of its inherent advantage, the DWBA is able to describe the scattering by objects with complex shapes, orientations, and inhomogeneous material properties, and has been successfully applied to the zooplankton scattering.<sup>2-5</sup>

## II. PHASE-COMPENSATED DWBA

Careful inspection of Fig. 4 reveals that for a weakly scattering object, the modulus of the scattering amplitude computed from the DWBA agrees reasonably well with the exact solution. Since Eq. (7) provides only the real part of the scattering amplitude, it indicates that the real component

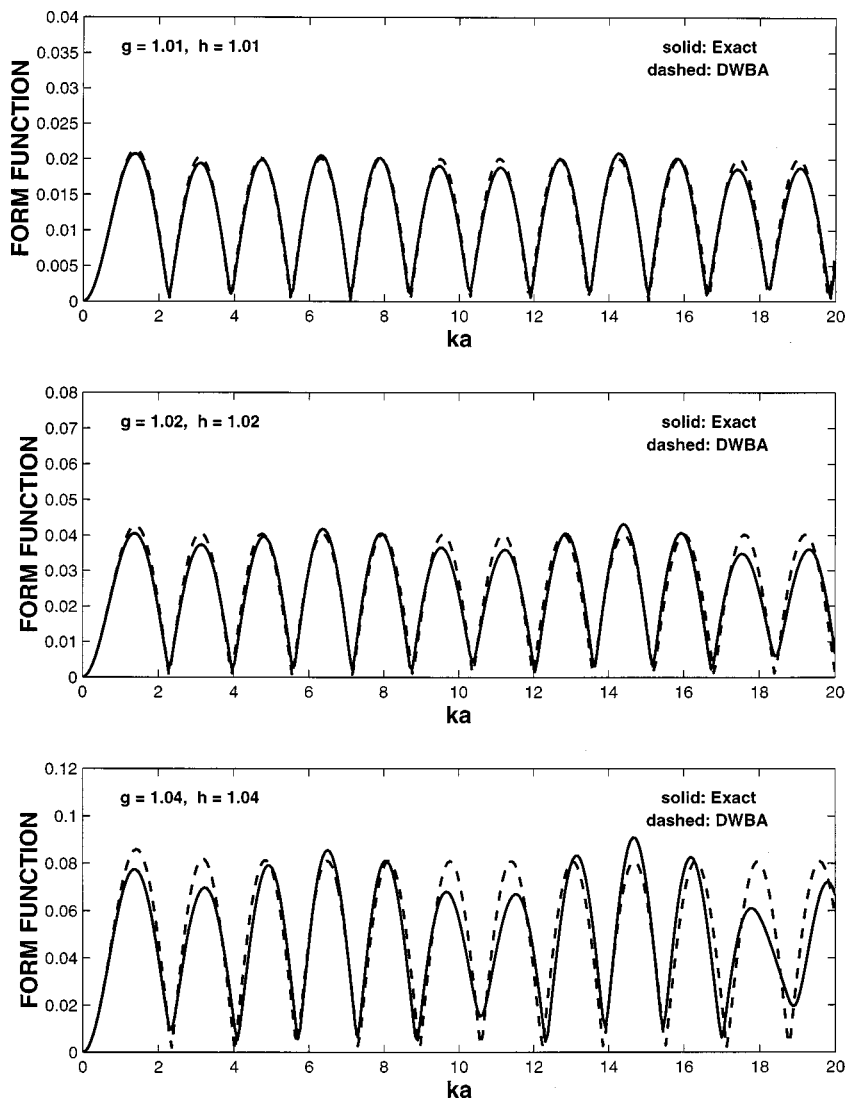


FIG. 4. Comparison of the backscattering form function ( $\theta_s = 180^\circ$ ) of a fluid sphere between the exact solution and the distorted wave Born approximation (DWBA) with different combinations of the density and sound-speed contrasts ( $g$  and  $h$ ). The solid lines are computed from the exact solution (Ref. 9) and the dashed lines are computed from Eq. (7).

of the scattering amplitude dominates the scattering process, while the imaginary component is much smaller and is ignored in the DWBA model. For the forward scattering,  $\mathbf{k}_i = \mathbf{k}_s$ , the integrand term  $e^{i(\mathbf{k}_i - \mathbf{k}_s) \cdot \mathbf{r}}$  in both Eq. (5) and Eq. (7) is equal to 1, and the resultant scattering amplitude  $f_{\text{scat}}^{\text{DWBA}}(0)$ , regardless of the shape and orientation of the object, is always a real function. For the bistatic scattering, the imaginary part is canceled out due to the symmetric geometry of a sphere. Clearly, in its present form, the DWBA cannot predict the scattering-induced attenuation characterized by the imaginary part of the scattering amplitude used in Eq. (2). For a general 3-D scattering problem, it is not easy, at least not obvious, to find a complex scattering amplitude representation which is capable of describing both modulus and phase reasonably well.

Without losing generality, we start with the simplest 1-D problem as shown in Fig. 5: an infinite medium containing a finite homogeneous layer with a thickness of  $2H$ . First, we use the DWBA method to derive a solution for this 1-D problem. Under DWBA, the internal field  $P_{\text{int}}(z_0)$  and its Green's function  $G(z|z_0)$  can be expressed as

$$P_{\text{int}}(z_0) = e^{ik_1 z_0}, \quad G(z|z_0) = \frac{i}{2k} e^{\pm ik_1 |z_0 - z|}, \quad (8)$$

where  $\pm$  signs correspond to  $z_0 > z$  and  $z_0 < z$ , respectively. By substituting Eq. (8) into the original integral Eq. (3), the backscattering and forward scattering can easily be obtained by integrating Eq. (3) directly,

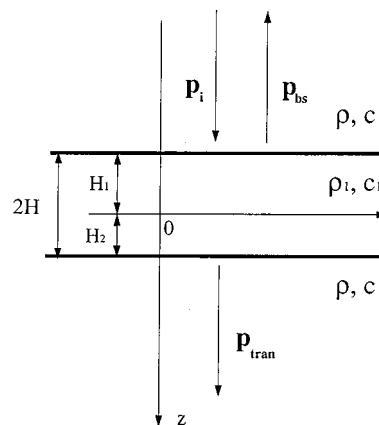


FIG. 5. Geometry of scattering by a homogeneous fluid layer. The incident wave is a plane wave. The boundaries of the layer are located at  $z = -H_1$  and  $z = H_2$ , respectively. The thickness of the layer is  $2H$ .

$$p_{bs}^{DWBA} = -\frac{h}{4}(h^2\gamma_k - \gamma_\rho)e^{-i2k_1H_1}(1 - e^{i4k_1H})e^{-ikz}$$

$$\approx \frac{1}{2}(\Delta h + \Delta g)e^{-i2k_1H_1}(1 - e^{i4k_1H})e^{-ikz}, \quad (9)$$

$$p_{fwd}^{DWBA} = -ihkH(h^2\gamma_k + \gamma_\rho)e^{ikz} \approx -i\Delta hkHe^{ikz}, \quad (10)$$

where  $h = c_1/c$  and  $g = \rho_1/\rho$  are the sound-speed and density contrasts, respectively.  $\Delta h = h - 1$  and  $\Delta g = g - 1$  are the deviations of the sound-speed and density contrasts from unity.

Now, consider the exact solution. It is easy to show that the exact solutions for backscattered (reflected) and forward scattered (transmitted–incident) fields are

$$p_{bs}^{EXACT} = Re^{-i2kH_1} \frac{(1 - e^{i4k_1H})e^{-ikz}}{1 - R^2e^{i4k_1H}}, \quad (11)$$

$$p_{fwd}^{EXACT} = Te^{i2(k_1 - k)H} \frac{(1 - R)e^{ikz}}{1 - R^2e^{i4k_1H}} e^{ikz}, \quad (12)$$

where  $R$  and  $T$  are reflection and transmission coefficients defined as

$$R = \frac{gh - 1}{gh + 1}, \quad T = \frac{2gh}{gh + 1}. \quad (13)$$

It is not yet obvious to directly see the differences between the exact and the DWBA solutions. To relate the two results, we take the advantage of the weak scattering assumption:  $g \approx 1$  and  $h \approx 1$ , which leads to  $R \ll 1$ ,  $T \approx 1$ . With some straightforward mathematical manipulations, we obtain

$$p_{bs}^{EXACT} \approx \frac{1}{2}(\Delta h + \Delta g)e^{-i2kH_1}(1 - e^{i4k_1H})e^{-ikz}$$

$$= e^{i\Phi_{bs}} p_{bs}^{DWBA}$$

$$p_{fwd}^{EXACT} \approx -i\Delta hk(1 - i\Delta hk_1H)e^{ikz}$$

$$\approx -i\Delta hke^{-i\Delta hk_1H}e^{ikz} = e^{i\Phi_{fwd}} p_{fwd}^{DWBA}, \quad (14)$$

where the two phase terms are

$$\Phi_{bs} = -2\Delta hk_1H_1, \quad \Phi_{fwd} = -\Delta hk_1H. \quad (15)$$

The amplitudes computed from both exact and the DWBA models are about the same, but their phases differ by factors of  $-2\Delta hk_1H_1$  and  $-\Delta hk_1H$  for the backscattering and forward scattering, respectively. If we choose the center of the layer as the origin, i.e.,  $H_1 = H$ , Eqs. (15) can be formulated in a more general form,

$$\Phi_{1d} = -[1 + \sin(\theta_s/2)]\Delta hk_1H, \quad (16)$$

where  $\theta_s$  is the scattering angle (angle between incident and scattering directions),  $\theta_s = 0^\circ$  for forward scattering, and  $\theta_s = 180^\circ$  for backscattering. The factor  $\Phi_{1d}$  is the compensated phase.

For a three-dimensional finite object, the phase compensation becomes more complicated. Let us consider the case of a sphere. Because of the finite dimension and the curved surface, the analytical solution, unlike the 1-D case, is a summation of an infinite series of partial waves (modes) involving Bessel and Legendre functions and is difficult to manipulate mathematically. However, we may utilize the result derived for the one-dimensional case. We take the following

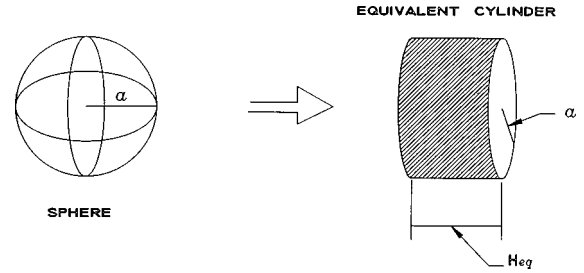


FIG. 6. Schematic of an equal volume cylinder as an approximate to a sphere, where the cylindrical radius is the same as the radius of the original sphere  $a$  and the geometric cylindrical cross section in the direction of incident is  $\pi a^2$ .

heuristic approach: first, transform the 3-D problem to a quasi-1-D problem, then seek an approximate phase compensation term similar to Eq. (16). To do this, in the incident direction we squeeze the sphere from both ends into a cylinder whose cylindrical radius is  $a$ , while keeping the volume and the maximum geometric cross section unchanged as illustrated in Fig. 6. The backscattering and the forward scattering from such an equal volume cylinder can be approximately regarded as a quasi-1-D problem when  $ka \gg 1$  (wavelength  $\lambda \ll a$ ). The equivalent thickness of the cylinder is found to be  $H_{eq} = 2a/3$ . If we choose the center of the sphere of radius  $a$  as the origin of the coordinates, following Eq. (16), we obtain the quasi-1-D phase compensation for a sphere,

$$\Phi_{sphere} = -\frac{2}{3}[1 + \sin(\theta_s/2)]\Delta hk_1a, \quad (17)$$

where scattering angle  $\theta_s$  is defined in Fig. 2. Substituting the phase compensation term Eq. (17) into the original DWBA solution Eq. (7), we obtain a phase-compensated DWBA representation of the scattering by a weakly scattering fluid sphere

$$f_{scat}^{PC-DWBA}(\theta_s, ka)$$

$$= k_1^2 a^3 (h^2 \gamma_k + \gamma_\rho \cos \theta_s) \frac{j_1(2k_1 a \sin(\theta_s/2))}{2k_1 a \sin(\theta_s/2)}$$

$$\times e^{-i2/3[1 + \sin(\theta_s/2)]\Delta hk_1a}. \quad (18)$$

Note that the above equation is applicable not only to the forward scattering and backscattering, but also to a more general bistatic scattering problem.

To validate this approach, the above solution is compared with the exact solution. Figure 7 shows the comparison between the exact solution based on the Anderson's fluid sphere model<sup>9</sup> (solid) and the phase-compensated DWBA solution (dashed) given by Eq. (18). It is found that both real and imaginary components of the scattering amplitude agree with the exact solution reasonably well. The sound-speed and density contrasts,  $h$ ,  $g$ , used in the comparison are both 1.04. For zooplankton applications,  $h = 1.04$  and  $g = 1.04$  are considered to be quite reasonable.<sup>4,10,11</sup> Because of the nature of the DWBA, it is expected that smaller  $h$  and  $g$  will give a better fit, while larger  $h$  and  $g$  will degrade the agreement.

It is obvious that the simple phase compensation Eq. (17) basically captures the phase of the scattering amplitude well. More importantly, the phase-compensated DWBA is

## SCATTERING BY A FLUID SPHERE

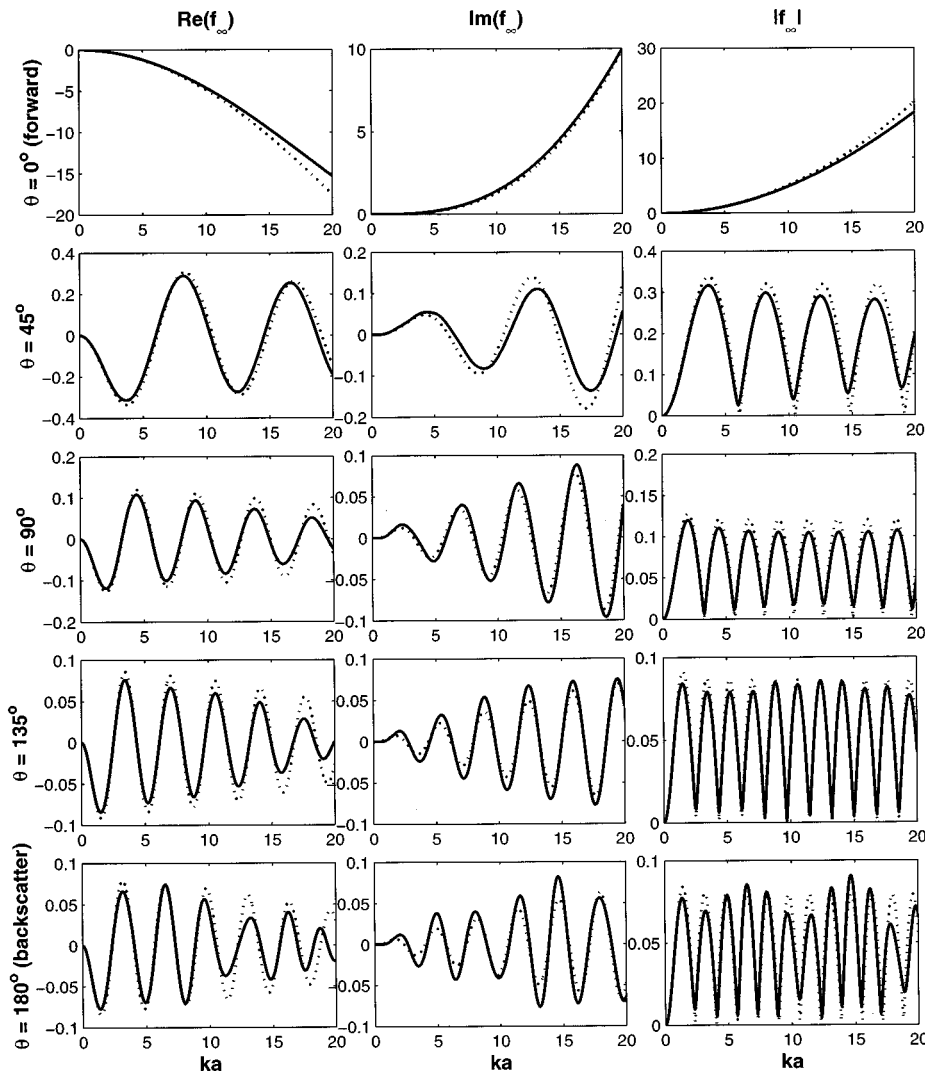


FIG. 7. Comparison of bistatic scattering form function of a fluid sphere between the exact solution and the phase-compensated DWBA solution. The solid lines are computed from the exact solution (Ref. 9) and the dashed lines are computed from Eq. (18). The simulation parameters are  $g=h=1.04$ .  $\theta_s=0^\circ$  for forward scattering and  $\theta_s=180^\circ$  for backscattering.

now capable of predicting the scattering-induced attenuation when used in describing propagation problems.

### III. APPLICATIONS TO ZOOPLANKTON

Although many marine animals such as copepods, an important zooplankton class, have been successfully modeled as spheres in predicting the biomass distributions,<sup>12,13</sup> many marine organisms have elongated shapes and have been successfully modeled as deformed cylinders for both laboratory experiments and field applications.<sup>2-5</sup> However, these models are restricted to the backscattering case and are not able to predict any scattering-induced attenuation.

In the previous section, we have presented the phase-compensated DWBA solution for a fluid sphere. To make it more general, we model the elongated marine organisms as prolate spheroids. First, we need to find the DWBA representation of the bistatic scattering by performing a 3-D integral of Eq. (7). For a broadside incidence of a plane wave on a fluid prolate spheroid shown in Fig. 8(a), the DWBA solution is [Appendix, Eq. (A20)]

$$f_{\text{scat}}^{\text{DWBA}}(\theta_s, ka, e) = k_1^2 a^3 e (h^2 \gamma_k + \gamma_p \cos \theta_s) \times \frac{j_1(k_1 a \Phi(\theta_s, e))}{k_1 a \Phi(\theta_s, e)}, \quad (19)$$

$$\Phi(\theta_s, e) = \sqrt{(1 - \cos \theta_s)^2 + e^2 \sin^2 \theta_s},$$

where  $e=b/a>1$  is the aspect ratio of the prolate spheroid, and  $a$  and  $b$  are its semiminor and semimajor axes, respectively. For  $e=1$ , Eq. (19) reduces to Eq. (7). Clearly, Eq. (19) has only the real part. To obtain the imaginary part of the scattering amplitude, we have to seek a phase compensation term as in the previous sphere case. One possible solution would be simply following the procedures that lead to the phase compensation term for the sphere case by squeezing the scatterer in the direction of the incident wave while maintaining the volume ( $V=4/3\pi a^2 b$ ) and leaving the maximum geometric cross section in that direction ( $S=\pi ab$ ) unchanged. The equivalent thickness of such a quasi-1-D problem would be  $H_{\text{eq}}=2a/3$ . However, because of the complicated nature of the scattering problem, the above solution does not provide the best result. For a prolate spheroid, the curvature on the surface is not a constant;



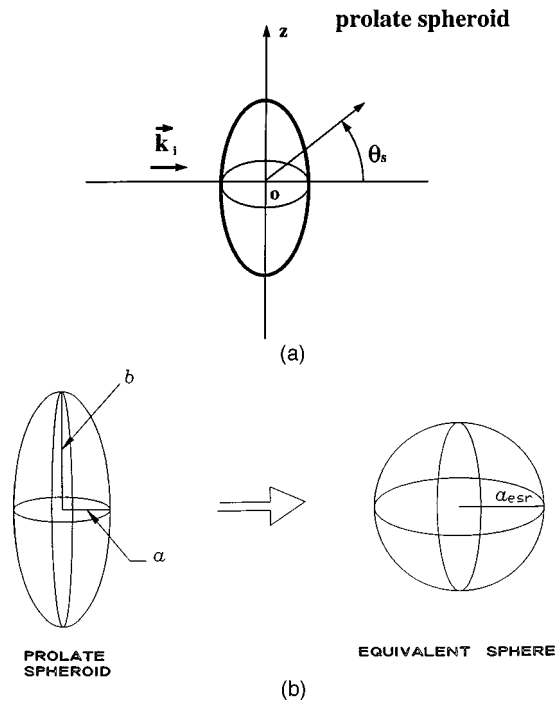


FIG. 8. Scattering by a prolate spheroid. (a) Geometry of bistatic scattering subject to a plane wave for a broadside incidence.  $\theta_s = 0^\circ$  for forward scattering and  $\theta_s = 180^\circ$  for backscattering. (b) Schematic of conversion from a prolate spheroid to an equivalent sphere of equal volume.

squeezing the prolate spheroid along the minor axis (direction of incidence) while keeping the maximum geometric cross section unchanged can be regarded as only using a constant radius of curvature,  $a$ . To overcome such a deficiency, the concept of the equivalent spherical radius commonly used in describing the scattering by irregular-shaped objects can be used, and is shown in Fig. 8(b). Since the equivalent spherical radius  $a_{\text{esr}} = ae^{1/3}$  is between  $a$  and  $b$ , using  $a_{\text{esr}}$  implies using an average radius of curvature between  $a$  and  $b$ . The resultant equivalent thickness for this quasi-1-D problem is then found to be

$$H_{\text{eq}} = \frac{2a_{\text{esr}}}{3} = \frac{2(a^2b)^{1/3}}{3} = \frac{2a}{3}e^{1/3}. \quad (20)$$

As a result, the phase compensation term for a fluid prolate spheroid can be obtained by replacing  $H$  in Eq. (16) with  $H_{\text{eq}}$  given by Eq. (20)

$$\Phi_{\text{spheroid}} = -\frac{2}{3}[1 + \sin(\theta_s/2)]\Delta hk_1 a e^{1/3} = e^{1/3}\Phi_{\text{sphere}}. \quad (21)$$

By combining Eq. (19) with Eq. (21), the complete solution for a prolate spheroid can be expressed as

$$\begin{aligned} f_{\text{scat}}^{\text{PC-DWBA}}(\theta_s, ka, e) &= k_1^2 a^3 e (h^2 \gamma_k + \gamma_\rho \cos \theta_s) \frac{j_1(k_1 a \Phi(\theta_s, e))}{k_1 a \Phi(\theta_s, e)} \\ &\quad \times e^{-i2/3[1 + \sin(\theta_s/2)]\Delta hk_1 a e^{1/3}}. \end{aligned} \quad (22)$$

Note that the above solution reduces to the sphere case, Eq. (18), if  $e = 1$ . The comparison of the phase-compensated DWBA solution with the exact modal series solution for a

broadside incidence is shown in Fig. 9, where the density and sound-speed contrasts are both chosen to be 1.04, and the aspect ratio  $e$  is 5. The exact solution is based on the result from Ref. 14. Apparently, the agreement is, in general, very promising given the fact that only a simple phase term is added. Both real and imaginary parts are described reasonably well by the phase-compensated representation, not only for monostatic scattering but also for bistatic scattering.

For objects with more complicated shapes such as zooplankton, the quasi-1-D approach may still work reasonably well provided that the scattering is weak. To show this, we express a general representation of the acoustic scattering by an arbitrarily shaped target as

$$f_{\text{scat}}^{\text{EXACT}} = |f_{\text{scat}}^{\text{EXACT}}| e^{i\phi}, \quad (23)$$

where the phase  $\phi$  is in general a function of the shape, orientation, and the material properties of the scatterer. For weakly scattering, the DWBA can predict the absolute value (module) of the complex scattering amplitude very well, i.e.,

$$|f_{\text{scat}}^{\text{EXACT}}| \approx |f_{\text{scat}}^{\text{DWBA}}|. \quad (24)$$

It is reasonable to assume that the dependence of  $\phi$  on density contrast ( $g$ ) is very weak and can be ignored; the phase term can be then be expressed as the Taylor series expansion in terms of  $\Delta h$ , sound-speed contrast deviation from unity,

$$e^{i\phi} = \exp\left(i \sum_{j=1}^{\infty} \alpha_j \Delta h^j\right) = \sum_{n=0}^{\infty} \frac{i^n}{n!} \left(\sum_{j=1}^{\infty} \alpha_j \Delta h^j\right)^n, \quad (25)$$

where  $\alpha_j$  is the coefficient to be determined by the scattering geometry. For weakly scattering,  $\alpha_j \Delta h^j \ll 1$ , Eq. (25) can be approximated by the first two terms,

$$e^{i\phi} \approx 1 + i \alpha_1 \Delta h. \quad (26)$$

This is the first-order perturbation solution of the phase-compensated DWBA and is consistent with Eqs. (18) and (22) by setting  $\alpha_1 = -2/3[1 + \sin(\theta/2)]k_1 a$  and  $\alpha_1 = -2/3[1 + \sin(\theta/2)]k_1 a e^{1/3}$ , respectively.

In the case of the forward scattering, which is crucial in studying the scattering-induced attenuation [see Eq. (2)], it can be shown that the real part of  $f_{\text{scat}}^{\text{DWBA}}$  is independent of the shape and orientation of the scatterer, and depends only on its volume. By letting the scattering angle  $\theta \rightarrow 0$ , in both Eqs. (7) and (19), we arrive at

$$f_{\text{scat}}^{\text{DWBA}}(0, ka) = \frac{k_1^2 V}{4\pi} (h^2 \gamma_k + \gamma_\rho), \quad (27)$$

for both sphere and prolate spheroid cases. To obtain the above equation, we have used the property of the spherical Bessel function

$$\lim_{x \rightarrow 0} \frac{j_1(x)}{x} = \frac{1}{3}. \quad (28)$$

It is reasonable to further assume that the imaginary part of the scattering amplitude is also independent of the shape and orientation. In other words, for the forward scattering, the solution would be the same for objects with different shapes and/or orientations as long as the total volume is kept

## SCATTERING BY A FLUID PROLATE SPHEROID

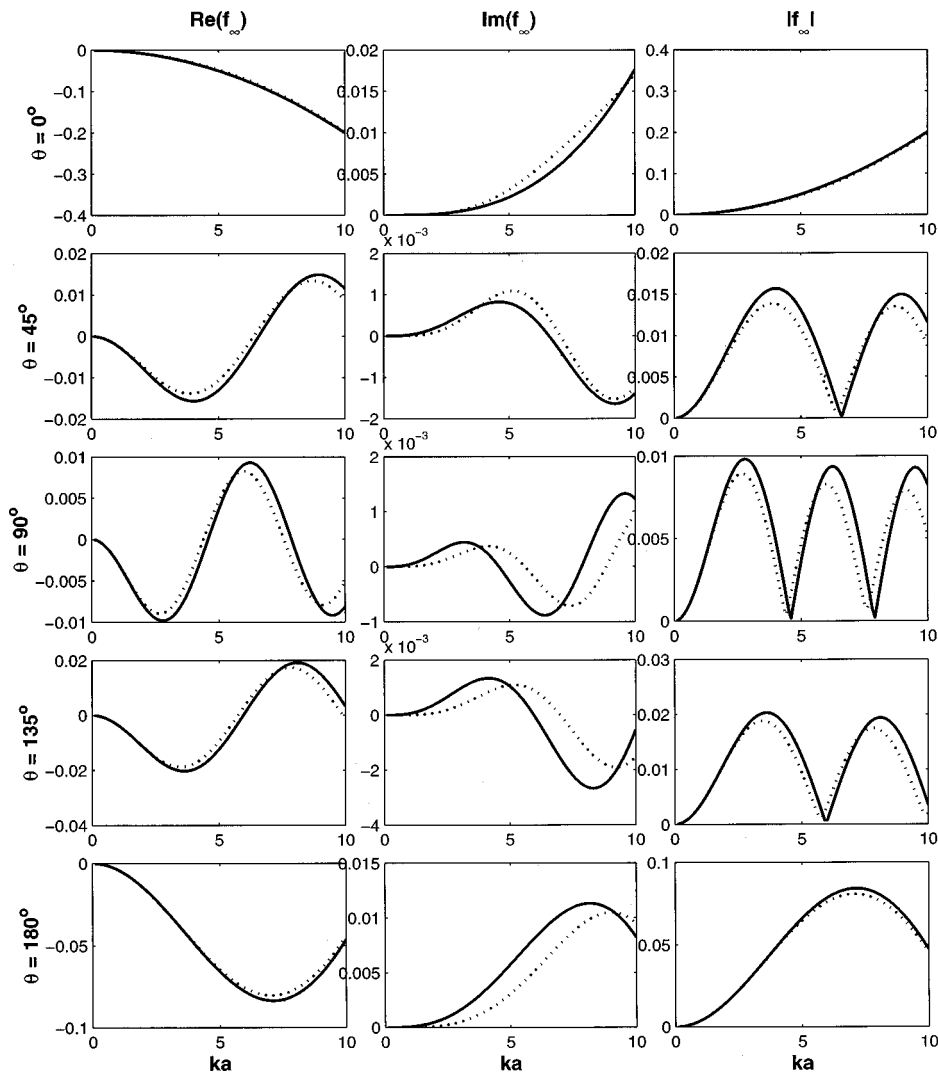


FIG. 9. Comparison of the bistatic scattering form function of a fluid spheroid between the exact solution and the phase-compensated DWBA solution. The solid lines are computed from the exact solution (Ref. 14) and the dashed lines are computed from Eq. (22).  $a$  and  $b$  are semiminor and semimajor axes of the prolate spheroid, respectively. The simulation parameters are  $g=h=1.04$ , and the aspect ratio is  $e=5$ .

the same. The quasi-1-D approach is basically using the concept of “equal volume.”

In contrast, for a general bistatic scattering,  $\alpha_j$  is a function of the shape and orientation of the scatterer, resulting in possible larger errors (compared to the forward scattering), but might not be unacceptable giving the condition of weakly scattering (see Fig. 9). The quantitative evaluation of such errors would involve tremendous numerical efforts and will not be discussed here.

Before applying the theory to evaluate the scattering-induced attenuation, according to the forward-scattering theorem given by Eq. (2), we need to first understand how the complex forward-scattering amplitude depends on various parameters. For simplicity, the sphere model will be used in the following analysis. Figure 10 shows how the scattering amplitude changes as sound-speed and density contrast change. The exact solution is based on the Anderson’s fluid sphere model.<sup>9</sup> It is found that for a fixed  $\Delta g$ , the real part of the scattering amplitude of a fluid sphere depends on  $\Delta h$  almost linearly ( $\langle \text{slope} \rangle \approx 0.9$ ), while the imaginary part approximately depends on  $\Delta h$  quadratically ( $\langle \text{slope} \rangle \approx 1.64$ ). The slopes marked in the figure are the average slopes over a range of  $\Delta h$  from 0.01 to 0.08. To better understand such

functional dependencies, we need to analysis the phase-compensated DWBA model given by Eq. (18). For forward scattering,  $\theta_s=0$ , Eq. (18) reduces to

$$\begin{aligned}
 f_{\text{scat}}^{\text{PC-DWBA}}(0,ka) &= \frac{k_1^2 a^3}{3} (h^2 \gamma_k + \gamma_p) e^{-i2/3 \Delta h k_1 a} \\
 &\approx -\frac{k_1^2 V}{4\pi} (2\Delta h + \Delta h^2) \left( 1 - i \frac{2}{3} \Delta h k_1 a \right),
 \end{aligned}
 \tag{29}$$

where  $V=(4\pi/3)a^3$  is the volume of the sphere. In expanding the phase term,  $e^{-i2/3 \Delta h k_1 a}$ , the terms of  $\Delta h$  with orders higher than 1 have been ignored. As mentioned previously, the forward-scattering amplitude is independent of the shape orientation, and is proportional to the volume of the scatterer. This may be a piece of useful information that can be used in detecting marine animals by forward scatter.

To determine analytically how the real and imaginary parts of the scattering amplitude depend on  $\Delta h$ , we assume the relations can be approximately described by a power law,  $(\Delta h)^p$ . The power  $p$  can be determined by finding the slope on a logarithm plot of the scattering amplitude versus  $\Delta h$ ,

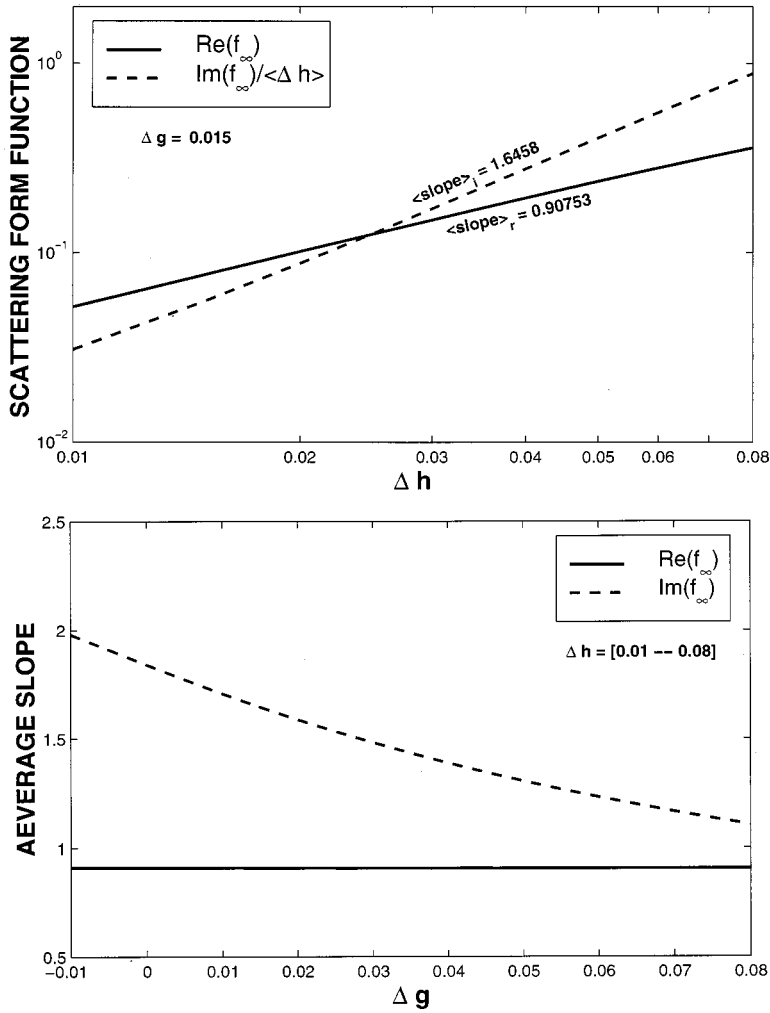


FIG. 10. Scattering form function of a fluid sphere in the forward direction  $\theta=0^\circ$  as a function of sound-speed contrast deviation ( $\Delta h=h-1$ ) and density contrast deviation ( $\Delta g=g-1$ ). The computations are based on the exact modal series solution for a fluid sphere. The solid and dashed lines correspond to the real and the imaginary parts of the scattering form function, respectively. (a) Real and imaginary parts of the forward scattering form functions as function of  $\Delta h$  with fixed  $\Delta g=0.015$  and  $ka=2.0$ . Since  $\text{Im}(f_\infty) \ll \text{Re}(f_\infty)$ , the imaginary part has been normalized by  $\langle \Delta h \rangle$ ; (b) average slope for real and imaginary parts of forward scattering form functions versus  $\Delta g$ . The average over  $\Delta h$  is from 0.01 to 0.08.

$$p = \frac{d \ln f_{\text{scat}}}{d \ln \Delta h}, \quad (30)$$

where  $\ln$  is the natural logarithm and  $f_{\text{scat}}$  could be either the real or imaginary part of the scattering amplitude. In order to apply a logarithmic operation, both real and imaginary components need to be replaced by their absolute values. From Eq. (29), by applying the chain rule of derivative, we obtain for the real part

$$p_r = \frac{d \ln |\text{Re}(f_{\text{scat}}^{\text{PC-DWBA}})|}{d \ln \Delta h} = \frac{d \ln |\text{Re}(f_{\text{scat}}^{\text{PC-DWBA}})|}{d \Delta h} \frac{d \Delta h}{d \ln \Delta h}, \quad (31)$$

$$= \left( \frac{1}{\Delta h} + \frac{1}{2 + \Delta h} - \frac{2}{1 + \Delta h} \right) \Delta h = 1 - \frac{3}{2} \Delta h. \quad (32)$$

Since  $\Delta h$  varies from 0.01 to 0.08 in the simulations,  $\langle \Delta h \rangle = 0.045$ , the average slope or power,  $p_r$ , is then

$$\langle p_r \rangle = 1 - \frac{3}{2} \langle \Delta h \rangle \approx 0.93. \quad (33)$$

Similarly, for the imaginary part of the scattering amplitude

$$p_i = \frac{d \ln |\text{Im}(f_{\text{scat}}^{\text{PC-DWBA}})|}{d \ln \Delta h} = \frac{d \ln |\text{Im}(f_{\text{scat}}^{\text{PC-DWBA}})|}{d \Delta h} \frac{d \Delta h}{d \ln \Delta h}, \quad (34)$$

$$= \left( \frac{2}{\Delta h} + \frac{1}{2 + \Delta h} - \frac{3}{1 + \Delta h} \right) \Delta h \quad (35)$$

$$= 2 - \frac{5}{2} \Delta h, \quad (36)$$

hence,

$$\langle p_i \rangle = 2 - \frac{5}{2} \langle \Delta h \rangle \approx 1.89. \quad (37)$$

Comparing  $\langle p_r \rangle$  and  $\langle p_i \rangle$  in Eqs. (33) and (37) with those computed from the exact solutions given in Fig. 10(a), we find that the analysis based on the phase-compensated DWBA model overestimates the power  $p_r$  and  $p_i$ , but with an error less than 4% for the real part and 16% for the imaginary part. The mismatch implies a more complicated nature of the scattering process, which depends not only on sound-speed contrast but also on density contrast. Note that, from Eqs. (33) and (37), as  $\Delta h$  approaches zero,  $\langle p_r \rangle \rightarrow 1$  (linear) and  $\langle p_i \rangle \rightarrow 2$  (quadratic).

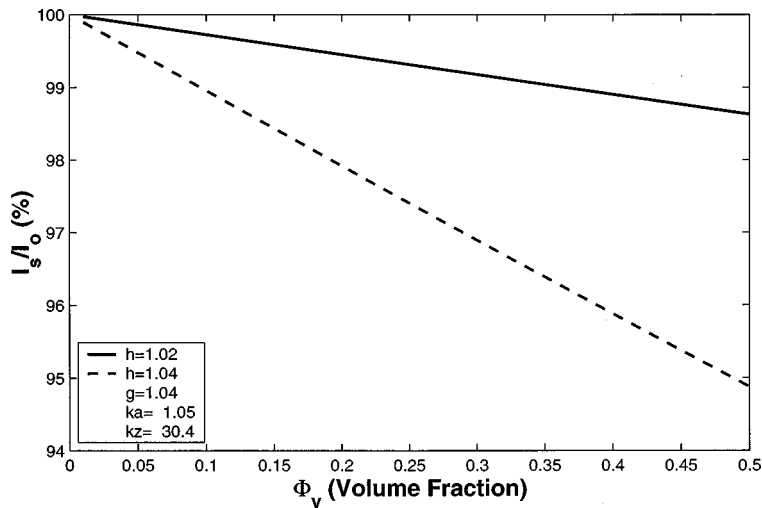


FIG. 11. Percentage of the scattering intensity ( $I_s/I_0$ ) observed at distance  $z$  shown in Fig. 1 as a function of volume fraction of scatterers,  $\Phi_v$ . The scatterers are assumed to be randomly distributed identical prolate spheroids. Two different values of  $h$  have been chosen to show the dependence of attenuation on  $h$ . The solid line corresponds to  $h=1.02$  and the dashed line corresponds to  $h=1.04$ , while  $\Delta g$  is held constant at 0.04. The other simulation parameters are: semiminor axis  $a_0=0.5$  mm, aspect ratio  $e_0=5$ , acoustic frequency is 500 kHz, sound speed in water  $c_w=1500$  m/s, the distance  $z=1.45$  cm. Corresponding  $ka$  and  $kz$  are 1.05 and 30.4, respectively. The PDF of the scatterers is chosen to be  $P(a,e)=\delta(a-a_0)\delta(e-e_0)$ .

Figure 10(b) illustrates how average slopes,  $\langle p_r \rangle$  and  $\langle p_i \rangle$ , change as density contrast deviation,  $\Delta g$ , changes. As in Fig. 10(a), the computations are based on the Anderson's fluid sphere model. Obviously, the slope for the real part is essentially a constant while the slope for the imaginary part decreases from 2 to 1.1 as  $\Delta g$  increases from  $-0.01$  (a negative  $\Delta g$  implies a positively buoyant scatterer) to 0.08. This result is expected since DWBA itself is valid only for weakly scattering objects, i.e., density and sound-speed contrasts are close to unity ( $\Delta g$  and  $\Delta h$  are close to zeros).

Including the imaginary part in the scattering amplitude, Eqs. (18) and (22) are able to predict the scattering-induced attenuation. Consider a unit plane wave propagation through a medium containing densely aggregated animals such as the case shown in Fig. 1. To estimate the acoustic intensity at a distance  $z$ , we can apply Eq. (1) and the forward-scattering theory Eq. (2) to compute the scattering-induced attenuation. Assuming the joint size and aspect ratio probability density function (PDF) of the animals is  $P(a,e)$ , where  $a$  is the equivalent spherical radius. The average extinction cross section per unit volume due to the scattering by animals can be expressed as

$$\langle \sigma_e \rangle = \frac{4\pi n_0}{k} \int \text{Im}(f_{\text{scat}}^{\text{PC-DWBA}}(0,ka,e))P(a,e)da de, \quad (38)$$

where  $n_0$  is the number of scatterers in a unit volume and

$$\int P(a,e)da de = 1. \quad (39)$$

To evaluate Eq. (38) analytically, we choose the simplest case by setting  $P(a,e)=\delta(a-a_0)\delta(e-e_0)$ , where  $\delta(x)$  is the Dirac delta function. The intensity at  $z$  is then

$$I_s(z) = I_0 e^{-\langle \sigma_e \rangle z} = I_0 e^{-4\pi\phi_v/k \text{Im}(f_{\text{scat}}^{\text{PC-DWBA}}(0,ka_0,e_0)/V)}, \quad (40)$$

where  $\phi_v = n_0 V$  is the volume fraction of the animal. The imaginary part of the phase-compensated DWBA solution for a prolate spheroid,  $\text{Im}[f_{\text{scat}}^{\text{PC-DWBA}}(0,ka_0,e_0)]$  is given by Eq. (22). Figure 11 demonstrates the dependence of the observed intensity at location  $z$  on the volume concentration  $\phi_v$  for two different sound-speed contrasts. The intensity at  $z$

decreases about 1.4% for  $h=1.02$  ( $\Delta h=0.02$ ), and decreases about 5.1% for  $h=1.04$  ( $\Delta h=0.04$ ). The decrease in intensity is nearly quadrupled for  $\Delta h=0.04$  as compared to  $\Delta h=0.02$ , an outcome stemming from the approximate quadratic relation between the extinction cross section and  $\Delta h$  as shown in Fig. 10. The implication of Fig. 11 is that once we know the volume fraction  $\Phi_v$ , by measuring the intensity attenuation, it is possible to infer the material properties of zooplankton whose influence on the acoustical scattering is significant and has been a very important issue in studying zooplankton acoustics.<sup>15</sup>

#### IV. CONCLUSIONS

A phase-compensated DWBA-based model to describe the bistatic scattering by weakly scattering objects has been developed. It has been shown that the approximate phase compensation term is simple, but reasonably characterizes the general feature of the scattering for both sphere and prolate spheroid. One of the major achievements of including a phase compensation term is that the solution is able to predict the scattering-induced attenuation, which could be important when dealing with densely aggregated zooplankton.

From our studies, it is found that the imaginary part of the scattering amplitude is much smaller than the real part, approximately by a factor of  $\Delta hka$ . The dependence of the scattering amplitude on the material properties of the scatterers has been investigated. The real part of the scattering amplitude is found to approximately depend on  $\Delta h$  linearly, while the imaginary part is found to approximately depend on  $\Delta h$  quadratically.

In reality, when the scattering-induced attenuation is noticeable, it is more convenient to measure the acoustic intensity which is directly related to the imaginary part of the forward scattering amplitude and infer the acoustic properties of the scatterers statistically.

#### ACKNOWLEDGMENTS

The authors would like to thank Emile Hoskinson for providing the numerical results for scattering by a prolate spheroid using the exact solutions. This work was partially



supported by the National Science Foundation under Grant No. OCE-9730680. This is Woods Hole Oceanographic Institution Contribution No. 9819.

## APPENDIX

By choosing the  $z$ -axis coinciding with the major axis of the prolate spheroid and defining

$$C_b = h^2 \gamma_k + \hat{k}_i \cdot \hat{k}_s \gamma_\rho, \quad (\text{A1})$$

which is a constant within the prolate spheroid, the first line of Eq. (7) can then be rewritten as

$$\begin{aligned} f_{\text{scat}}^{\text{DWBA}} &= \frac{k_1^2}{4\pi} \int_V C_b e^{i(\mathbf{k}_i - \mathbf{k}_s) \cdot \mathbf{r}_0} dV_0 \\ &= \frac{k_1^2 C_b}{4\pi} \int_0^{2\pi} d\phi_0 \int_0^{r(z_0)} \rho_0 d\rho_0 \int_{-b}^b e^{i(\mathbf{k}_i - \mathbf{k}_s) \cdot \mathbf{r}_0} dz_0, \end{aligned} \quad (\text{A2})$$

where  $b$  is the semimajor axis, and

$$\mathbf{r}_0 = x_0 \hat{x} + y_0 \hat{y} + z_0 \hat{z}, \quad \rho_0 = \sqrt{x_0^2 + y_0^2}, \quad (\text{A3})$$

$$\mathbf{k}_i - \mathbf{k}_s = k_1 (\hat{k}_i - \hat{k}_s) = k_1 (\mu_x, \mu_y, \mu_z),$$

where

$$\begin{aligned} \mu_x &= \cos \phi_i \sin \theta_i - \cos \phi_s \sin \theta_s \\ \mu_y &= \sin \phi_i \sin \theta_i - \sin \phi_s \sin \theta_s \\ \mu_z &= \cos \theta_i - \cos \theta_s, \end{aligned} \quad (\text{A4})$$

with the incident and scattered unit vectors  $\hat{k}_i$  and  $\hat{k}_s$  defined as

$$\begin{aligned} \hat{k}_i &= \cos \phi_i \sin \theta_i \hat{x} + \sin \phi_i \sin \theta_i \hat{y} + \cos \theta_i \hat{z} \\ \hat{k}_s &= \cos \phi_s \sin \theta_s \hat{x} + \sin \phi_s \sin \theta_s \hat{y} + \cos \theta_s \hat{z}. \end{aligned} \quad (\text{A5})$$

Using Eqs. (A3)–(A5), we have

$$\begin{aligned} (\mathbf{k}_i - \mathbf{k}_s) \cdot \mathbf{r}_0 &= k_1 (\mu_x \rho_0 \cos \phi_0 + \mu_y \rho_0 \sin \phi_0 + \mu_z z_0) \\ &= k_1 (\mu_r \rho_0 \cos(\phi_0 - \phi') + \mu_z z_0), \end{aligned} \quad (\text{A6})$$

where

$$\begin{aligned} \phi_0 &= \tan^{-1} \frac{y_0}{x_0}, \\ \mu_r &= \sqrt{\mu_x^2 + \mu_y^2}, \\ \phi' &= \tan^{-1} \frac{\mu_y}{\mu_x}. \end{aligned} \quad (\text{A7})$$

Substituting Eq. (A6) into Eq. (A2), we have

$$\begin{aligned} f_{\text{scat}}^{\text{DWBA}} &= \frac{k_1^2 C_b}{4\pi} \int_0^{2\pi} e^{ik_1 \mu_r \rho_0 \cos(\phi_0 - \phi')} d\phi_0 \\ &\quad \times \int_0^{r(z_0)} \rho_0 d\rho_0 \int_{-b}^b e^{ik_1 \mu_z z_0} dz_0, \end{aligned} \quad (\text{A8})$$

$$= \frac{k_1^2 C_b}{2} \int_0^{r(z_0)} \rho_0 J_0(k_1 \mu_r \rho_0) d\rho_0 \int_{-b}^b e^{ik_1 \mu_z z_0} dz_0, \quad (\text{A9})$$

$$= \frac{k_1 C_b}{2\mu_r} \int_{-b}^b r(z_0) J_1(k_1 \mu_r r(z_0)) e^{ik_1 \mu_z z_0} dz_0, \quad (\text{A10})$$

where  $J_0(x)$  and  $J_1(x)$  are Bessel functions of order 0 and 1, respectively. To obtain the last equation, we have used integral properties of the Bessel function.<sup>16</sup> For a spheroid, the  $z$ -dependent radius  $r(z_0)$  can be expressed as

$$r(z_0) = a \sqrt{1 - \left(\frac{z_0}{b}\right)^2}, \quad (\text{A11})$$

where  $a$  and  $b$  are the semi-minor and-major axes of the prolate spheroid, respectively. Let

$$u = \frac{z_0}{b}, \quad du = \frac{dz_0}{b}, \quad (\text{A12})$$

Eq. (A10) becomes

$$\begin{aligned} f_{\text{scat}}^{\text{DWBA}} &= \frac{k_1 a b C_b}{2\mu_r} \int_{-1}^1 \sqrt{1-u^2} J_1(\mu_r k_1 a \sqrt{1-u^2}) e^{i\mu_z k_1 b u} du \\ &= \frac{k_1 a b C_b}{\mu_r} \int_0^1 \sqrt{1-u^2} J_1(\mu_r k_1 a \sqrt{1-u^2}) \\ &\quad \times \cos(\mu_z k_1 b u) du. \end{aligned} \quad (\text{A13})$$

The result of the integral can be found in Ref. 17,

$$\begin{aligned} I &= \int_0^1 \sqrt{1-u^2} J_1(k_1 a \mu_r \sqrt{1-u^2}) \cos(ik_1 \mu_z b u) du \\ &= \mu_r k_1 a \frac{j_1(k_1 a \sqrt{\mu_r^2 + e^2 \mu_z^2})}{k_1 a \sqrt{\mu_r^2 + e^2 \mu_z^2}}, \end{aligned} \quad (\text{A14})$$

where  $e = b/a$  is the aspect ratio of the spheroid and  $j_1(x)$  is the spherical Bessel function of order 1. To obtain Eq. (A14), a conversion from the cylindrical Bessel function (taken directly from Ref. 17) to the spherical function is needed by using the following relation:

$$j_1(x) = \sqrt{\frac{\pi}{2x}} J_{3/2}(x). \quad (\text{A15})$$

Substituting Eqs. (A14) and (A1) into Eq. (A13), the final solution can be written as

$$f_{\text{scat}}^{\text{DWBA}} = k_1^2 a^3 e (h^2 \gamma_k + \cos \Theta \gamma_\rho) \frac{j_1(k_1 a \sqrt{\mu_r^2 + e^2 \mu_z^2})}{k_1 a \sqrt{\mu_r^2 + e^2 \mu_z^2}}, \quad (\text{A16})$$

where  $\mu_r$  and  $\mu_z$  are defined in Eqs. (A7) and (A1), and  $\Theta$  is the angle between incident  $\hat{k}_i$  and scattered  $\hat{k}_s$  waves

$$\cos \Theta = \cos \theta_i \cos \theta_s + \sin \theta_i \sin \theta_s \cos(\phi_i - \phi_s). \quad (\text{A17})$$

Although Eq. (A16) is originally derived by assuming a prolate spheroid, it is also applicable to oblate spheroids by simply allowing  $e < 1$ .

Consider a special case of broadside incidence, in which  $\theta_i = \pi/2$  and  $\phi_i = 0$ . If we further assume the reception is in the same plane as the incidence, we obtain

$$\mu_r = 1 \pm \sin \theta_s, \quad \mu_z = -\sin \theta_s \cos \phi_s, \quad (\text{A18})$$

where  $\pm$  signs correspond to the scattering in forward directions ( $\phi = 0$ ) and backward directions ( $\phi = \pi$ ), respectively. If we define a new angle  $\theta'_s = \pi/2 \mp \theta_s$ , where  $\mp$  signs correspond to the scattering in forward and backward scattering, respectively, Eq. (A18) becomes

$$\mu_r = 1 - \cos \theta'_s, \quad \mu_z = -\cos \theta'_s. \quad (\text{A19})$$

where  $\theta'_s$  is 0 for forward scattering and  $\pi$  for backscattering, respectively. In addition, it can be deduced that  $\Theta = \theta'_s$ , implying that this new angle  $\theta'_s$  is just the angle between the incident and scattering directions as shown in Fig. 8(a) [ $\theta_s$  in the figure is equivalent to  $\theta'_s$  in Eq. (A19)] ranging from 0 to  $\pi$ . The DWBA representation for this special case is then

$$f_{\text{scat}}^{\text{DWBA}} = k_1^2 a^3 e (h^2 \gamma_k + \gamma_\rho \cos \theta'_s) \times \frac{j_1(k_1 a \sqrt{(1 - \cos \theta'_s)^2 + e^2 \sin^2 \theta'_s})}{k_1 a \sqrt{(1 - \cos \theta'_s)^2 + e^2 \sin^2 \theta'_s}}. \quad (\text{A20})$$

Specifically, for  $e = 1$ , Eq. (A20) reduces to the DWBA representation of the scattering amplitude by a fluid sphere,

$$f_{\text{scat}}^{\text{DWBA}} = k_1^2 a^3 (h^2 \gamma_k + \gamma_\rho \cos \theta'_s) \frac{j_1(2k_1 a \sin(\theta'_s/2))}{2k_1 a \sin(\theta'_s/2)}. \quad (\text{A21})$$

For the simple Born approximation (BA), Eq. (A21) reduces to

$$f_{\text{scat}}^{\text{BA}} = k^2 a^3 (\gamma_k + \gamma_\rho \cos \theta'_s) \frac{j_1(2ka \sin(\theta'_s/2))}{2ka \sin(\theta'_s/2)}. \quad (\text{A22})$$

<sup>1</sup>C. F. Greenlaw, "Backscattering spectra of preserved zooplankton," J. Acoust. Soc. Am. **62**, 44–52 (1997).

<sup>2</sup>D. Chu, K. G. Foote, and T. K. Stanton, "Further analysis of target strength measurements of Antarctic krill at 38 kHz and 120 kHz: Comparison with deformed cylinder model and inference of orientation distribution," J. Acoust. Soc. Am. **93**, 2985–2988 (1993).

<sup>3</sup>T. K. Stanton, D. Chu, P. H. Wiebe, and C. S. Clay, "Average echoes from randomly oriented random-length finite cylinders: Zooplankton models," J. Acoust. Soc. Am. **94**, 3463–3472 (1993).

<sup>4</sup>T. K. Stanton, D. Chu, and P. H. Wiebe, "Sound scattering by several zooplankton groups. II. Scattering Models," J. Acoust. Soc. Am. **103**, 236–253 (1998).

<sup>5</sup>P. H. Wiebe, T. K. Stanton, M. C. Benfield, D. G. Mountain, and C. H. Greene, "High-frequency acoustic volume backscattering in the Georges Bank coastal region and its interpretation using scattering models," in "Shallow water acoustics, geophysics, and oceanography," IEEE J. Ocean Eng. **22**(3), 445–464 (1997).

<sup>6</sup>P. M. Morse and U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 8, pp. 407–414.

<sup>7</sup>A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. I, Chap. 1, pp. 14–15, Vol. II, Chap. 14, pp. 265–268.

<sup>8</sup>M. Born and W. Wolf, *Principles of Optics* (Macmillan, New York, 1964).

<sup>9</sup>V. C. Anderson, "Scattering by a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431 (1950).

<sup>10</sup>K. Foote, "Speed of sound in *Euphausia superba*," J. Acoust. Soc. Am. **87**, 1405–1408 (1990).

<sup>11</sup>K. G. Foote, T. Knutsen, A. E. Bekkevold, P. Dalpadado, and S. E. Johannessen, "Initial, collateral measurements of some properties of *calanus finmarchicus*," ICES C.M. 1996/L.21 Ref. B, p. 23.

<sup>12</sup>D. V. Holliday, R. E. Pieper, and G. S. Kleppel, "Determination of zooplankton size and distribution with multi-frequency acoustic technology," J. Cons., Cons. Int. Explor. Mer **46**, 52–61 (1989).

<sup>13</sup>D. V. Holliday and R. E. Pieper, "Bioacoustical oceanography at high frequencies," ICES J. Mar. Sci. **52**, 279–296 (1995).

<sup>14</sup>Z. Ye, E. Hoskinson, R. K. Dewey, L. Ding, and D. M. Farmer, "A method for acoustic scattering by slender bodies. I. Theory and verification," J. Acoust. Soc. Am. **102**, 1964–1976 (1997).

<sup>15</sup>Z. Ye and S. McClatchie, "On inferring speed of sound in aquatic organisms," J. Acoust. Soc. Am. **103**, 1667–1670 (1998).

<sup>16</sup>I. S. Gradshteyb and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Corrected and Enlarged Edition (Academic, New York, 1980).

<sup>17</sup>A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transformations* (McGraw-Hill, New York, 1954), Vol. I, p. 57.

# Low grazing angle bistatic sea floor scattering on the Florida Atlantic coastal shelf

Christopher M. Day and Tokuo Yamamoto

University of Miami Geoacoustical Laboratory, Division of Applied Marine Physics, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 6 January 1998; revised 27 April 1999; accepted 1 June 1999)

Observations of bistatic scattering at moderate frequencies 3750 Hz, 7500 Hz, and 15 000 Hz and low grazing angles ( $2^\circ$ – $20^\circ$ ) are presented with a geoacoustical description of the site. Given a source–receiver separation of 80.7 m, bistatic scattering strengths in the sandy sediments at the measurement site range from  $-70$  to  $-15$  dB. The observed bistatic scattering at low grazing angles depends primarily on the incident bearing angle. Bistatic scattering values are highest along the transect between the source and receiver. Maximum scattering strengths ( $-20$  to  $-15$  dB) appear 30 m from the receiver. Regions of high scattering ( $>-30$  dB) strength are centered along the direct path between the source and receiver. The contribution of volume scattering to the observed scattering pattern is not very clear. However, given the sandy sediments found at the site, roughness scattering is expected to be a major contributor to the observed bistatic scattering patterns. © 1999 Acoustical Society of America. [S0001-4966(99)00910-8]

PACS numbers: 43.30.Ft, 43.30.Ky, 43.30.Gr, 43.30.Ma [DLB]

## INTRODUCTION

The interaction of sound with the sea floor is essential for the understanding of sound propagation in shallow water. When the sea floor is ensonified from the water column, several mechanisms may result in a scattering of acoustical energy: bottom roughness, sub-sea floor interfaces (i.e. shallow bedrock), gas bubbles (Boyle and Chotiros, 1995), and inhomogeneities within the sediment volume (Lyons *et al.*, 1994). Of these mechanisms, the most common are roughness and volume scattering. Jackson and Briggs (1992) state that bottom roughness scattering is expected to be the primary scattering mechanism in coarse sands and gravels at frequencies in the 15 to 45-kHz range. Lyons, *et al.* (1994) observe that in high backscatter areas ( $-35$  to  $-25$  dB, grazing angle= $40^\circ$ ,  $f=6500$  Hz) volume scattering is the primary mechanism, while in low backscatter areas ( $<-40$  dB, grazing angle= $40^\circ$ ,  $f=6500$  Hz), roughness scattering dominates. Scattering due to gas bubbles can occur in fine grained, organic sediments (Boyle and Chotiros, 1995).

A large number of theoretical models have been developed to describe these scattering mechanisms. Moe *et al.* (1995), Swift and Stephen (1994), etc., apply propagation models, using their description of the scattered wave field to estimate scattering strengths. Others, such as Hines (1996), Williams and Jackson (1996), Yamamoto (1996), Boyle and Chotiros (1995), Essen (1994), Jackson *et al.* (1986), etc., have developed analytical expressions for the scattering strength as a function of the scattering geometry and the roughness, sound speed, and density statistics of the sea floor.

Numerous backscatter data have been collected and used to verify analytical and numerical scattering models (Rogers, 1996; Lyons *et al.*, 1994; Mourad and Jackson, 1993; Gensane, 1993; Jackson and Briggs, 1992; Merklinger, 1968). However, forward, “side,” or bistatic scattering measurements have been limited. Stanic *et al.* (1991) note that the

bistatic scattering experiments have often been costly and complex. Their 20 to 180 kHz bistatic scattering measurements near Panama City, Florida held the source and receiver in close (0.50 m) proximity. Scattering strengths are treated as backscatter and presented as functions of frequency, nominal grazing angle, the differences between incident and scattered grazing angle, and the difference between incident and scattered bearing angle. Urick (1960) presents early 22-kHz measurements of forward or “side” scattering, also near Panama City, Florida. His experiment was conducted using a receiver mounted 2700 m from the source. Moe *et al.* (1995) have used the data of Chotiros (1995) to verify a bistatic scattering propagation model. Despite these examples, the amount of bistatic scattering measurements is small in comparison to backscatter measurements; additional data is desirable.

The objective of this work is to present measurements of low grazing angle bistatic scattering on the Florida Atlantic coastal shelf using a three dimensional, bilinear acoustic array. Backscatter measurements and crosswell tomography are used to provide a geoacoustic description of the measurement site for use in future modeling efforts.

## MEASUREMENTS

### Measurement sites

Measurements of acoustic sea floor scattering were taken aboard the RV SEWARD JOHNSON between Nov. 17 and Nov. 24, 1996. The six measurement sites, listed in Table I and shown in Fig. 1, were located approximately 5 to 9 nautical miles east of Fort Pierce, Florida. Backscattering was sampled at sites S1, S2, S3, and S4, while bistatic scattering was sampled at sites FS1 and FS2. Except for site FS2, all of the measurement sites fell on or near an east–west transect approximately 4 nautical miles long.

Gravity cores were taken at each of the measurement

TABLE I. Location and depth of measurement sites.

Site	Latitude	Longitude	Depth (m)
S1	27° 27.50' N	80° 12.1' W	10.7
S2	27° 27.40' N	80° 12.3' W	11.3
S3	27° 27.60' N	80° 13.6' W	7.9
S4	27° 27.85' N	80° 8.8' W	16.8
FS1	27° 27.80' N	80° 8.7' W	16.2
FS2	27° 27.40' N	80° 8.8' W	18.3

sites. At all sites, gravity cores revealed the sediments to be composed of coarse sands mixed with shell fragments.

Water depths at the measurement sites ranged from 8 to 18 m. In general, water depths increased towards the east. However, two of the sites, S1 and S2, were located near a dredge disposal site centered at 27° 27.5' N, 80° 11.5' W. The shallower depth at site S1 reflects the presence of this disposal site. The presence of a dredge disposal site also suggests that sediment properties near sites S1 or S2 could differ from those of the surrounding region.

Wave action was characterized by 100 to 500 foot long, 3 foot swells from the northeast and small 1 to 3 foot chop from the west during the beginning of the experiments and 6 foot breaking seas from the north near the end of the experiments. Several strong storms had taken place in the vicinity during October and November 1996. This sea state resulted in well mixed temperature and salinity profiles, as well as turbid water conditions. The turbidity of the water made visual assessment of the bottom roughness difficult. At several of the sites (Fig. 1), no light reached the bottom, and divers were not able to examine the sea floor. However, at site S2, divers were able to identify a large sand ripple running in the north/south direction.

**Apparatus**

Measurements were taken using an ITC Model 6121 source and a bilinear receiver array. This source is cylindrical, 10 cm in diameter and 40 cm long, thus exhibits directivity. Source levels were approximately 182 dB *re* 1  $\mu$ Pa in the horizontal direction (perpendicular to the axis) and were reduced according to the directivity for grazing angles not equal to zero. The directivity of the ITC Model 6121 source is given as a function of the angle and frequency in Table II.

The bilinear receiver array, shown in Fig. 2, consisted of two horizontal, orthogonal linear arrays mounted 3.0 m and

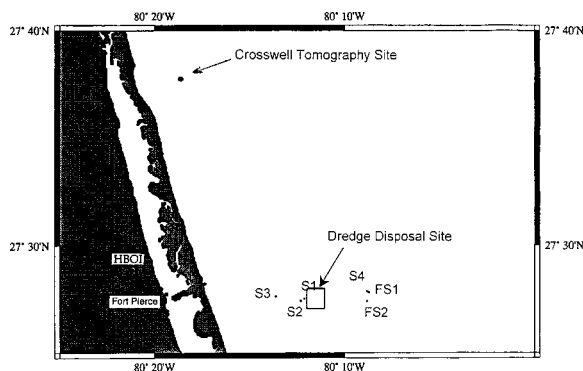


FIG. 1. Location of measurement sites.

TABLE II. Directivity of ITC 6121 cylindrical source.

Grazing angle (degrees)	Source directivity index (dB)		
	3750 Hz	7500 Hz	15 000 Hz
0	0.0	0.0	0.0
5	0.0	-0.3	-1.0
10	-0.3	-0.8	-3.5
15	-0.7	-2.2	-8.0
20	-1.1	-4.3	-15.0
25	-1.6	-6.2	-16.0
30	-2.4	-8.2	-14.0
35	-3.3	-10.4	-14.0
40	-4.7	-14.2	-16.0
45	-5.5	-16.5	-21.0
50	-6.1	-17.7	-24.0
55	-6.6	-18.1	-23.0
60	-6.9	-17.3	-19.5
65	-7.7	-16.7	-18.0
70	-7.9	-15.7	-17.5
75	-8.2	-14.8	-17.8
80	-8.3	-14.4	-18.0
85	-8.5	-14.4	-18.3
90	-8.5	-14.5	-18.5

4.5 m above the sea floor. Each linear array was encased in a sealed, oil filled PVC pipe housing 16 omnidirectional Benthos AQ-4 hydrophones. Spacing between the array elements was 10 cm. Each hydrophone, with a sensitivity of -201 dB *re* 1  $\mu$ Pa per volt, was hardwired to a preamplifier with a gain of 50.

For backscattering experiments, the source was mounted in the center of the bilinear receiver array, as shown in Fig. 2. For bistatic scattering experiments, the source was removed from the array and mounted 3.75 m above the sea floor on a similar tower. For either case, the source was fixed

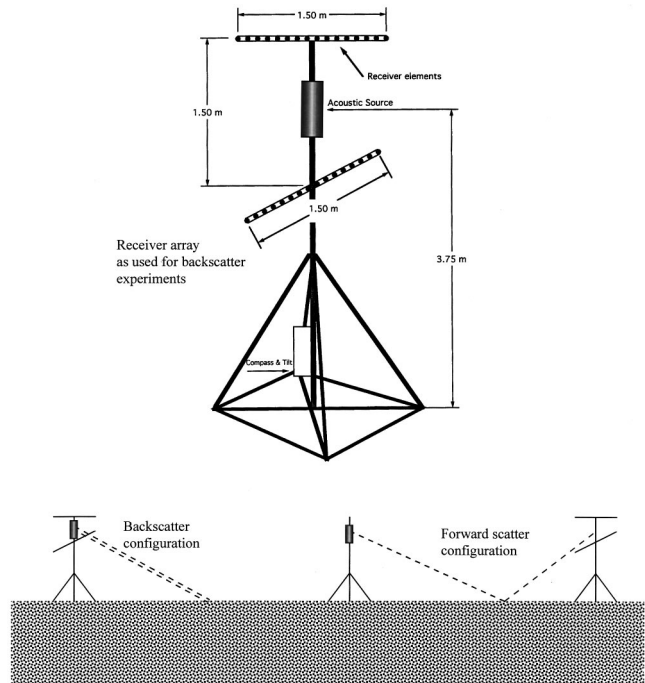


FIG. 2. Source and receiver array towers.



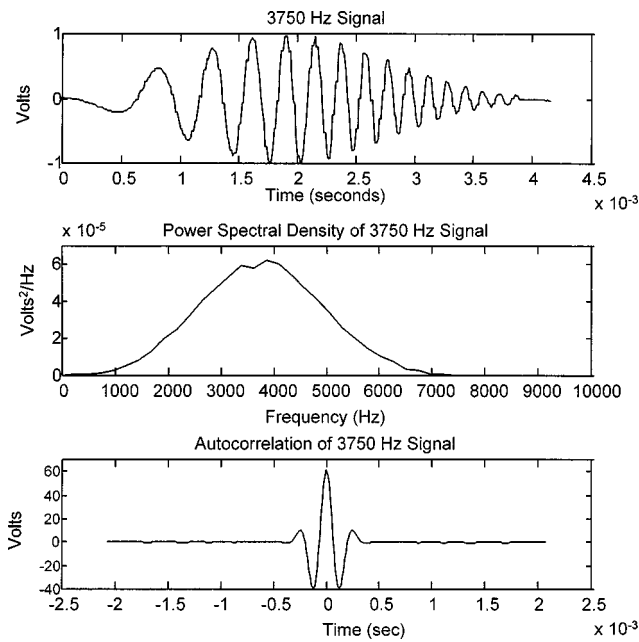


FIG. 3. 3750 Hz FM sweep signal.

to the vertical tower with the source axis in the vertical direction.

During all experiments, the source and receiver arrays were tethered to the ship via electrical cables. Source levels were controlled using a signal amplifier, which fed 100 V into the source. A function generator governed the shape and duration of source pulses. Output from the hydrophones and preamplifiers was amplified by a gain of 100. Experiments were controlled using an IBM compatible PC, which sampled the hydrophone output at 66.7 kHz.

### Broadband signal processing

At each measurement site, experiments were conducted using broadband FM sweeps centered at three frequencies, 3750 Hz, 7500 Hz, and 15000 Hz. The 3750 Hz pulse appears in Fig. 3. To remove ambient noise, 200 signals were emitted by the source, detected by the hydrophones, recorded by the data acquisition system, and averaged. Averaging a large number of shots was also an effective means of removing the more temporal returns due to sea surface reverberation (Rogers, 1996; Rogers and Yamamoto, under review).

Received signals were then beamformed. Figure 4 shows the array response (Appendix) at 15000 Hz for two incoming waves at 30° grazing angle and 0° bearing and at 30° grazing angle and 30° bearing. Bearing angle beamwidths are on the order of 5 to 10°, while grazing angle beamwidths are on the order of 2°. At most grazing angles, only small portions of the records were needed to estimate the scattering strength. Thus, beamforming was performed in the time domain:

$$s(t) = \frac{1}{N} \sum_{n=1}^N s_n(t + \mathbf{v} \cdot \mathbf{x}_n / C_w), \quad (1)$$

where  $s(t)$  is the beamformed signal in volts,  $N$  is the number of hydrophones,  $s_n(t)$  is the averaged, correlated signal at hydrophone  $n$ ,  $\mathbf{v}$  is a unit vector parallel to direction of

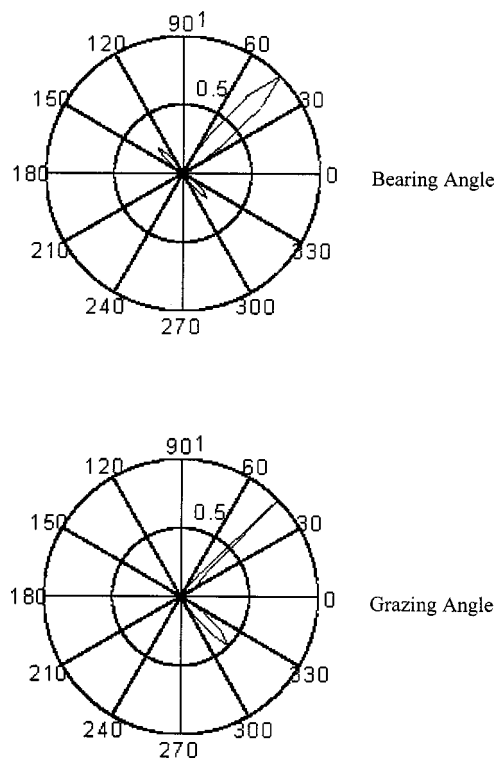


FIG. 4. Array response to an incoming beam at 45° bearing angle (top) and 45° grazing angle.

propagation,  $\mathbf{x}_n$  is a vector specifying the hydrophone location, and  $C_w$  is the sound speed in the water column. The mean square signals in the time interval  $t_1$  to  $t_2$  were then calculated given the beamformed signals:

$$\langle v^2 \rangle = \frac{1}{\Delta t} \int_{t_1}^{t_2} s^2(t) dt, \quad (2)$$

$$t_1 = t_{\text{travel}} - \tau_p/2, \quad t_2 = t_{\text{travel}} + \tau_p/2, \quad \Delta t = t_2 - t_1,$$

where  $\tau_p$  is the length of the pulse, which varied from  $1 \times 10^{-3}$  s for the 15000-Hz signal to  $4 \times 10^{-3}$  s for the 3750-Hz signal.

Backscattering and bistatic scattering strengths ( $10 \log_{10} \sigma_a$ ) were then calculated using the modified sonar equation:

$$\begin{aligned} 10 \log_{10} \sigma_a &= 10 \log_{10}(\sigma_{va} + \sigma_r) \\ &= 10 \log_{10} \langle v^2 \rangle \\ &\quad - \text{SI} + \text{DI} - \text{HS} - \text{PAmG} - \text{AmG} - \text{CCG} \\ &\quad - \text{ArG} + \text{TL}_{\text{inc}} + \text{TL}_{\text{scat}} - 10 \log_{10} \delta A, \end{aligned} \quad (3)$$

where  $\sigma_{va}$  and  $\sigma_r$  are the contributions of volume and roughness scattering to total scattering strength  $\sigma_a$ , SL is the source level in dB *re* 1  $\mu\text{Pa}^2$ , DI is the directivity index of the source, HS = -201 dB *re* 1  $\mu\text{Pa}$  per volt is the hydrophone sensitivity, PAmG is the preamplifier gain in dB, AmG is the amplifier gain, CCG is the cross correlation gain, ArG is the array gain,  $\text{TL}_{\text{inc}}$  and  $\text{TL}_{\text{scat}}$  are the incident and scattered transmission losses, and  $\delta A$  is the scattering area. The factor HS converted the mean square signal from volts to Pa. The parameter  $\delta A$  was determined using the broad-

band array response as shown in Fig. 4 and detailed in the Appendix.

In addition to providing the reverberation levels, broadband signal processing methods were used to establish the location of the source relative to the receiver array during the bistatic scattering experiments. The separation between the source and receiver was established by searching for the largest peaks in the averaged, cross-correlated time series. The arrival times of these peaks, corresponding to the direct propagation of the signal between the source and receiver array, were then used to establish the distance between the source and receiver. To determine the direction of the direct arrival, the signal was beamformed for various bearing angles and a grazing angle of zero. At site FS1, source and receiver were separated by 250 m. At site FS2, the source and receiver were approximately 80.7 m apart.

## RESULTS

### Geoacoustical description of measurement site

Sea floor scattering is most often described in terms of interface roughness spectra (Williams and Jackson, 1996), mean sound speeds, attenuation, and sound speed fluctuation spectra (Yamamoto, 1996). According to Williams and Jackson (1996), sea floor interface roughness scattering may be described in terms of the roughness spectrum

$$S_r(k) = \omega_2 k^{-\gamma}, \quad (4)$$

$$\sigma_r = f(k, \theta_{w1}, \theta_{w2}, \omega_2, \gamma, \phi_b, C_s, \rho_0), \quad (5)$$

where  $\omega_2$  is spectral strength,  $\gamma$  is a spectral exponent,  $C_s$  is the average sediment sound speed,  $\rho_0$  is the average sediment density,  $\mathbf{k}$  is the wave number,  $\phi_b$  is the difference between the scattered and incident bearing angles,  $\theta_{w1}$  is the incident grazing angle, and  $\theta_{w2}$  is the scattered grazing angle. A complete description of the roughness scattering model may be found in Williams and Jackson (1996).

Due to poor visibility of waters during our experiments, interface roughness statistics were not measured. Gravity cores taken during the experiments showed the sediments at the interface to be coarse sands mixed with significant amounts of shell fragments. Measured roughness statistics of similar sediments reported by Jackson and Briggs (1992) are fairly constant;  $\omega_2 = 0.0032$  to  $0.0042 \text{ cm}^4$  and  $\gamma = 3.18$  to  $3.25$ . These roughness statistics values were shown to closely predict the measured backscatter from the shell rich sandy sediments near the experiment sites (Rogers, 1996). The reader may use these values of roughness statistics as a first approximation input in models to compare with the forward scattering data.

According to Yamamoto (1996), plane-wave, sediment volume backscattering may be described in terms of the mean sediment sound speed  $C_s$ , the mean sediment attenuation  $\alpha$ , the mean sediment bulk density  $\rho_0$  the water density  $\rho_w \sim 1025 \text{ kg/m}^3$ , the sediment grain density  $\rho_r \sim 2650 \text{ kg/m}^3$ ,

$$\sigma_v = 2\pi |\mathbf{k}|^4 \Gamma^2 S_\mu(-2\mathbf{k}), \quad (6)$$

$$\Gamma = 1 + 2\zeta, \quad (7)$$

$$\zeta \approx \frac{(\rho_r - \rho_0)}{2\rho_0 - \rho_r}, \quad (8)$$

$$\sigma_{va} = \sigma_v \left( \frac{\sin \theta_s}{4\alpha} \right) T_{ws}^2 T_{sw}^2 \left( \frac{\cos^2 \theta_s \sin^2 \theta_w}{\cos^2 \theta_w \sin^2 \theta_s} \right), \quad (9)$$

$$\mu = (C - C_s)/C_s, \quad (10)$$

where  $\mathbf{k}$  is the wave vector,  $S_\mu$  is the energy spectrum of the sound speed fluctuation  $\mu$  within the sediments,  $T_{ws}$  and  $T_{sw}$  are the transmission coefficients between the water column and the sediment volume,  $\theta_w$  is the grazing angle in the water, and  $\theta_s$  is the grazing angle in the sediment, calculated based on  $C_s$  and Snell's law. The volume scattering coefficient as observed in the sediment volume is  $\sigma_v$ . The volume scattering coefficient as observed in the water column is  $\sigma_{va}$  (denoted as " $\sigma_a$ " in Yamamoto, 1996), and accounts for the effects of transmission, spreading, and attenuation within the sediment volume. The parameter  $\Gamma$  represents the contribution of density fluctuations to the volume scattering strength relative to that of the sound speed fluctuations. A high value of  $\Gamma$  indicates that a large portion of the scattering strength is due to density fluctuations.

In most cases, a power law can approximate the sediment sound speed fluctuation spectrum  $S_\mu(k)$ :

$$S_\mu(\mathbf{k}) = (\beta \Lambda^2 B / 2\pi) [\Lambda^2 (k_x \cos \phi \cos \delta + k_y \sin \phi \cos \delta - k_z \sin \delta)^2 + \Lambda^2 (-k_x \sin \phi + k_y \cos \phi)^2 + (k_x \cos \phi \sin \delta + k_y \sin \phi \sin \delta + k_z \cos \delta)^2]^{-(\beta/2+1)}, \quad (11)$$

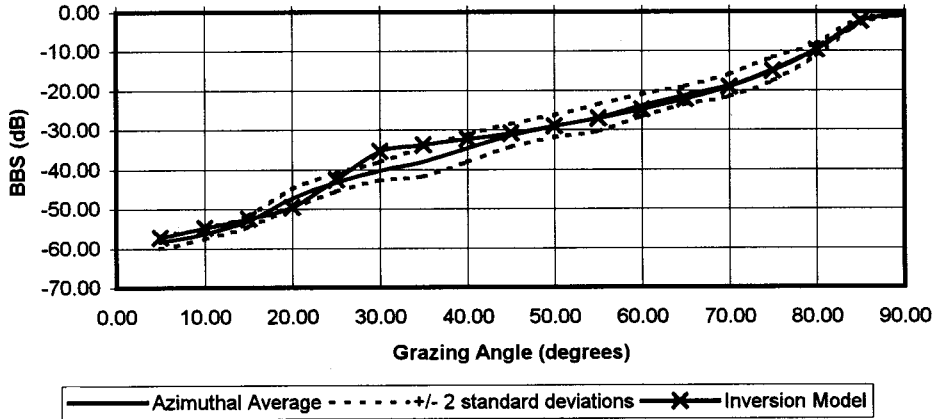
where  $B$  is the spectral strength,  $\beta$  is the spectral exponent,  $\Lambda$  is the aspect ratio,  $\delta$  is the dip angle, and  $\phi$  is the dip bearing, measured counterclockwise from due north. The spectrum described by Eq. (11) results in elliptical surfaces of equal  $S_\mu(k_x, k_y, k_z)$ . The aspect ratio  $\Lambda$  is the ratio of the major axis of an elliptical surface to its minor axes (Yamamoto, 1996), and represents the anisotropy of the sound speed fluctuation intensity. Sediments with well-defined layers have high values of  $\Lambda$ . The dip angle represents the deviation of the major axes from the vertical and, consequently, the minor axis from the horizontal. In sediments with well-defined layers, dip may be observed in the inclination of the layers. If the sediment layers tilt due north, the dip bearing is  $0^\circ$ , etc. For  $\delta=0$ ,  $k_z$  becomes the major axis, while  $k_x$  and  $k_y$  become the minor axes.

A similar model plane wave could be developed for bistatic scattering by generalizing the proportionality constant [Eq. (9)] between  $\sigma_{va}$  and  $\sigma_v$  for bistatic geometry (Yamamoto, 1996). However, when grazing angles are low ( $<28^\circ$  for the present experiments), the transmission of acoustic energy to the sediment volume usually occurs in the form of evanescent waves or shear waves. Thus, an adaptation of the Yamamoto (1996) volume scattering model [Eqs. (6)–(10)] for bistatic geometry would only be valid for geometries in which the incident and scattered grazing angles are above the critical grazing angle.

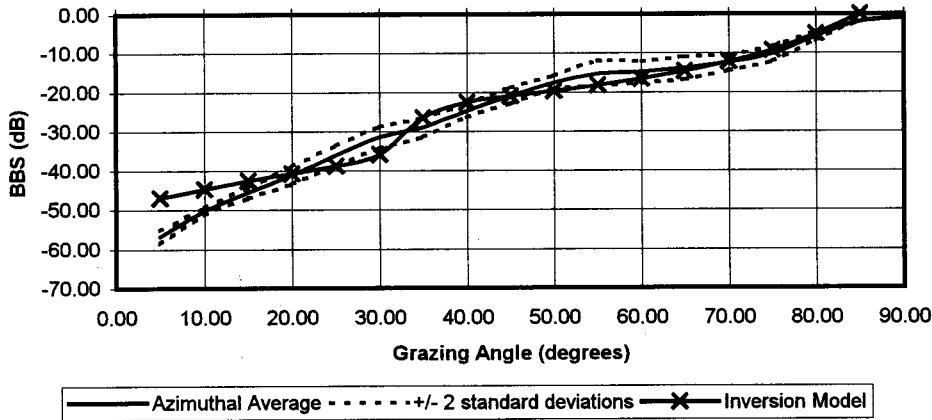
TABLE III. Volume scattering parameters estimated by crosswell tomography at 27° 37.4' N, 80° 18.6' W, 14 nautical miles northwest of the site FS2 (Rapids *et al.*, 1997; Rogers and Yamamoto, 1997).

Section	$B$	$\beta$	Aspect ratio	Dip (deg)	Dip bearing (deg)	Mean sound speed (m/s)	Atten. (dB/m/kHz)
A-B	$4.98 \times 10^{-5}$	0.312	4.6	14.10	110	1559	0.15
A-C	$1.27 \times 10^{-5}$	0.214	4.4	1.36	20	1571	0.15
A-D	$2.29 \times 10^{-5}$	0.778	6.8	1.39	20	1565	0.15

### 3750 Hz Bottom Backscatter (BBS), Fort Pierce Site S4



### 7500 Hz Bottom Backscatter (BBS), Fort Pierce Site S4



### 15000 Hz Bottom Backscatter (BBS), Fort Pierce Site S4

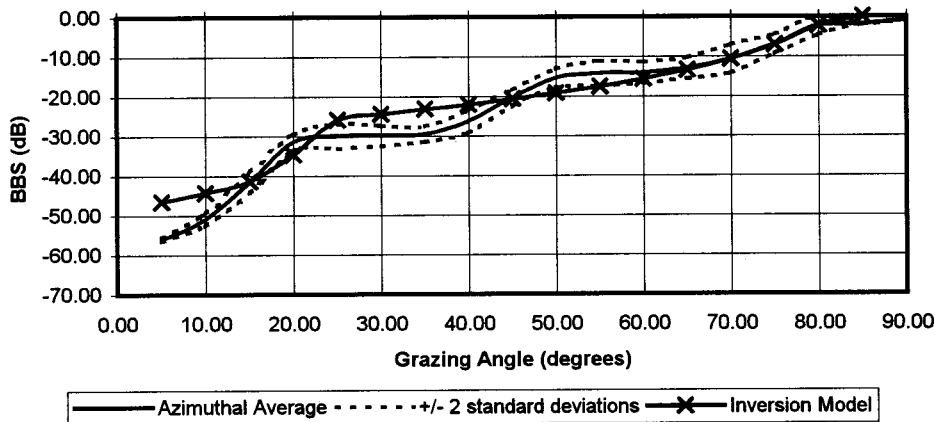


FIG. 5. Measured and theoretical backscatter at Site S4.

TABLE IV. Volume scattering parameters estimated by inversion of azimuthally averaged backscattering strength at site S4, adjacent to site FS2.

Frequency (Hz)	3750	7500	15 000	Mean	$\sigma$	Lower	Upper
Porosity	38%	30%	42%	37%	6%	30%	42%
$\alpha$ (dB/m/kHz)	0.4	0.5	0.28	0.39	0.11	0.28	0.50
$\log_{10}(B)$	-2.2	-2.8	-2.4	-2.47	0.31	-2.80	-2.20
$\beta$	1.43	0.43	0.83	0.90	0.50	0.43	1.43
$\Lambda$	11.8	20.2	34.6	22.2	11.5	11.8	34.6
$\log_{10}(\omega_2 \text{ in cm}^4)$	-2.8	-2.6	-3.2	-2.9	0.3	-3.2	-2.6
$\gamma$	2.3	2.9	1.6	2.3	0.6	1.6	2.9
Density (kg/cubic m)	2033	2163	1968	2054	99	1968	2163
Sound speed (m/s)	1726	1883	1668	1759.0	111	1668	1883
$B$	0.006 31	0.00 58	0.003 98	0.003 96	0.002 36	0.001 58	0.006 31
$\omega_2$ (cm <sup>4</sup> )	0.001 45	0.002 75	0.000 63	0.001 61	0.001 07	0.000 63	0.002 75
Root mean square error between data and model	5%	10%	14%				

Given the Williams and Jackson (1996) roughness scattering model and the current Yamamoto (1996) backscattering model, nine parameters are necessary for a satisfactory geoacoustical description of a given site:

- The roughness spectrum parameters  $\omega_2$  and  $\gamma$ .
- The mean sediment sound speed  $C_s$ .
- The mean attenuation  $\alpha$ .
- The sound speed fluctuation spectrum parameters  $B$ ,  $\beta$ , and  $\Lambda$ .
- The dip angle  $\delta$  and dip bearing  $\phi$ .

Three independent methods of estimating the volume scattering parameters were available for the bistatic scattering measurement site: core samples, crosswell tomography measurements (Rapids *et al.*, 1998), and the inversion of backscattering measurements (Rogers, 1996; Rogers and Yamamoto, under review). A second set of core samples and *in-situ* sound speed and shear wave measurements at the site were taken by Richardson and Briggs (1999) following the experiment. The data extracted from those samples give an average sound speed of approximately 1700 m/s. Future modelers may use these geoacoustic data in their models of bistatic scattering.

Crosswell tomography measurements taken at 27° 37.4' N, 80° 18.6' W are discussed in Rapids *et al.* (1998), and are also summarized in Rogers (1996). Sound speeds at that site average 1565 m/s and range from 1500 to 1700 m/s over a depth of 0 to 2 m below the sea floor. The sound speeds are characterized by a slower layer near the sea floor, and a faster layer approximately 1.5 m below the sea floor. Volume scattering parameters derived from these measurements were estimated by Rogers and Yamamoto (under review) appear in Table III.

Figure 5 shows the bottom backscattering strength at site S4, adjacent to site FS2. Backscattering measurements are averaged with respect to bearing and are given for every 5 degrees of grazing angle  $\theta$ . Between  $\theta=5^\circ$  and  $85^\circ$ , backscattering strength increases from -59 dB to -2 dB at  $f=3750$  Hz, from -47 dB to -3 dB at  $f=7500$  Hz, and from -46 dB to -2 dB at  $f=15000$ .

Volume and roughness scattering parameters representing the best fit to the data in Fig. 5 appear in Table IV. They have been estimated by adapting the inversion method of

Rogers and Yamamoto (under review). Both the  $\omega_2$  and  $\gamma$  values are slightly smaller than those of Jackson and Briggs (1992). However, the inverted roughness statistics are reasonable values for sands mixed with shell fragments. The inversion shows that the aspect ratio  $\Lambda$  is much higher than that of Rogers and Yamamoto (under review); sound speed fluctuations at site S4 are highly anisotropic. In addition, the attenuation and the spectral parameter  $B$  are considerably higher than the range of values calculated by Rogers and Yamamoto (under review). However, the values of  $\beta$  are of the same order of magnitude. Furthermore, the range of inverted sound speeds (1668 to 1883 m/s) encompasses the 1700 m/s average sound speed estimated via core samples and *in-situ* measurements (Richardson and Briggs, 1999) and the upper range of sound speeds given in Rapids *et al.* (1998).

Given both the geographical separation between the sites, the dredging activity at site S4, and the presence of storm activity prior to the experiment, differences between the tomographic measurements and the inversion results are reasonable.

### Bistatic scattering time series data

Cross-correlated time series of the bistatic scattering data at site FS2 appear in Figs. 6–8. The time series are characterized by weak first arrivals, followed by strong arrivals and subsequent weaker arrivals. The first arrivals correspond to head waves arriving at the receiver array. The second set of arrivals corresponds to the direct propagation of acoustic energy through the water column from the source to the receiver. The third set of arrivals corresponds to energy that has been scattered by the sea floor.

A closer view of the 15 000 Hz signal at hydrophone 6 appears in Fig. 9. Given averaging and cross-correlation, noise levels as observed by the data acquisition system are 0 to 5 dB *re* 1 volt<sup>2</sup>. Head wave arrivals are 15 to 20 dB above the ambient noise. Direct arrivals are 35 to 40 dB above the noise levels. Later arrivals, corresponding to scattered energy, are 5 to 15 dB above the ambient noise.

Time series taken at site FS1 were not used for the estimation of bistatic scattering strengths. Except for direct ar-



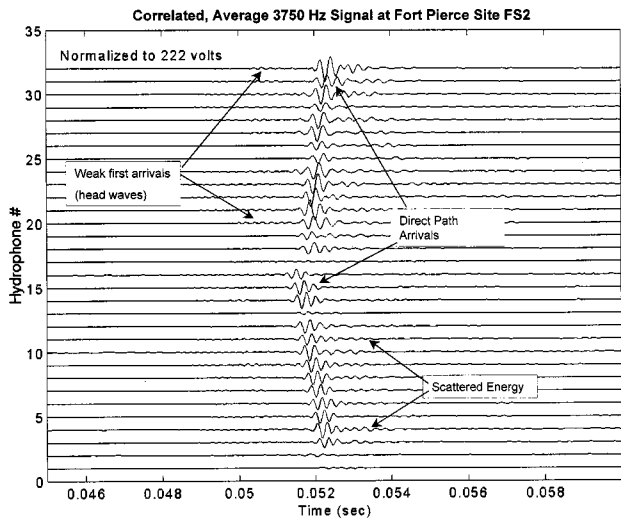


FIG. 6. Received 3750 Hz signal at bistatic scattering measurement site FS2. Figure represents an average of 200 shots, cross-correlated with the source signal.

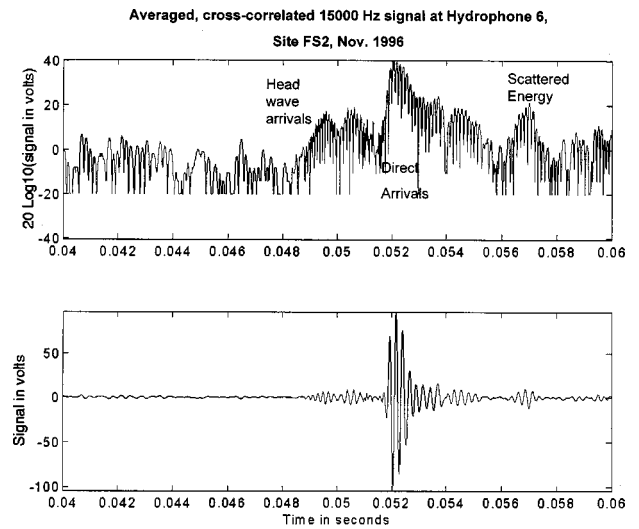


FIG. 9. Received 15 000 Hz signal at hydrophone 6 at bistatic scattering measurement site FS2. Figure represents an average of 200 shots, cross-correlated with the source signal.

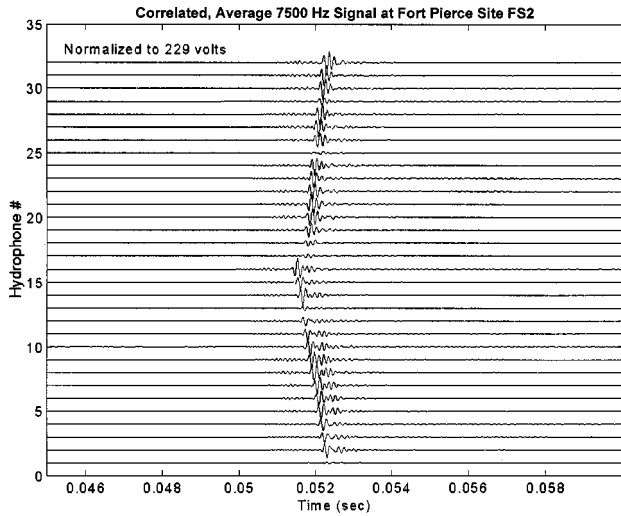


FIG. 7. Received 7500 Hz signal at bistatic scattering measurement site FS2. Figure represents an average of 200 shots, cross-correlated with the source signal.

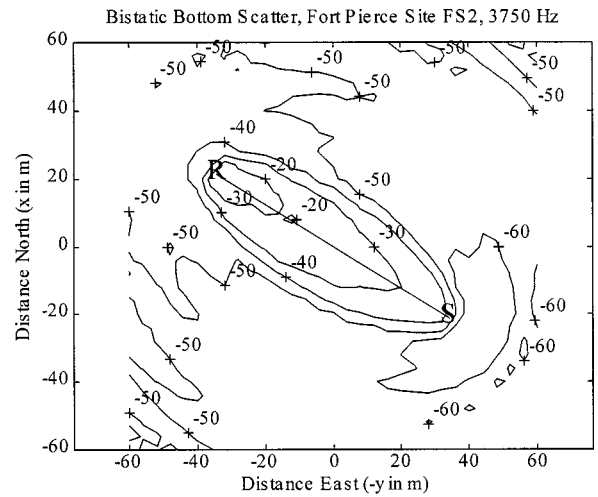


FIG. 10. 3750 Hz bottom bistatic scattering strength at site FS2.

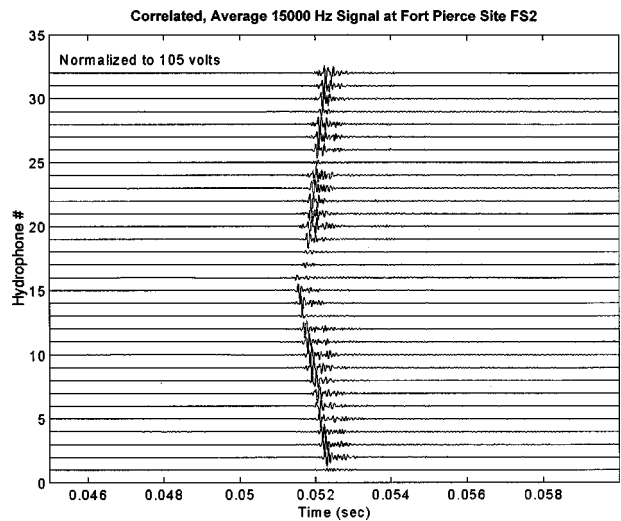


FIG. 8. Received 15 000 Hz signal at bistatic scattering measurement site FS2. Figure represents an average of 200 shots, cross-correlated with the source signal.

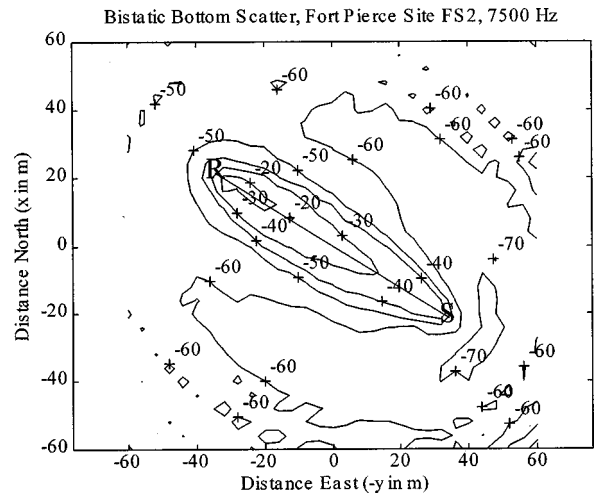


FIG. 11. 7500 Hz bottom bistatic scattering strength at site FS2.

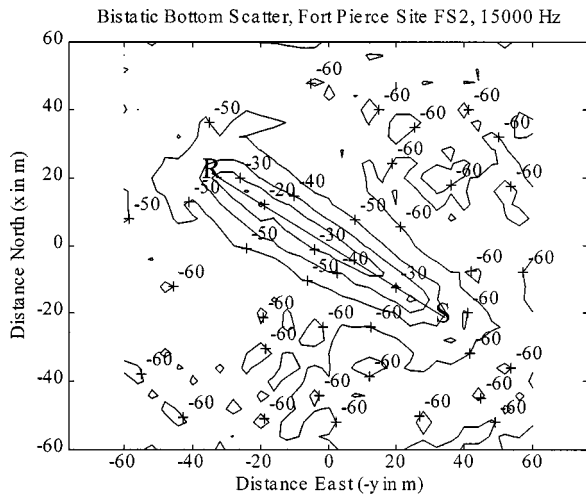


FIG. 12. 15 000 Hz bottom bistatic scattering strength at site FS2.

rivals, signal-to-noise ratios were not high enough to accurately estimate bistatic scattering strengths.

### Bistatic scattering levels

Bottom bistatic scattering strength (BSS) levels at site FS2 are shown as a function of position in Figs. 10–12. The source position  $S$  is 34.2 m east, 21.4 m south, while the receiver position  $R$  is 34.2 m west, 21.4 m north. The source–receiver separation is 80.7 m. For each of the frequencies shown, the bistatic scattering pattern is characterized by a 60 m long ellipsoidal region of high scattering strength (BSS  $> -30$  dB) centered approximately 30 m from the receiver along transect  $S-R$ , the line between the source

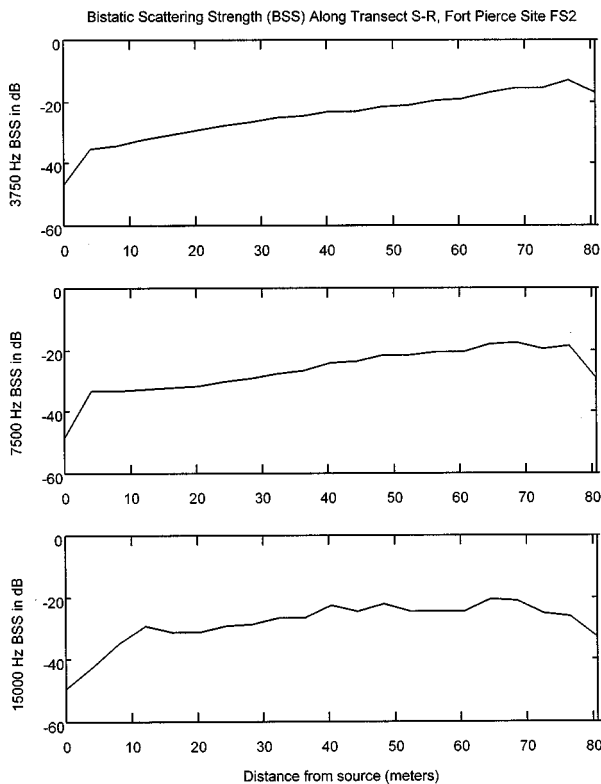


FIG. 13. Bottom bistatic scattering strength at site FS2 along transect  $S-R$ .

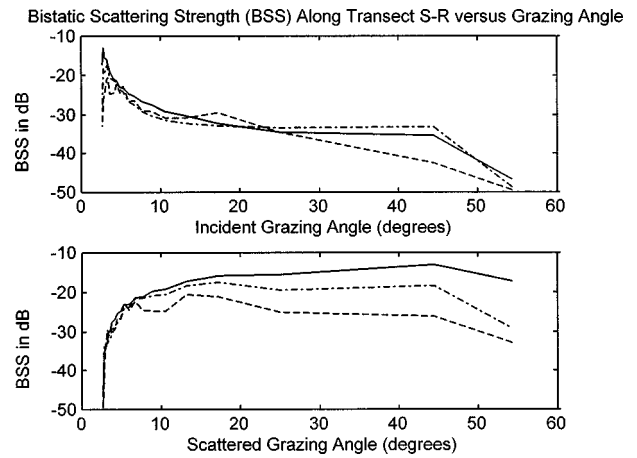


FIG. 14. Bottom bistatic scattering strength at site FS2 along transect  $S-R$  versus angle.

and receiver. The width of the high BSS region at  $f = 3750$  Hz is 30 m wide, at  $f = 7500$  Hz is 20 m wide, and at  $f = 15000$  Hz is 10 m wide.

The bistatic scattering patterns are not symmetric about the midpoint between the source and receiver. Scattering strength is stronger near the receiver than it is near the source. Surrounding the high BSS region at each of the frequencies shown is an irregularly shaped region of moderate ( $-50$  to  $-30$  dB) BSS. At  $f = 3750$ , the shape of this region, roughly symmetric about transect  $S-R$ , is elliptical between the source and the midpoint of transect  $S-R$ . The corresponding width of the elliptical section of the regions is approximately 40 m, with the remainder of the region approximating a semicircle approximately 85 m in diameter (Fig. 10). At  $f = 7500$  Hz, the moderate BSS region is elliptical, being approximately 35 m wide. At  $f = 15000$  Hz, the moderate BSS region, centered on transect  $S-R$ , is roughly trapezoidal, being 20 m near the source and 35 to 50 m wide near the receiver.

The remainder of the domain shown in Figs. 10–12 is

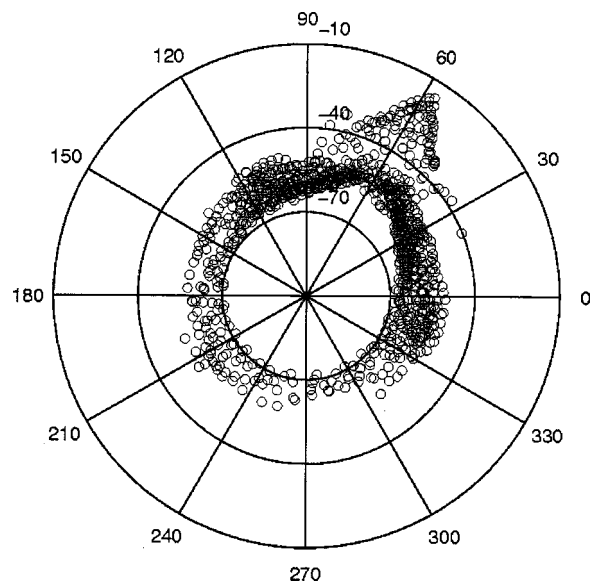


FIG. 15. 7500 Hz bottom bistatic scattering strength at site FS2 as a function of incident bearing angle.

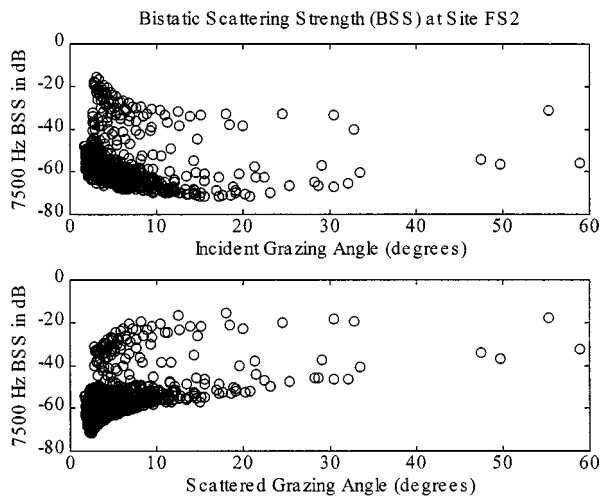


FIG. 16. 7500 Hz bottom bistatic scattering strength at site FS2 as a function of grazing angle.

characterized by low BSS values on the order of  $-70$  to  $-60$  dB. The smallest BSS values appear approximately 12 to 15 m southeast of the source.

Bistatic scattering levels along the transect between the source and receiver are shown in Fig. 13. Bistatic scattering strengths are strongly dependent on position but show no consistent variation with frequency. Within 4 m of the source, the bistatic scattering strength increases from  $-50$  to  $-47$  dB to the range from  $-40$  to  $-36$  dB. Between 4 and 64 m from the source, the bistatic scattering strength exhibits a more gradual increase to the range from  $-20$  to  $-15$  dB. In the remaining 16.5 m distance to the receiver, the bistatic scattering level drops to the range of  $-17$  to  $-34$  dB.

Figures 10–13 show that bistatic scattering patterns along transect  $S-R$  is not symmetric with respect to the midpoint of the transect. This lack of symmetry becomes evident in Fig. 14, which displays bistatic scattering strength along transect  $S-R$  as a function of grazing angle. Near the source (incident grazing angle  $>45^\circ$ ), the BSS decreases with increasing grazing angle. This is because a large portion of acoustic energy is transmitted into the sediment and either absorbed or trapped. Thus, only a portion of the incident acoustic energy is scattered by surface roughness. Near the receiver (incident grazing angle  $<45^\circ$ ), the opposite is true. Most or all of the incident energy is scattered by surface roughness. Contrary to the incident grazing angle dependence, acoustic scattering is almost independent of scattered grazing angle.

Figure 15 shows the bistatic scattering strength at 7500 Hz as a function of incident bearing given the entire range of grazing angles. The maximum scattering strengths occur at  $58^\circ$  incident bearing angle, the direct bearing from the source to the receiver. Within  $\pm 30^\circ$  of the bearing from source to receiver, there is a noticeable degree of spread in the scattering data due to a dependence on grazing angle. However, outside these limits, the scattering strength is low ( $<-50$  dB). Figure 16 shows the bistatic scattering strength at 7500 Hz as a function of grazing angle given the entire range of bearing angles. For steeper ( $>30^\circ$ ) incident grazing angles, the BSS is generally lower, while for steeper scattered graz-

ing angles, the BSS is generally higher. Nevertheless, due to the dependence of the BSS data on bearing angle, its degree of spread is very large (20 to 60 dB). Given the 80.7-meter separation between source and receiver, the primary angular dependence of the bistatic scattering data is by incident bearing angle. The asymmetry of the scattering pattern about the midpoint between the source and receiver is due to the incident grazing angle and resulting transmission of energy into the sea bed.

## DISCUSSION

The bistatic scattering data is difficult to compare with previous bistatic experiments. The 20 to 180 kHz data presented by Stanic *et al.* (1991) was taken using a source and receiver in very close proximity (50 cm) and is of very different geometry than the data in Figs. 10–12. The 22 kHz data presented by Urick (1960) was taken using a source and receiver located 2.7 km apart. The scattering pattern appearing in Urick (1960) varies only by 12 dB over the entire domain of his experiment; bistatic scattering strengths are on the order of  $-30$  dB. In contrast, the scattering patterns in Figs. 10–12 vary widely with position.

The bistatic scattering patterns are an important indicator of how sound propagates near the sea floor at the measurement site. Figures 10–16 demonstrate that between the source and receiver, the majority of energy propagating between the source and receiver travels along the direct line between the source and receiver when scattered by the sea floor sediments. The observed bistatic scattering patterns are clearly asymmetric with respect to the center of the domain (Figs. 10–12), being larger near the receiver than near the source. At a steeper incident grazing angle, more energy is transmitted into and absorbed or trapped in the sediments. Thus, less energy is scattered by the interface roughness. At a smaller incident grazing angle, most or all energy is scattered by roughness.

It is important to note that most of the incident and scattered grazing angles in this particular bistatic scattering experiment are below critical ( $\sim 28^\circ$ ). In this case, most of the energy enters the sediment either in the form of evanescent waves or shear waves. Some authors have stated that given the transfer of energy in the form of evanescent waves, no volume scattering would occur (Jackson *et al.*, 1986). However, Hines (1996) has developed a volume scattering theory which shows that the contribution of evanescent waves to the bistatic volume scattering strength can be significant. If the transfer of acoustic energy to the sediment at subcritical grazing angles occurs in the form of shear waves, volume scattering of shear waves may occur. An accurate delineation of the roughness and volume scattering mechanisms in this experiment will require a better understanding of volume scattering at subcritical grazing angles.

## RECOMMENDATIONS

A forward volume scattering theory that describes scattering below the critical grazing angle will be needed to interpret the results of this experiment and similar experiments. Although Hines (1996) volume scattering model describes

such scattering along the transect between source and receiver, it cannot describe bistatic volume scattering at other locations. For frequencies and sediment types where acoustic energy is transferred to the sediments in the form of shear waves, volume scattering models may be applicable. A conclusive evaluation of the scattering physics in this experiment will require a thorough understanding both evanescent scattering and the conversion of plane waves in the water column to shear waves within the sediments.

## CONCLUSIONS

Observations of bistatic scattering at moderate frequencies 3750 Hz, 7500 Hz, and 15 000 Hz and low grazing angles ( $2^\circ$ – $20^\circ$ ) have been presented with a geoacoustical description of the site. Given a source–receiver separation of 80.7 m, bistatic scattering strengths in the sandy sediments at the measurement site range from  $-70$  to  $-15$  dB. Bistatic scattering at low grazing angles depends primarily on the incident bearing angle. However, the asymmetry of the scattering pattern about the midpoint between the source and receiver is due to the incident grazing angle and resulting transmission of energy into the sea bed. Bistatic scattering values are highest along the transect between the source and receiver. Maximum scattering strengths ( $-20$  to  $-15$  dB) appear near the receiver. Regions of high scattering ( $> -30$  dB) strength are centered 30 m from the receiver along the direct path between the source and receiver. However, given the sands and shell fragments lining the sea floor at the measurement site, roughness scattering is a major contributor to the observed bistatic scattering patterns.

## ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research Ocean Acoustic Program. The authors would also like to thank Captain Dan Schwarz and the crew of the R/V SEWARD JOHNSON, for leading a productive cruise and improving the deployment method of the backscatter array. The authors are grateful for the assistance of Murat Kuru, Dr. Tom Nye, and Dr. Andrew Rogers, who assisted with the operation of the data acquisition systems and deployment of the acoustic array. Finally, the authors would like to thank Dr. F. D. Tappert for his recommendations and suggestions on the presentation of the measurements appearing in this paper.

## APPENDIX: ARRAY RESPONSE FUNCTION

The broadband array response function  $A(\phi_n, \theta_n, \phi, \theta, s(t))$  represents the ability of a receiver array to distinguish an acoustic pulse  $s(t)$  traveling in a given direction. The bearing angle  $\phi_n$  and grazing angle  $\theta_n$  represent the direction of propagation. The bearing angle  $\phi$  and grazing angle  $\theta$  represent the direction to which the signals received by the array elements are beamformed.

Let  $s(t)$  be a given acoustic pulse, such as the FM sweep shown in Fig. 3. For a given direction of propagation, the signal as observed at each hydrophone “ $i$ ” would be

$$s_i(t) = s(t - \mathbf{v} \cdot \mathbf{x}_i / C_w), \quad (\text{A1})$$

$$\mathbf{v} = \mathbf{i}_x \cos \theta_n \cos \phi_n + \mathbf{i}_y \cos \theta_n \sin \phi_n + \mathbf{i}_z \sin \theta_n,$$

where  $\mathbf{x}_i$  is the vector specifying the location of hydrophone  $i$ ,  $C_w$  is the sound speed, and  $\mathbf{i}_x$ ,  $\mathbf{i}_y$ , and  $\mathbf{i}_z$  are unit vectors in the  $x$ ,  $y$ , and  $z$  directions. Given zero ambient noise, the signal  $s(t)$  may be resolved by beamforming the observed signals  $s_i(t)$ :

$$s(t) = \frac{1}{N} \sum_{i=1}^N s_i(t + \mathbf{v} \cdot \mathbf{x}_i / C_w). \quad (\text{A2})$$

When the observed signal is beamformed to a direction  $(\phi, \theta)$  other than the direction of propagation  $(\phi_n, \theta_n)$ , the result is not zero. Instead, the apparent beamformed signal is

$$r(t, \phi, \theta) = \frac{1}{N} \sum_{i=1}^N s_i(t + \mathbf{u} \cdot \mathbf{x}_i / C_w), \quad (\text{A3})$$

$$\mathbf{u} = \mathbf{i}_x \cos \theta \cos \phi + \mathbf{i}_y \cos \theta \sin \phi + \mathbf{i}_z \sin \theta.$$

Given  $r(t, \phi, \theta)$  and  $s(t)$ , the array response pattern may be given as

$$A(\phi_n, \theta_n, \phi, \theta, s(t)) = V_r(\phi, \theta) / V_s, \quad (\text{A4})$$

$$V_r(\phi, \theta) = \int_{f_1}^{f_2} G_r(f) df, \quad V_s = \int_{f_1}^{f_2} G_s(f) df, \quad (\text{A5})$$

where frequencies  $f_1$  and  $f_2$  define the bandwidth of interest, and  $G_r(f)$  and  $G_s(f)$  are the power spectra density of  $r(t, \phi, \theta)$  and  $s(t)$ .

The parameter  $\delta A$ , used in the modified sonar equation [Eq. (3)] may be determined using the array response function. If a signal  $s(t)$  originating from the sea floor is known to be propagating in a direction  $(\phi_n, \theta_n)$ , let  $I(\phi_n, \theta_n, t)$  be the intensity of that signal. If the hydrophone array is beamformed to a direction  $(\phi, \theta)$  other than  $(\phi_n, \theta_n)$ , let  $I(\phi, \theta, t)$  be the intensity of the apparent signal  $r(t, \phi, \theta)$ . The intensities  $I(\phi_n, \theta_n, t)$  and  $I(\phi, \theta, t)$  are proportional to  $s^2(t)$  and  $r^2(t, \phi, \theta)$ , respectively. The parameter  $\delta A$  may be estimated as

$$\delta A(\phi_n, \theta_n) = \frac{\int_{\phi=0}^{2\pi} \int_{\theta=-\pi/2}^{\pi/2} \langle I(\phi, \theta, t) \rangle R^2 \cos \theta d\theta d\phi}{\langle I(\phi_n, \theta_n, t) \rangle \sin \theta_n}, \quad (\text{A6})$$

$$R = Z / \sin \theta_n, \quad (\text{A7})$$

where  $Z$  represents the height of the array above the sea floor, and the brackets  $\langle \rangle$  represent the average over the pulse length. In this estimation of the scattering area,  $\delta A$  is

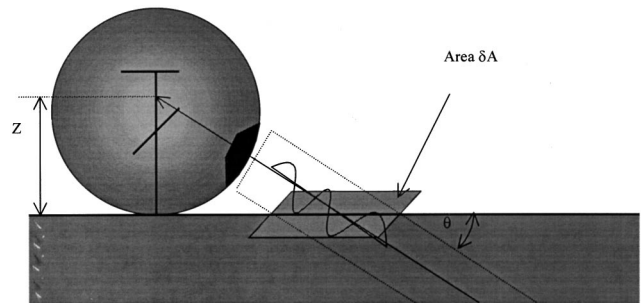


FIG. A1. Definition of scattering area  $\delta A$ .



patch of area on a sphere of radius “ $R$ ,” centered at grazing angle  $\theta_n$ , and projected onto the seafloor (Fig. A1).

If the bandwidth of the signal  $s(t)$  is limited to the range of frequencies between  $f_1$  and  $f_2$ ,  $I(\phi_n, \theta_n, t)$  and  $I(\phi, \theta, t)$  become proportional to  $V_s$  and  $V_r(\phi, \theta)$ . As  $V_s$  is a constant with respect to  $(\phi, \theta)$  (but not  $\phi_n, \theta_n$ ), the area  $\delta A$  may be calculated as

$$\delta A(\phi_n, \theta_n) = \int_{\phi=0}^{2\pi} \int_{\theta=-\pi/2}^{\pi/2} A(\phi_n, \theta_n, \phi, \theta, s(t)) R^2 \times \cos \theta d\theta d\phi. \quad (\text{A8})$$

For the array used in this experiment, the scattering area  $\delta A$  as approximated in Eqs. (A6)–(A8) depends primarily on the grazing angle  $\theta_n$ , with bearing angle dependence  $\phi_n$  being only on the order of 1 to 2 dB over the range  $\phi_n = 0$  to  $360^\circ$ . To simplify the processing of data, a value of  $\delta A$  averaged with respect to  $\phi_n$  was used.

Boyle, F. A., and Chotiros, N. P. (1995). “A Model for Acoustic Backscatter from Muddy Sediments,” *J. Acoust. Soc. Am.* **98**, 531–541.

Chotiros, N. P. (1995). “Biot Model of Sound Propagation in Water Saturated Sand,” *J. Acoust. Soc. Am.* **97**, 199–214.

Essen, H. H. (1994). “Scattering from a Rough Sedimental Seafloor Containing Shear and Layering,” *J. Acoust. Soc. Am.* **95**, 1299–1310.

Gensane, M. (1993). “Sea Bottom Reverberation: the Role of Volume Inhomogeneities of the Sediment,” in *Ocean Reverberation*, edited by D. D. Ellis *et al.* (Kluwer Academic, Dordrecht, The Netherlands), pp. 59–64.

Hines, P. C. (1996). “Theoretical Model of In-Plane Scatter from a Smooth Sediment Seabed,” *J. Acoust. Soc. Am.* **99**, 836–844.

Jackson, D. R., and Briggs, K. (1992). “High Frequency Bottom Backscattering: Roughness Scattering versus Volume Scattering,” *J. Acoust. Soc. Am.* **92**, 962–977.

Jackson, D. R., Winebrenner, D. P., and Ishimaru, A. (1986). “Application of the Composite Roughness Model to High Frequency Bottom Backscattering,” *J. Acoust. Soc. Am.* **79**, 1410–1422.

Lyons, A. P., Anderson, A. L., and Dawn, F. S. (1994). “Acoustic Scattering from the Seafloor-Modeling and Data Comparison,” *J. Acoust. Soc. Am.* **95**, 2441–2541.

Merklinger, H. M. (1968). “Bottom Reverberation Measured with Explosive Charges Fired in the Deep Ocean,” *J. Acoust. Soc. Am.* **44**, 508–513.

Moe, J. E., Thorsos, E. I., Jackson, D. R., and Williams, K. R. (1995). “The Effect of Roughness on Acoustic Penetration of the Seafloor as Given by a Fluid-Fluid Perturbation Model and Comparison with Recent Sediment Penetration Experiments,” *J. Acoust. Soc. Am.* **97**, 3315–3316.

Mourad, P. D., and Jackson, D. R. (1993). “A Model/Data Comparison for Low Frequency Backscatter,” *J. Acoust. Soc. Am.* **94**, 344–358.

Richardson, M. D., and Briggs, K. B. (1999). Personal communication.

Rapids, B. R., Nye, T., and Yamamoto, T. (1998). “High Resolution Acoustical Crosshole Tomography Experiment in a Marine Sediment,” *J. Acoust. Soc. Am.* **103**, 212–224.

Rogers, A. K. (1996). “Acoustic Scattering from Marine Sediment: Modeling with Experimental Verification,” Ph.D. Dissertation, University of Miami.

Rogers, A. K., and Yamamoto, T. (1999). “Analysis of High Frequency Acoustic Scattering Data Measured in the Shallow Waters of the Florida Straits,” *J. Acoust. Soc. Am.* (to be published).

Stanic, S., Kennedy, E., and Ray, R. I. (1991). “Variability of Shallow Water Bistatic Bottom Backscattering,” *J. Acoust. Soc. Am.* **90**, 547–553.

Swift, S., and Stephen, R. A. (1994). “The scattering of a low-angle pulse beam from seafloor volume heterogeneities,” *J. Acoust. Soc. Am.* **96**, 991–1001.

Urlick, R. J. (1960). “Side Scattering of Sound in Shallow Water,” *J. Acoust. Soc. Am.* **32**, 351–355.

Williams, K. L., and Jackson, D. R. (1996). “A Model for Bistatic Scattering into Ocean Sediments for Frequencies from 10–100 kHz,” Technical Report No. APL-UW TR 9505, Applied Physics Laboratory—University of Washington, Seattle, Washington.

Wood, A. B. (1941). *A Textbook of Sound* (Bell, London).

Yamamoto, T., Turgut, A., Shulkin, M., and Bennet, R. (1991). “Geoacoustic Properties of the Seabed Sediment Critical to Acoustic Reverberation at 50 to 500 Hz: A Preliminary Data Set,” RSMAS Technical Report No. TR-91-001, GeoAcoustical Laboratory Report No. 1014, University of Miami, Miami, Florida.

Yamamoto, T., Nye, T., and Kuru, M. (1994). “Porosity, Permeability, Shear Strength: Crosswell Tomography Below an Iron Foundry,” *Geophysics* **59**, 1530–1541.

Yamamoto, T. (1995). “Velocity Variabilities and Other Physical Properties of Marine Sediments Measured by Crosswell Acoustic Tomography,” *J. Acoust. Soc. Am.* **98**, 2235–2248.

Yamamoto, T. (1996). “Acoustic Scattering in the Ocean from Velocity and Density Fluctuations in the Sediments,” *J. Acoust. Soc. Am.* **99**, 866–879.

Yamamoto, T., Day, C. M., Kuru, M., and Nye, T. (1997). “Acoustic Scattering from Sediment Volume in Shallow Water,” Proceedings of the International Conference on Shallow Water Acoustics, April 21–25, 1997, Beijing, China.

# Monostatic and bistatic reverberation statistics west of the Mid-Atlantic Ridge

Yevgeniy Y. Dorfman<sup>a)</sup> and Ira Dyer

Department of Ocean Engineering, MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

(Received 23 February 1999; revised 2 June 1999; accepted 11 June 1999)

Monostatic and bistatic reverberation of highly resolved signals from very rough bottoms at one site is statistically analyzed. Scattering at a mean frequency of about 230 Hz from a large number of bottom footprints is considered. The reverberation envelope is found to be non-Rayleigh, with the degree of departure from Rayleigh dependent upon the bottom grazing angle for two cases considered,  $\theta_g \approx 5^\circ$  and  $40^\circ$ , and upon the bistatic angle in the entire range,  $\phi \approx 0^\circ$  to  $180^\circ$ . These rough bottom observations can be explained by adopting a continuous scattering model having a Rayleigh envelope, added to a discrete scattering model (arising from a small number of individual features within the sonar footprint) having a distinctly non-Rayleigh envelope. These models, plus a heuristic mechanism of self-selection within the discrete scattering model, arguably explain the observed angle dependence of the reverberation statistics. © 1999 Acoustical Society of America. [S0001-4966(99)05010-9]

PACS numbers: 43.30.Gv, 43.30.Re, 43.30.Hw [DLB]

## INTRODUCTION

Scattering of sound from rough ocean bottoms is a major source of active sonar interference. As measured and affected by properties of the sonar, such interference is known as reverberation. Its statistical properties are typically used to design and operate signal processors intended to counter its effects. Rough bottom reverberation is commonly assumed, to zero order, to be a Gaussian stochastic process. Correspondingly, the probability density function (PDF) of the reverberation envelope is assumed to be Rayleigh and this, indeed, is observed in low-resolution measurements.<sup>1,2</sup> In recent research with high-resolution signals, however, it is found that reverberation envelopes are non-Rayleigh,<sup>3-6</sup> with no universal PDF that fits all data. Generally, sonar parameters (frequency, sonar beamwidth, grazing angle, sonar pulse length, etc.) are found to affect the reverberation PDF.

In microwave backscattering from the ocean surface, radar signals can be represented as a combination of two stochastic processes.<sup>7</sup> One process is continuous, describes contributions from a large number of scatterers within the radar footprint, and is taken as Gaussian, i.e., with a Rayleigh envelope. The other process is discrete, describes contributions from the fewer but stronger scatterers (facets) within the footprint, and can be statistically matched to a  $K$ -distribution with a distribution parameter selected to best fit the data. This model is used to adaptively suppress radar scattering attributed to facets. Recently, it was hypothesized that discrete backscatter also is important in acoustic reverberation associated with very rough bottoms.<sup>8,9</sup> With use of this hypothesis, reverberation observed in experiments was successfully described.

In this article, we analyze reverberation measured at a particular site in deep water west of the Mid-Atlantic Ridge,

and on simple physical mechanisms that may be responsible for its statistical properties. Both monostatic and bistatic reverberation is included. We restrict attention to reverberation obtained with linear FM chirps (LFM) used during the ARSRP-93 experiment.<sup>10,11</sup> In the data considered, the signal auto-correlation time is  $\tau_p \approx 20$  ms, equivalent to a path-wise interval of  $c\tau_p/2 \approx 15$  m for backscatter, and of  $c\tau_p \approx 30$  m for forward scatter, where  $c$  is the sound speed. We characterize the signal as “high resolution” because these path-wise intervals are not large compared to the scales of the bottoms west of the Mid-Atlantic Ridge, which is said to have rms bottom elevations  $\zeta > 100$  m and rms bottom slopes  $\gamma \approx 30^\circ$  (e.g., Refs. 12, 13). The product of acoustic wave number  $k$  with the largest scales of the roughness is  $k\zeta \gg 1$ , so that we also characterize the bottom as “very rough.”

In this paper, we: (a) summarize the data obtained at a particular site; (b) report on their statistical characteristics; (c) identify independent parameters of the experiment that affect the statistical properties of the reverberation; and (d) outline a physically useful approach to understanding reverberation from very rough bottoms. While the ARSRP-93 database is large, entailing diverse bottom conditions, the particular data we use are drawn from a relatively small area. Thus we portray our statistical results as particular to the selected area, but with the belief that the underlying physical ideas can be more generally applicable.

We find that the envelope statistics deviate significantly from Rayleigh. We show that the degree of departure from Rayleigh is dependent upon the vertical grazing angle  $\theta_g$  and the horizontal bistatic angle  $\phi$ . We use a discrete scattering model containing a self-selection mechanism<sup>8,14</sup> as outlined in Refs. 8, 9, 15, together with a continuous scattering model, to explain the statistical properties of the observed scattering. Adoption of the discrete scattering model follows from inspection of temporal data; for data considered here bottom scattering displayed in the time domain consists of well-separated and strong brief peaks ( $\geq 20$  ms separation

<sup>a)</sup>Now at BBN Systems and Technologies, 77 Fawcett St., Cambridge, MA 02138.

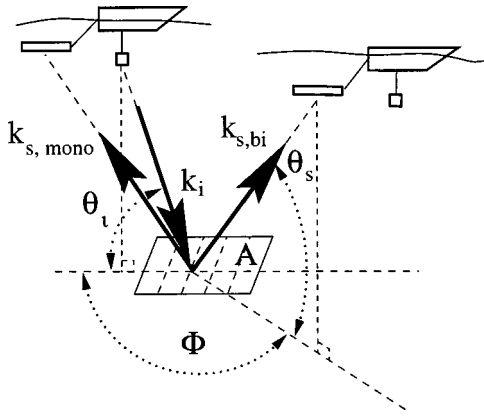


FIG. 1. Basic bistatic scattering experiment using two ships. The actual area  $A$  and shape of the ensonified patch, or footprint, on the bottom, depends on the bistatic geometry, local bathymetry and source and receiver beam patterns. The incident wave vector  $k_i$ , and scattered wave vectors into the two receiver directions  $k_{s,mono}$  and  $k_{s,bi}$  are shown. For simplicity, refraction is ignored in this picture.

and  $\approx 20$  ms brevity). We attribute the highest peaks observed in the data to the discrete self-selection model, while we ascribe the lower levels between the highest peaks to the continuous scattering model. We show that the additive discrete and continuous scattering models are consistent with both the monostatic and bistatic ARSRP reverberation statistics.

## I. DATA ACQUISITION AND PROCESSING

### A. ARSRP experiment

The Acoustic Reverberation Special Research Program (ARSRP) experiments<sup>10,11,16</sup> were carried out in part to explore low grazing angle bottom reverberation. In the following, we analyze and interpret the envelope statistics of such data acquired during the ARSRP-93 cruise.<sup>10,11</sup>

Experimenters using two ships, R/V CORY CHOUET and R/V Alliance, operating in deep water (3–5 km) over the western flank of the Mid-Atlantic Ridge, obtained data on monostatic and bistatic bottom reverberation. We analyze the latter with use of LFM chirps radiated by the CORY CHOUET. The source was a vertical line array (VLA), about  $4\lambda$  long ( $\lambda$  = acoustic wavelength), towed at 180 m depth, with a measured on-axis source level of 232 dB *re*: 1  $\mu$ Pa and 1 m, at 240 Hz, as a representative frequency. For the data analyzed here, the source was steered to the horizontal. The scattered acoustic signals were received and recorded on both the CORY CHOUET and ALLIANCE, thus providing both monostatic and bistatic data. Each receiver was a horizontal line array (HLA), approximately  $50\lambda$  long, towed at about 200 m depth. A conventional 128-beam HLA beamformer with a Hanning window was used on both ships.

The basic two-ship experiment, with use of the CORY CHOUET as the source ship, is diagrammed in Fig. 1. In the experiment, a 5-s LFM chirp (up-sweep from 200 to 255 Hz) was radiated. The equivalent post-processing resolution is  $\tau_p = 20$  ms in time, or  $c\tau_p/\cos\theta_g \approx 30$  m in range-wise horizontal path for forward scatter, and  $\approx 15$  m for backscatter, where  $\theta_g$  is the vertical grazing angle at the bottom.

The HLAs received the chirps scattered from the ocean bottom. Reception time defines both  $\theta_g$  and the horizontal

range  $R$  to the ensonified bottom footprint. In forward scatter, the footprint area is  $c\tau_p/\cos\theta_g$  by  $2R\sin(\psi_e/2)$ , where  $\psi_e$  is the azimuthal resolution of the HLA. At broadside,  $\psi_e \approx 1.1^\circ$ , resulting in a footprint of about  $30 \times 2000$  m at the maximum direct range to the bottom. For backscatter, the range-wise footprint length is  $c\tau_p/2\cos\theta_g$  so that the footprint area at maximum is about  $15 \times 2000$  m. For the more general bistatic cases, the footprint's long dimension can be, at maximum  $R$ , smaller or larger than 2000 m, and is a function of both the HLA steering angle  $\psi$  and the bistatic angle  $\varphi$ .

Temporal displays of the scattering appear discrete in that the matched filter output shows discrete or brief peaks ( $\approx 20$  ms wide) with considerable separation between adjacent peaks (typically  $\geq 20$  ms).<sup>8,9,15</sup> Although the experiment reported here is characterized as high resolution because of the short range-wise extent of the footprint, scattering from bottom features in the footprint's long dimension is potentially more poorly resolved. An interpretation we favor of scattering from such a high aspect-ratio footprint follows. One or more features in a given footprint, with high scattering strength, cause high-amplitude brief peaks. The number of such peaks is likely small, given the large observed separation between them in the range-wise direction. Such an interpretation implies at least approximate horizontal isotropy in the scattering process, evidence for which is presented, at the smaller roughness scales of the bottom, in the next paragraph.

Bathymetry charts where the present reverberation data were acquired are shown in Fig. 2.<sup>13</sup> The bottom is highly lineated, with many major features oriented with their longer dimension parallel to about  $020^\circ$  or  $200^\circ$  (north is the origin at  $000^\circ$ ). Horizontal resolution of the bathymetry data in Fig. 2 is 200 m. While not evident in this figure, features smaller than 200 m are part of the bottom, as various bottom size distribution studies have shown (e.g., Refs. 17–19). At horizontal scales much smaller than 200 m, the bathymetry tends toward isotropy, and loses its strong lineated character. This tendency is shown in Fig. 3, from high-resolution (2-m) bathymetric data<sup>13</sup> over a small area at the experiment site.

### B. AGC processing

It is well known that reverberation in experiments such as reported upon here is a nonstationary stochastic process, because its mean is strongly dependent on sonar properties, on experiment geometry, and on environment parameters.<sup>20,21</sup> Of course, the nonstationarity of the mean can be used to image large-scale features of the bottom (e.g., Refs. 21–23), which is an experimental virtue. But we remove variations in the mean, and thus produce a stationary-mean process, so that the higher order statistics become evident. To do so, we separate the sought-after higher order statistical properties of the scattering data from its mid-time average (MTA), as defined in the next paragraph. The procedure is equivalent to automatic gain control (AGC) processing.

A sample monostatic reception versus time (equivalently, range) is shown in Fig. 4. Its instantaneous value var-



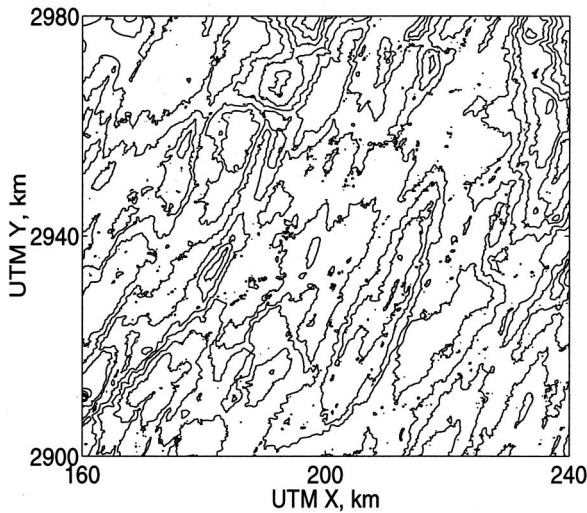
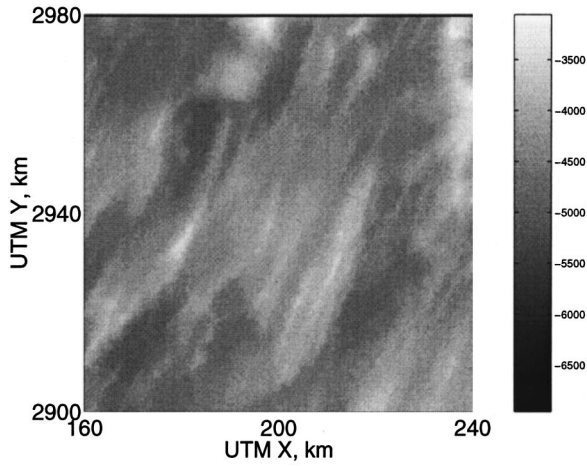


FIG. 2. ARSRP bathymetry charts in the vicinity of site  $B'$ , the latter centered at about UTM  $X=185$  km, UTM  $Y=2915$  km (Refs. 10, 13). The upper chart provides an overall impression of the lineated bottom topography at the site, and the lower chart provides detailed depth contours, at 300-m intervals, with the minimum depth at 3059 m.

ies sharply with time corresponding to the sonar pulse auto-correlation time,  $\tau_p \approx 20$  ms. We see, however, that the local MTA of the received signal, averaged here in a sliding window of width  $T_w = 781$  ms, varies with a longer time scale,  $T_e \geq 10$  s. Therefore, in order to remove the local mean from consideration, while leaving the longer term variations of the received signal intact, we subtract from the time series its MTA in the window  $T_w$ , the latter therefore being bounded by  $\tau_p \ll T_w \ll T_e$ . For all monostatic data averaging, we subsequently used  $T_w = 781$  ms (and as described later for bistatic data averaging, we used three values for  $T_w$ ). A sample of the normalized (AGC) monostatic data is also shown in Fig. 4. For comparison, a realization of noise (reception with no transmission present) is shown as well. In the following, we investigate the differences between the statistical measures of the normalized reverberation and the noise, and compare each to the Rayleigh statistical process.

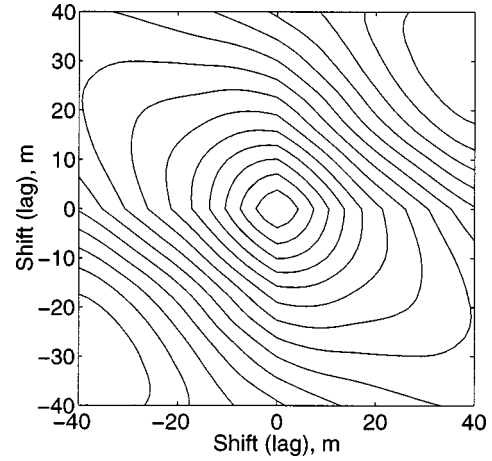


FIG. 3. Maximum correlation contours of roughness heights in a  $100\text{ m} \times 100\text{ m}$  patch centered at about UTM  $X=221$ ,  $Y=3019$  km, from high-resolution bathymetry measurements in the vicinity of site  $B'$  (Ref. 13). Contour intervals are 0.1, with the maximum at 1.0. The horizontal lag radius,  $r$ , has its ordinate in the north direction. At  $r \approx 2$  m, the roughness appears isotropic. At  $r \approx 40$  m, lineation is definite, but not parallel to the macro-scale lineation evident in Fig. 2. (Other  $100\text{ m} \times 100\text{ m}$  patches measured at high resolution, relatively far away ( $\geq 10$  km), also are isotropic at  $r \approx 2$  m, and lineated at  $r \approx 40$  m, but its direction varies from patch-to-patch; Ref. 13).

### C. Statistical measures

The PDF is commonly used to characterize stochastic signals. It is known that low probability, large-amplitude tails in the PDF can be very important in sonar performance (e.g., Refs. 4, 6) and, to focus on such tails, other measures are also employed. Common ones are probability of false alarm (PFA) and separation between adjacent peaks exceeding a threshold (which is directly related to PFA at the threshold). For simplicity and its sensitivity to the PDF tails, peak separations are used here, along with PDFs, to probe the overall physical message contained in the data.

## II. STATISTICAL PROPERTIES OF THE DATA

All received pings were match-filtered, AGC-processed, absolute-valued, dB-transformed, and then analyzed statisti-

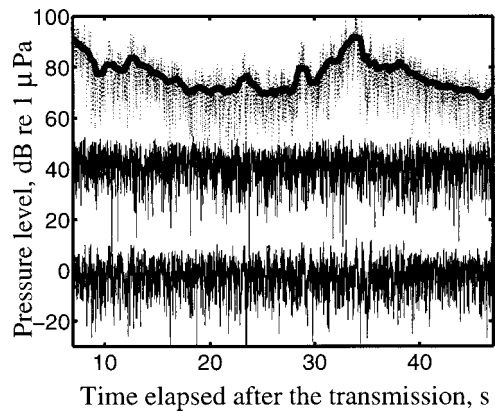


FIG. 4. Top curve (dot line): signal received from Ping 412 at the CORY CHOUET, on Julian day 197 (J197) starting at 05:36 Z; (solid line) 781-ms sliding average (in the logarithmic domain) of the received signal. Middle curve: noise received in the band 200–255 Hz, delayed to the Ping 413 time-slot, so as to avoid ping reverberation; the noise is shifted 12 min and plotted on the same time axis as the signal. Bottom curve: received signal with its 781 ms sliding average subtracted.



cally. ARSRP-93 Pings 411, 412, and 414 were used for monostatic data analyses. Ambient noise was obtained from Ping 413, with no transmission present. And seven nonconsecutive Pings (430, 436, 469, 480, 486, 492 and 499) were used for bistatic data analyses.

Our objective is to estimate relevant statistical properties of reverberation, at an acceptable level of confidence, for the monostatic and the bistatic case. To achieve confidence, a large number of data points, for all scattering levels of interest, are required. Because high-level events have low probability, large numbers are difficult to obtain. Indeed, with the small range resolution afforded by the signal used, and with the narrow horizontal resolution of the HLA, too few data points at high levels tend to be observed. Instead we adopt various approaches to increase the number of points or otherwise to reach acceptable levels of statistical confidence. We describe these approaches subsequently for each form of data analysis entered into.

Increasing the number of data points increases the sample space, and thereby seems counter to the high-resolution capabilities of the measurement system. The conflict between a confident statistical result, and a highly specific experimentally observed attribute, has no easy remedy. We take comfort, however, that the compromises we later describe to resolve the conflict do not totally thwart basic understanding of reverberation at this site.

In measuring statistical properties of a random signal such as reverberation, one aggregates all relevant observations to form its statistical ensemble. Reverberation depends both on experiment geometry and on bottom properties. For monostatic geometry, we take the ensemble to encompass all data in the approximate range interval of  $2.5 \leq R \leq 30$  km, over the HLA main lobe steering angle interval of  $0 \leq \Psi \leq 180^\circ$ , that is, the ensemble entails a bottom area of about  $2800 \text{ km}^2$ . All pings used for the monostatic data essentially pertain to one highly overlapped site of area about  $3000 \text{ km}^2$ . For bistatic geometry, data were obtained from the same site, but the ensemble of samples encompassed an area of about  $600 \text{ km}^2$ . With respect to the latter, we hypothesize that the samples in this smaller area are a spatially ergodic subset of the ensemble over the larger area, so that bistatic reverberation may be reasonably compared with monostatic. That is, the data are hypothesized to form a spatially ergodic process; we show this to be in agreement with the data, at least for the various data analysis procedures used in the paper.

### A. Monostatic spatial ergodicity

Monostatic reverberation PDFs from Ping 412, resulting from substantially different geometrical conditions and bottom footprint locations, are compared in Fig. 5. To obtain the requisite number of data points for the 95% confidence intervals shown, two steps were taken. First, we used contiguous beams to produce HLA sensing cones, each with  $20^\circ$  azimuthal coverage (see Fig. 6). One cone was steered to the two directions at  $90^\circ$ , and the other to the two directions at  $40^\circ$ , both measured from the HLA velocity vector. Second, we aggregated the data into two 10-s groups: (1) data designated as “early arrivals” were recorded 7–17 s after the ping was radiated, during which reverberation was obtained

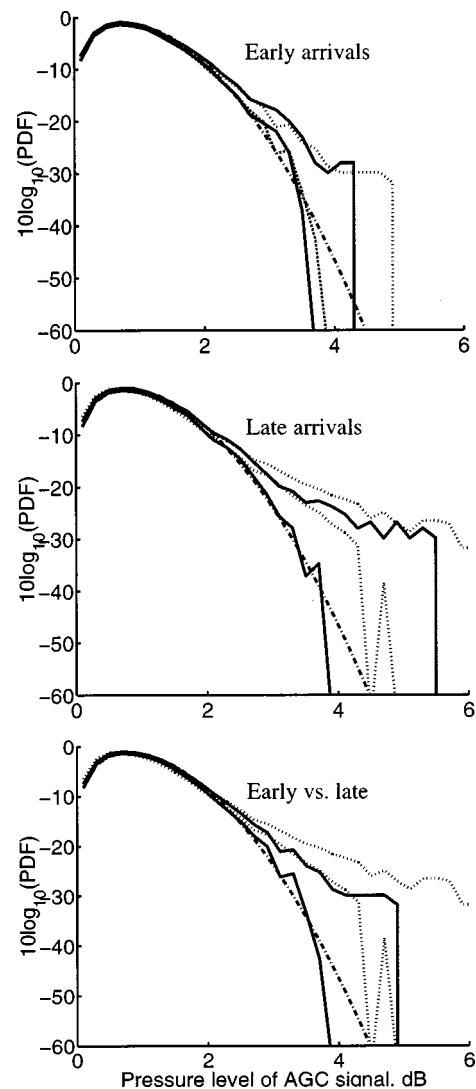


FIG. 5. Reverberation PDFs for Ping 412, sampled in the conical HLA beams shown in Fig. 6. Conical beams centered at  $90^\circ$  (dot line), and at  $40^\circ$  (solid line), in which the two lines shown for each delimit the 95% confidence interval of the corresponding estimate. (Sharp termination of the PDFs at a high level is an artifact of the data analysis procedure.) Top graph shows early arrivals ( $\theta_g \approx 40^\circ$ ), middle graph late arrivals ( $\theta_g \approx 5^\circ$ ), and bottom graph early/late comparison for the  $90^\circ$  conical beam. In all graphs, the Rayleigh PDF is the dash-dot line.

at short ranges ( $2.5 < R < 10$  km) and at relatively large bottom grazing angles ( $\theta_g \approx 60^\circ$  to  $20^\circ$ , stated for brevity as  $\theta_g \approx 40^\circ$ ); and (2) data designated as “late arrivals” were recorded from 30 to 40 s during which scattering originated at long ranges ( $20 < R < 30$  km) and at small grazing angles ( $\theta_g \approx 8^\circ$  to  $1^\circ$ , for brevity  $\theta_g \approx 5^\circ$ ).<sup>24</sup> We label each of these groups by its nominal arrival time or nominal  $\theta_g$ , and analyze these data separately in each sensing cone. The resulting sample space on the bottom for each cone is  $O(10^4)$  larger than that intrinsic to the base measurement system.

At the time of Ping 412, the HLA heading was about  $160^\circ$  true, and was located at about UTM  $Y=2954$  and  $X=219$  km. The axes of the HLA sensing cones thus intersect the bottom at paired directions of  $070^\circ$  and  $250^\circ$ , and of  $120^\circ$  and  $200^\circ$ , respectively. It can be concluded from Fig. 2 that these intersections cross the major bottom lineational direction at significantly different angles. Also, the left/right foot-

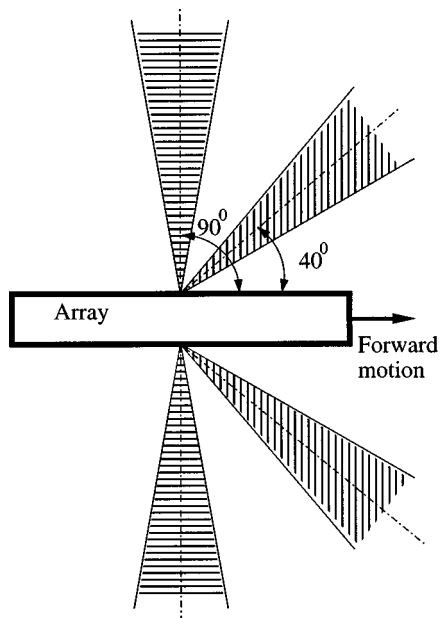


FIG. 6. HLA conical beams for the Fig. 5 data, each with a  $20^\circ$  included angle.

prints of each paired beam direction encompass significantly different bottom contours.

At high levels, each data set shown in Fig. 5 departs from the Rayleigh PDF, an observation that we will later revisit. For now, note that the PDF for each arrival group, in the  $90^\circ$  and  $40^\circ$  beam pairs, is closely the same. But there are differences. Also, there are differences in the PDFs between early and late arrival time, as shown for the  $90^\circ$  cones (but there are smaller if not negligible differences for the  $40^\circ$  cones). Thus it is reasonable to hypothesize from this figure that the data at this site could belong to a single statistical ensemble, in which the main parameter affecting its statistical character would be  $\theta_g$ . Because Fig. 5 is not totally convincing in this respect, a larger sample size is required to test the hypothesis. Correspondingly, if the hypothesis is accepted, such a larger sample size would be the basis of studies in which spatial ergodic behavior is a desired assumption or attribute. This is addressed next.

The PDFs of reverberation from Ping 412, given in Fig. 5 for the two-sided  $20^\circ$  HLA sample space, are now shown in Fig. 7 for the full  $180^\circ$  space. The resulting sample space

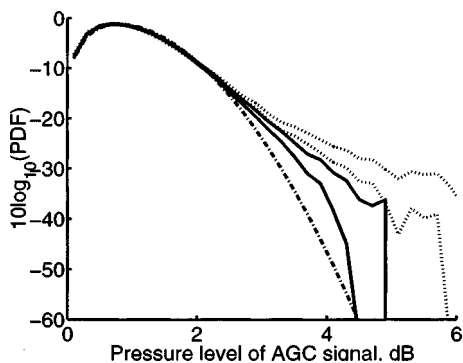


FIG. 7. Reverberation PDFs for Ping 412, with full  $180^\circ$  HLA conical beam sampling. Early arrivals ( $\theta_g \approx 40^\circ$ , solid) and late arrivals ( $\theta_g \approx 5^\circ$ , dot). Dash-dot: Rayleigh PDF (dash-dot).

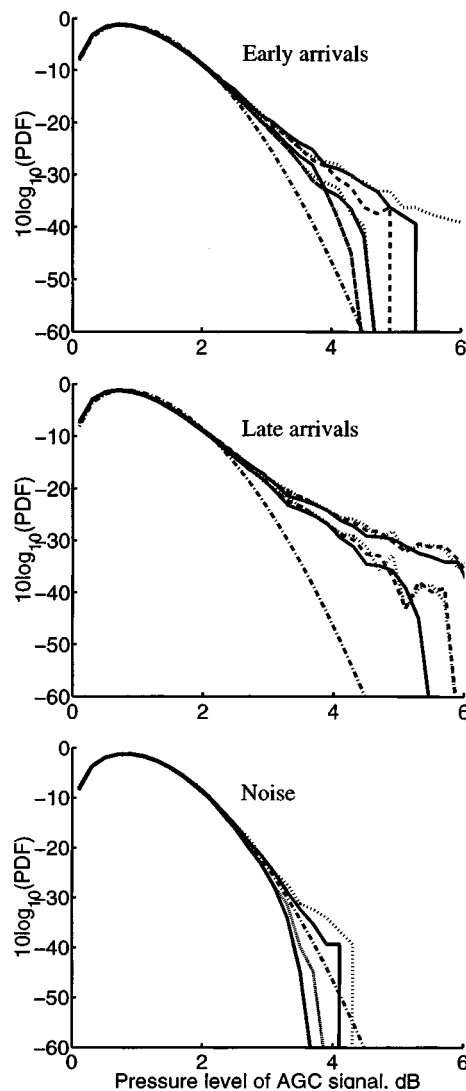


FIG. 8. PDFs for reverberation (top and middle graphs) and ambient noise (bottom graph), with full  $180^\circ$  HLA sampling. Ping 411 (solid), Ping 412 (dash), Ping 414 (dot), Rayleigh PDF (dash-dot). Top shows early arrivals ( $\theta_g \approx 40^\circ$ ), middle late arrivals ( $\theta_g \approx 5^\circ$ ), and bottom shows noise, measured in Ping 413 during two arbitrary well-separated 10-s intervals.

on the bottom is thereby larger by  $O(10^5)$  than that intrinsic to the base measurement system. This decreases the variance of the estimates, which is desired, in part, to check the parametric importance of  $\theta_g$ . As stated previously, early and late arrivals are separately analyzed,<sup>24</sup> and compared in Fig. 7. We conclude from Fig. 7 that monostatic reverberation statistics at high levels are unequivocally parametric in  $\theta_g$ .

To proceed to the spatial ergodicity question we first show, in Fig. 8, reverberation PDFs obtained separately from Pings 411, 412, and 414. Two-sided  $180^\circ$  HLA sample space, and separation in  $\theta_g$ , is used here too. Within each  $\theta_g$  data set, no significant difference is found in Fig. 8 among the reverberation PDFs for the three pings. The ship moved only about 4 km between Pings 411 and 414, thus entailing considerable overlap in bottom sample space, so that the similarity among the PDFs mostly demonstrates data reduction robustness. (The upper two graphs in Fig. 8 point, once again, to the parametric role of  $\theta_g$  in the statistics.)

Next, note that the PDF for each  $\theta_g$  has a spread in Fig.

8 substantially contained by those in Fig. 5. Stated differently, the differences pointed to earlier between the PDFs in Fig. 5 are fluctuations subsumed by the  $O(10)$  increase in the data sample space of Fig. 8. This means that each of the  $20^\circ$  sets is approximately acceptable as a sample of the  $180^\circ$  HLA data set. Further, we may conclude that spatial ergodicity is approximately achieved for a bottom area (calculated from one  $20^\circ$  HLA beam pair) that is as small as about  $35 \text{ km}^2$  for  $\theta_g \approx 40^\circ$ , or as small as about  $170 \text{ km}^2$  for  $\theta_g \approx 5^\circ$ .

The data we have analyzed were collected at just one site. Hence, the foregoing result should not be taken to be universal, or that spatial ergodicity would necessarily behave similarly at other sites in the ARSRP database.

## B. Monostatic statistics

Comparison with statistical results on noise, given in this section, puts reverberation statistics in a broader physical context. Both PDF and peak-separation analyses are also discussed, the main thrust of which is to gain understanding of reverberation from use of both tools.

Ambient noise PDFs from Ping 413 are given in Fig. 8. As expected, the noise is unequivocally Rayleigh.

Central to our thesis, we observe from Figs. 5, 7, and 8 that there are statistically significant departures of the reverberation PDFs from Rayleigh. Specifically, compared to Rayleigh, reverberation has slightly smaller probability near its mean (hard to see in these figures), but significantly larger probability at high levels (easy to see).

To investigate these results from the perspective of adjacent peak separation, we set a 2-dB threshold, a value 1 dB higher than the most probable value of the PDFs. At and somewhat above this threshold, the measured PDFs cannot readily be differentiated from Rayleigh, but the peak separation data can do so easily. For this threshold, the average measured temporal separation between adjacent peaks, as a function of each HLA beam, is shown in Fig. 9. HLA beam aggregations are thus not used but, as previously, range bins to separate  $\theta_g$  are used. For comparison, ambient noise was processed in the same way as the scattering, and is also shown. We see that the average separation between adjacent peaks for reverberation is significantly different than for noise. Since statistics of the ambient noise envelope is Rayleigh, the statistics of the reverberation envelope must therefore deviate from Rayleigh, a conclusion consistent with the PDF figures.

## C. Interpretation of monostatic statistics

The observed non-Rayleigh statistics of reverberation can be explained with use of the discrete scattering hypothesis. Large amplitudes can be found both in reverberation and noise. However, since noise is a Rayleigh process, its large amplitudes are generated when many small noise events contribute constructively. For noise, therefore, the probability of exceeding a large threshold is very small; it decays exponentially as the level above the threshold increases. For reverberation, as for noise, one can also get strong signals due to constructive addition from many weak

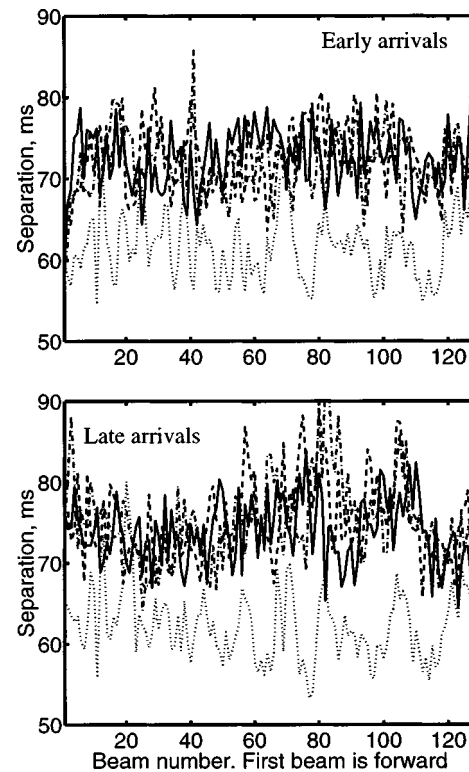


FIG. 9. Average temporal separation between adjacent peaks, above a 2-dB threshold, as a function of beam number. The HLA back direction is number 0, the forward direction number 127. Ping 411 (solid), Ping 412 (dash), and Ping 414 (dash-dot). Noise from Ping 413 (dot). Upper plot:  $\theta_g \approx 40^\circ$ . Lower plot:  $\theta_g \approx 5^\circ$ .

scatterers in the sonar footprint; the envelope statistics of these are also Rayleigh. With use of the discrete scattering idea, and with the assumption that the footprint contains a few strong scatterers, as well as many weak ones, high amplitudes in the reverberation can be caused by strong scattering from some features, with envelope statistics decidedly non-Rayleigh. The latter statistics reflect the distribution of these scatterer sizes on the bottom, and their corresponding target strengths, as affected by size, orientation, and acoustic wavelength.

The distribution of heights on the ocean bottom is known to be governed by the Goff-Jordan power law (e.g., Ref. 17). Large scatterers are less likely than small ones, expressed by a negative power law relationship between probability and scatterer size. Since the power law decays more slowly than the exponential, one expects that reverberation will have more high-level events, and smaller peak separation, than Rayleigh (i.e., noise). At first glance this is opposed to what the data in Fig. 9 show. But this figure is dominated by separation statistics at and slightly above the 2-dB threshold, and not by levels  $>3$  dB, for which the departure from Rayleigh becomes significant (see Figs. 5, 7, and 8). That is, since the levels not too far above the threshold compensate for the larger number at higher levels, the separation statistics of Fig. 9 are the reverse of that which would be true at high thresholds.

One would suppose that reverberation statistics of high-level events depend on area  $A$  of the bottom footprint because, for a constant scatterer density, the total number of

strong discrete scatterers  $N_s$  is proportional to  $A$ . In a given beam away from vertical,  $A$  varies with  $\theta_g$ . From Eq. (4) of Sec. III,  $A$  is about two times larger at  $\theta_g \approx 5^\circ$ , than at  $40^\circ$ . A difference this large for the two grazing angles contradicts Fig. 9, in which the peak separations for the two, and hence the implied areas, are roughly equal. This apparent contradiction is explained, once again, by the 2-dB threshold. The data in Fig. 9 are dominated by peaks at (and only somewhat above) 2 dB, corresponding to a large number of weaker scatterers in the footprint, both for smaller and larger  $A$  (observed at  $\theta_g \approx 40^\circ$  and  $\approx 5^\circ$ , respectively).

Another apparent difficulty in interpreting Fig. 9 is this: the beamwidth from Sec. III, and therefore  $A$  (with  $\theta_g$  fixed), theoretically increases as  $\csc(\psi)$  as the HLA is steered away from broadside, where  $\psi$  is the main lobe steering angle. Although some variation with  $\psi$  can be observed in Fig. 9, especially for  $\theta_g \approx 5^\circ$ , the expected  $\csc(\psi)$  variation is not observed. Once again, the separation statistics are dominated by peaks at (and only somewhat above) 2 dB. These obscure the few but strong discrete scatterers at levels  $>3$  dB, and thus the expected dependence  $A \sim \csc(\psi)$  is also obscured.

#### D. Bistatic statistics

For the bistatic data, too few data points are available at a given  $\phi$  in a single ping. Therefore, data acquired during seven nonconsecutive pings are used. Still, we did not obtain enough points to generate a good estimate of the bistatic PDF at the tails. Because of this, we use only peak separations to analyze bistatic statistical properties.

For bistatic data, adjacent points in the time series have no monotonic connection with scattering from a specific bottom footprint, or with scattering geometry (incidence and scattering grazing angles  $\theta_i$  and  $\theta_s$ , and bistatic angle  $\phi$ ). Therefore, via ray tracing, we relate the time and beam number of the data to each footprint, and calculate the scattering geometry,  $\theta_i$ ,  $\theta_s$ , and  $\phi$ . Then data are grouped from various parts of the time series, and data subsets uniform in scattering geometry are analyzed separately. Implicit in this grouping is the assumptions of spatial ergodicity and horizontal isotropy, the latter at least for the roughness scales of importance in bistatic scattering. These caveats need to be attached to the bistatic data reduction approach for later discussion.

Because the bistatic data subsets can still correspond to very different bottom locations within the particular site, we submerge (by temporal averaging) shorter scales of the process. This is a different form of the AGC processing that was used for the monostatic data. In absence of *a priori* knowledge of the appropriate length of the sliding time average, we used three windows ( $T_w = 500, 213,$  and  $107$  ms) to explore averaging effects. Additionally, since partitioning of each ping into data bins left many with too few points, we aggregated data acquired during the seven pings listed earlier. For these pings, both ships followed tracks around a relatively small region, approximately centered on the same site used for the monostatic data. Footprints of pings in this region were chosen within a bottom area of about  $600 \text{ km}^2$ , located near a strong deterministic bottom feature, known to ARSRP

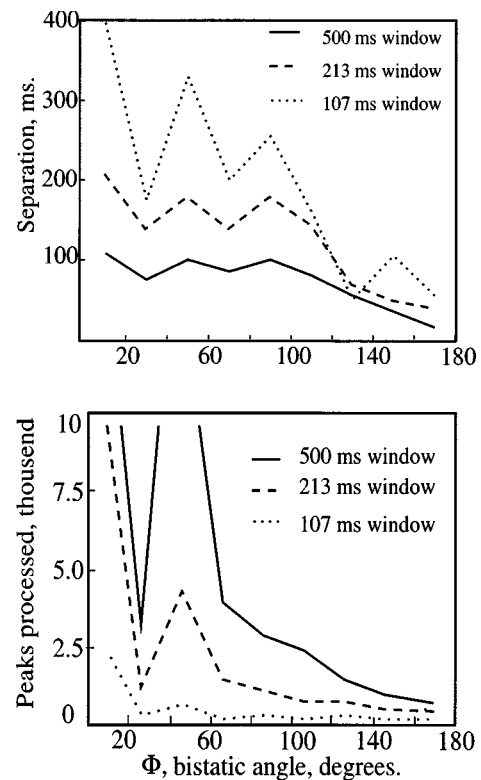


FIG. 10. Upper plot: average separation between directly adjacent peaks as a function of bistatic angle  $\phi$ , above a 2.5-dB threshold. Lower plot: number of peaks observed in the upper plot.  $\phi = 0^\circ$  corresponds to backscattering,  $\phi = 180^\circ$  to forward-scattering. Solid line is the signal normalized with a 500-ms sliding average, dash line with a 213-ms sliding average, and dot line with a 107-ms sliding average. (Data for low bottom grazing angles; see text for details.)

as  $B'.$ <sup>10,13</sup>  $B'$  was detected in each of the seven pings, which helped to group like data, but the analyzed footprints explicitly excluded  $B'$  to avoid its strong bias. Footprints adjacent to  $B'$ , however, were included. More details on the data partitioning can be found in Ref. 15.

In grouping points, we chose bins for  $\theta_i$ ,  $\theta_s$ , and  $\phi$ . Specifically, we processed only low grazing angle data ( $\theta_i$  and  $\theta_s$  from about  $2^\circ$  to  $10^\circ$ ), but for all bistatic angles ( $\phi$  in  $20^\circ$  bins). A value of 2.5 dB was chosen as the peak detection threshold. The mean interpeak separation, as a function of  $\phi$ , is plotted for the three sliding windows in Fig. 10. The observed peak separation trends, for at least the two larger windows, are roughly constant for  $\phi < 100^\circ$ , and decreasing with increasing  $\phi$ , for  $\phi > 100^\circ$ . The separations rapidly decrease and reach their minimum at the forward scattering direction.

As a rough gauge of statistical validity, the number of peaks detected is also shown in Fig. 10. The number is fairly large for the 500-ms window, but considerably smaller for the 107-ms window. The smaller the number of detected peaks, the larger the apparent statistical uncertainty. Normalized by the trends for  $\phi < 100^\circ$ , the fluctuations in mean separation from Fig. 10 are roughly 10%, 20%, and 40%, respectively, for the 500-, 213-, and 107-ms windows.

With respect to spatial ergodicity, and to confident estimates of bistatic PDFs, we would need data at each  $\phi$  from more than about  $4 \times 10^3$  footprints. This value follows from



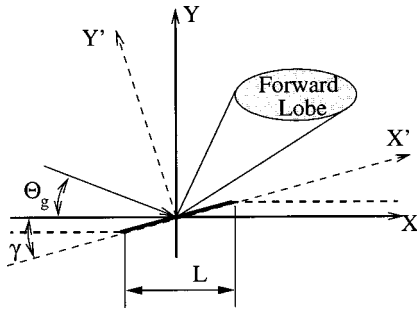


FIG. 11. Coordinates for scattering from a facet rotated by the angle  $\gamma$  from horizontal.  $X'Y'$  is the local coordinate system aligned with the facet. The observation direction is defined by angle  $\beta$  from forward, where  $\theta_s = \beta$ ,  $0 < \beta < \pi/2$ , and  $\theta_s = \pi - \beta$ ,  $\pi/2 < \beta < \pi$ .

170 km<sup>2</sup> (determined in Sec. II A for the monostatic data) as the minimum required area, and from  $4 \times 10^{-2}$  km<sup>2</sup> as an estimate of the mean bistatic footprint area, both for small  $\theta_g$ . With a probability  $< 0.03$  for peaks  $\geq 2.5$  dB (from Fig. 8, say), the number of detected peaks needed in each  $\phi$  bin is  $\geq O(10^2)$ . This condition is not met by the 107-ms data in Fig. 10. But the 213- and 500-ms data sets do; they are thus consistent with the ergodicity and statistical confidence obtained from the monostatic data.

### E. Interpretation of bistatic statistics, and dominant length scales for bistatic and monostatic scattering

Consider scatterers (facets) of size  $L$  in the footprint. On average, the bottom can be taken as horizontal, but because of roughness, the local slopes are nonzero and, therefore, each facet is rotated from the horizontal by an angle  $\gamma$ . One can expect strong scattering from a facet within its forward lobe, i.e., within an angular spread of  $\lambda/L$  from the local specular direction. (Direction refers to the local coordinate system  $X'Y'$  aligned with the facet, i.e., rotated to the angle  $\gamma$ , as shown in Fig. 11.) For a facet to contribute to scattering in the direction  $\beta$  from forward, it has to be rotated by the angle

$$\frac{\beta - \vartheta_i - \lambda/(2L)}{2} \leq \gamma \leq \frac{\beta - \vartheta_i + \lambda/(2L)}{2}. \quad (1)$$

Since the average slope  $\langle \gamma \rangle = 0$ , for most facets the local specular direction is close to the specular direction measured from the horizontal (in the  $XY$  coordinates). Therefore, most facets scatter into the specular direction, and only small ones (i.e., those with a wide forward lobe), or facets rotated to a relatively large angle given by Eq. (1) and thus improbable, can scatter away from the specular direction.

By taking into account the dependence of scattering cross section on facet size, and with the foregoing reasoning, a self-selection mechanism has previously been hypothesized, such that facets of size  $L \approx \lambda$  dominate backscatter.<sup>8,9</sup> Smaller facets are incapable of generating appreciable backscatter, and larger facets with narrow specular lobes are very unlikely to be rotated enough to contribute to backscatter. While the hypothesis was framed originally for backscatter, the preceding paragraph generalizes it to include the bistatic case.

From the present data, the statistical properties of bistatic scattering change at  $\phi \approx 100^\circ$ . For  $\phi > 100^\circ$ , where peak separations become distinctly smaller with increasing  $\phi$ , a continual change in envelope statistics is implied. Thus  $\phi \approx 100^\circ$  defines a half-width angular sector  $\Delta\phi \approx 80^\circ$ , in which there is an increasing merging or overlay of peaks, with attendant smaller interpeak separation that reaches a minimum as  $\phi \rightarrow 180^\circ$ . This is consistent with the self-selection idea from which, with use of  $\sin \Delta\phi = \lambda/L$ , we can say that the dominant scatterer roughness scales in this domain are given by  $L \geq O(\lambda)$ .

On the other hand, for  $0^\circ < \phi < 100^\circ$ , the approximately constant separation observed vs  $\phi$  means that this is a domain of constant statistics, in which smaller scales dominate scattering, specifically  $L \leq O(\lambda)$ . To match the self-selection condition at backscatter ( $\phi \rightarrow 0^\circ$ ) as in Refs. 8, 9, 15, we adopt the more stringent statement  $L \approx \lambda$  as an appropriate descriptor for this domain. By extension, such  $\lambda$ -scale scattering must be present at all  $\phi$ , but where  $\phi > 100^\circ$ , contributions from larger (more directive) scatterers, constrained to their forward lobes, are more important. Of course, the forward lobe width depends on  $\lambda/L$ : for  $L = \lambda$ , the forward lobe occupies  $\phi > 100^\circ$ , for  $L = 2\lambda$ ,  $\phi > 125^\circ$ , etc. Thus when large bistatic angles are considered, scattering from the more directive features adds to and dominates that from  $\lambda$ -scale features, and the various scales in the combined scattering for  $\phi > 100^\circ$  results in statistical properties that depend on  $\phi$ . For  $\phi < 100^\circ$  they do not.

In the present data set  $\lambda \approx 7$  m. For the range  $\phi < 100^\circ$  (including backscatter), the dominant scattering roughness scale is then  $L \approx 7$  m which, for large local slope  $\gamma$ , has a corresponding horizontal scale about  $r \leq 7$  m. With the guidance of Fig. 3, we can therefore conclude that the process is approximately horizontally isotropic. On the other hand, for  $\phi > 100^\circ$ , the dominant horizontal scale can be  $r > 7$  m. Thus in the forward scatter sector, the process tends toward increasing anisotropy as  $\phi \rightarrow 180^\circ$ , and thereby recedes from the simplicity of horizontal isotropy. Consequently, aggregation of the bistatic data as done by us, which assumes isotropy, is strictly appropriate only for the range  $\phi < 100^\circ$ . Also, the 2.5-dB threshold used for the bistatic peak separation analysis no doubt admits a higher number of smaller peaks than in retrospect we would have liked, and thus blunts the sharpness of these conclusions.

### III. A MODEL FOR DISCRETE BACKSCATTERING

With use of the conclusions drawn in the foregoing, we calculate the functional form of the number of discrete scatterers contributing to backscattering. While the bottom topography in the area of our acoustic observations is known to be highly lineated (see Fig. 2), at the  $\lambda$ -scale appropriate to backscatter the roughness is taken as horizontally isotropic (see Refs. 15, 18, 19, and Fig. 3). In this context, we neglect shadowing, take the remaining complexity of the rough bottom to be the local bottom slope  $\gamma$ , and assume it to follow a Gaussian distribution, consistent with classical rough surface statistics (e.g., Ref. 25)

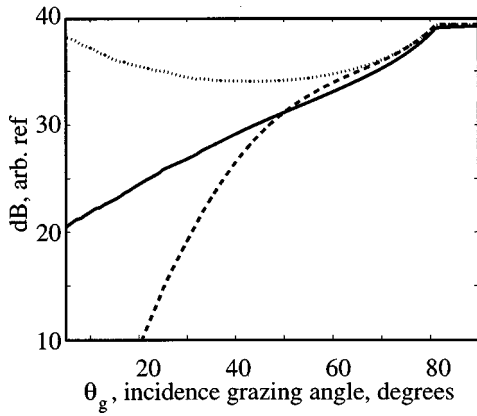


FIG. 12. Effective number (in  $10 \log_{10} N_{s,e}$ ) of backscattering facets of size  $L=\lambda$ , for unit facet density ( $n=1 \text{ m}^{-2}$ ), and for other conditions given in the text. Solid line:  $\gamma_{\text{rms}}=30^\circ$ . Dash line:  $\gamma_{\text{rms}}=15^\circ$ . Dot line: footprint area  $A$ .

$$P(\gamma) = \frac{1}{\sqrt{2\pi} \cdot \gamma_{\text{rms}}} \cdot \exp\left(-\frac{\gamma^2}{2\gamma_{\text{rms}}^2}\right) \quad (2)$$

in which  $P$  is the PDF of  $\gamma$ , and  $\gamma_{\text{rms}}$  is the rms slope. Then, the probability of a contribution from a facet to backscattering depends on its forward lobe width  $\alpha=\lambda/L$ , and on incidence grazing angle:<sup>26</sup>

$$P_1 = \int_{\pi/2-\theta_g-\alpha/4}^{\pi/2} P(\gamma) d\gamma. \quad (3)$$

For a macroscopically flat horizontal bottom and for  $\theta_g$  less than  $90^\circ$ , the footprint area can be stated as

$$A = \eta c(d) \tau_p S[\theta_g, \theta_0, c(z), d] / 2 \cos \theta_g \quad (4)$$

in which  $\eta=(\lambda/\Lambda)/\sin \psi$  is the HLA beamwidth (except at endfire),  $\Lambda$  the HLA length,  $\psi$  the HLA steering angle ( $\psi=90^\circ$  at broadside),  $c(z)$  the sound speed profile in depth  $z$ ,  $d$  the local ocean depth, and  $S$  the refracted pathlength which is a function of the path's ray angle  $\theta_0$  at the source (as well as of other previously defined quantities). Equation (4) was evaluated numerically for  $\lambda/\Lambda=0.02$ ,  $\psi=90^\circ$ ,  $c(z)$  as measured at the site,  $\tau_p=20$  ms,  $\theta_0=0$ ,  $d$  as obtained from ARSRP bathymetric data, and  $\theta_g$  from Snell's Law. Then, the number of facets contributing to backscatter (the effective number of facets), as a function of  $\theta_g$ , is given by

$$N_{s,e} = nAP_1, \quad (5)$$

where  $n$  is the areal density for a given scatterer size.

Results for scatterer size  $L=\lambda$ , and for  $\gamma_{\text{rms}}=15^\circ$  and  $30^\circ$ , are shown in Fig. 12. Because most facets are probabilistically steered away from the backscatter direction at small grazing angles [see Eqs. (2) and (3)], their number effectively decreases as  $\theta_g \rightarrow 0$ . Thus the model is in agreement with the larger departure from the Rayleigh PDF at smaller grazing angles, as shown in the data of Figs. 5, 7, and 8.

For  $\gamma_{\text{rms}}=30^\circ$ , a value we derived from ARSRP topographic data at 5-m horizontal resolution,<sup>23</sup>  $N_{s,e}(5^\circ)$  is predicted to be 7 dB less than  $N_{s,e}(40^\circ)$ . This is a significant change. But absolute values of the levels comprising the den-

sities shown in Figs. 5, 7, and 8 were not retained (and anyway, for  $\theta_g \approx 40^\circ$ , not robustly measured). Thus they are not used to additionally test the model.

Finally, for  $\gamma_{\text{rms}}=15^\circ$ , i.e., for a surface smoother than in ARSRP, smaller values of  $N_{s,e}$  are calculated for  $\theta_g < 50^\circ$  (see Fig. 12). Therefore, at a fixed  $\theta_g$ , smoother surfaces are expected to generate larger departures from Rayleigh envelope statistics, compared to the departures we have observed for  $\gamma_{\text{rms}} \approx 30^\circ$ . Such a smoothness effect is, of course, conditional; there must be at least a few  $\lambda$ -scale scatterers statistically present in the footprint. This effect gains support from measurements of radar backscatter from the sea surface;<sup>27</sup> these data have larger departures from the Rayleigh PDF compared to that for the ARSRP data, the sea surface  $\gamma_{\text{rms}}$  for radar being substantially smaller than  $30^\circ$ .

## IV. CONCLUSIONS

Reverberation envelopes for a high-resolution signal, and for the very rough bottom at a particular site west of the Mid-Atlantic Ridge, have non-Rayleigh probability densities. The departure from Rayleigh depends upon the bottom grazing angle  $\theta_g$ , and the bistatic angle  $\phi$ . The smaller the value of  $\theta_g$ , the larger the departure, corresponding to a larger probability of high-amplitude scattering (relative to the local mean). Peak separations on average in a monostatic or bistatic time-series are, in general, a function of a selected threshold, and of  $\theta_g$ , and can be useful along with probability densities, in understanding statistical processes affecting reverberation.

For monostatic reverberation, virtually the same peak separations were observed for small and large  $\theta_g$ , in which the peak threshold was set at about 1 dB above the amplitude's most probable value. But one could expect that for a threshold set much higher, say  $>3$  dB above the most probable value, the separations would be smaller for smaller  $\theta_g$ , as would follow from the measured probability densities.

Peak separations on average for bistatic reverberation is approximately constant for  $\phi < 100^\circ$  (including the back direction). In contrast, the average separation decreases monotonically for  $\phi > 100^\circ$  (toward the forward direction). From these observations, the bistatic envelope probability density is inferred to be constant for  $\phi < 100^\circ$  and variable for  $\phi > 100^\circ$ , neither of which would be Rayleigh.

These rough bottom reverberation observations can be explained, arguably, by adopting a continuous scattering model having a Rayleigh envelope, added to a discrete and self-selecting scattering model (arising from a small number of individual features within the sonar footprint) having a distinctly non-Rayleigh envelope. The latter self-selects a scatterer size  $L \approx \lambda$ , which dominates scattering within  $\phi < 100^\circ$ , including backscatter. It also self-selects scatterer sizes  $L \geq O(\lambda)$ , which dominate scattering within the remaining bistatic domain,  $\phi > 100^\circ$ , including forward scatter. Based on available bathymetric data, and on the wavelength used in the self-selection model, non-Rayleigh components of the reverberation at this site are likely dominated by horizontally isotropic roughness, both for monostatic and bistatic observations, provided the latter has bistatic angles within  $\phi < 100^\circ$ .

A large number of data points are required at this site to obtain statistically confident results for low-probability high-amplitude events. Monostatic reverberation from bottom areas, of about  $170 \text{ km}^2$  for  $\theta_g \approx 5^\circ$  and about  $35 \text{ km}^2$  for  $\theta_g \approx 40^\circ$ , and measured in paired beams of a line-array cone, had probability densities close to those measured from the bottom area over the full site (nine times larger). While no systematic effort was undertaken to sharpen this result, it is apparent that bottom variability in one cone pair encompassing these areas is virtually sufficient to statistically represent the full site.

For grazing angles  $\theta_g < 60^\circ$ , backscatter calculations neglecting shadowing show that the number of scatterers in a rough bottom footprint is effectively much less than that proportional to the footprint's area. Local bottom slope  $\gamma$  causes the reduction; it steers bottom scattering away from the nominal backscatter direction. These calculations show that the effective number of scatterers on average is sensitive to the rms slope. For example, a decrease of  $\gamma_{\text{rms}}$  from  $30^\circ$  to  $15^\circ$  reduces the backscatter at  $\theta_g = 20^\circ$  by 15 dB, an effect likely more pronounced than that of the neglected shadowing. But although smoother because of its smaller  $\gamma_{\text{rms}}$ , one can expect the monostatic reverberation probability density to have larger departures from Rayleigh (most likely with a smaller mean than that observed for rougher surfaces).

## ACKNOWLEDGMENTS

Data analyses were facilitated by the efforts of Dr. Joseph Bondaryk, then of MIT, and Edward Scheer of WHOI. Professors Arthur Baggeroer and Henrik Schmidt led MIT's overall ARSRP research. The considerable help given by all four is gratefully acknowledged. We also thank Drs. Marshall Orr, Mohsen Baidey, and Mr. Jeffrey Simmen, ONR, for their knowledgeable leadership of the overall ARSRP effort.

- <sup>1</sup>P. Faure, "Theoretical model of reverberation noise," *J. Acoust. Soc. Am.* **36**, 256–259 (1964).
- <sup>2</sup>D. Middleton, "A statistical theory of reverberation and similar first-order scatter yields. Part I: Waveforms and general processes," *IEEE Trans. Inf. Theory* **IT-13**, 372–392 (1967).
- <sup>3</sup>G. R. Wilson and D. R. Powell, "Probability estimates of surface and bottom reverberation," *J. Acoust. Soc. Am.* **73**, 195–200 (1983).
- <sup>4</sup>N. P. Chotiros, H. Boehme, T. G. Goldsberry, S. P. Pitt, R. A. Lamb, A. L. Garsia, and R. A. Altenburg, "Acoustic backscattering at low grazing angles from the ocean bottom. Part II. Statistical characteristics of bottom backscatter at a shallow water site," *J. Acoust. Soc. Am.* **77**, 975–982 (1985).
- <sup>5</sup>M. Gensane, "A statistical study of acoustic signals backscattered from the sea bottom," *IEEE J. Ocean Eng.* **14**, 84–93 (1989).
- <sup>6</sup>S. Stanic and E. G. Kennedy, "Reverberation fluctuations from a smooth sea floor," *IEEE J. Ocean Eng.* **18**, 95–99 (1993).
- <sup>7</sup>R. S. Raghavan, "Adaptive processing of microwave sea echo for the suppression of strong reflections from scatterers," *IEEE J. Ocean Eng.* **14**, 76–83 (1989).
- <sup>8</sup>I. Dyer, A. B. Baggeroer, H. Schmidt, J. R. Fricke, N. Ozluer, and D. Giannoni, "Discrete backscatter can be dominant in rough bottom scattering," in *Ocean Reverberation*, edited by D. D. Ellis, J. R. Preston, and

- H. J. Urban (Kluwer Academic, Boston, 1993), pp. 51–57, Sec. 1.
- <sup>9</sup>I. Dyer, "Heuristic facet backscatter theory: a plausible model for dominant backscatter events?" Woods Hole Oceanographic Institution, 1991. Office of Naval Research Acoustic Reverberation Special Research Project Symposium.
- <sup>10</sup>Office of Naval Research, Acoustic Reverberation Special Research Program Initial Report, R/V CORY CHOUET, July 5–26, 1993.
- <sup>11</sup>J. R. Preston, E. Michelozzi, L. Troiano, and R. Hollett, "Acoustic Reverberation Special Research Program Initial Report," R/V ALLIANCE. SACLANTEN and ONR ARSRP group joint experiment, August 30, 1993.
- <sup>12</sup>J. Goff, B. Tucholke, J. Lin, G. Jaroslow, and M. Kleinrock, "Quantitative analysis of abyssal hills in the Atlantic Ocean: A correlation between inferred crustal thickness and extensional faulting," *J. Geophys. Res.* **100**, 22309–22322 (1995).
- <sup>13</sup>B. Tucholke of WHOI provided ARSRP topographic data.
- <sup>14</sup>Briefly, the self-selection mechanism within the discrete scattering model posits that discrete backscattering is dominated by wavelength-scale features of the bottom (Refs. 8, 9). Scattering from smaller features is weak due to their small scattering strength (which decreases with the fourth power of size). Scattering from larger features is strong but directive, i.e., concentrated within their forward scattering lobe. This makes the probability of observing such a feature small. In Refs. 8, 9 it was shown that contributions to the backscattering strength maximizes for sizes  $O(\lambda)$ .
- <sup>15</sup>Y. Y. Dorfman, "Bistatic scattering of acoustic waves from a rough ocean bottom," Ph.D. thesis, Massachusetts Institute of Technology, 1997.
- <sup>16</sup>Office of Naval Research, Acoustic Reverberation Special Research Program Initial Report, R/V CORY CHOUET, August 19, 1991.
- <sup>17</sup>J. A. Goff and T. H. Jordan, "Stochastic modeling of sea floor morphology: Inversion of Sea Beam data for second order statistics," *J. Geophys. Res.* **93**, 13589–13608 (1988).
- <sup>18</sup>T. H. Bell, Jr., "Mesoscale sea floor roughness," *Deep-Sea Res.* **26A**, 65–76 (1979).
- <sup>19</sup>J. Goff and B. Tucholke, "Multiscale spectral analysis of bathymetry on the flank of the Mid-Atlantic Ridge: Modification of the seafloor by mass wasting and sedimentation," *J. Geophys. Res.* **102**, 15447–15462 (1997).
- <sup>20</sup>J. E. Bondaryk, I. Dyer, and Y. Dorfman, "Estimation of bottom scattering characteristics from acoustic reverberation special research program bistatic data," *J. Acoust. Soc. Am.* **98**, 2988 (1995).
- <sup>21</sup>N. C. Makris, L. Z. Avelino, and R. Menis, "Deterministic reverberation from ocean ridges," *J. Acoust. Soc. Am.* **97**, 3547–3574 (1995).
- <sup>22</sup>I. Dyer, A. B. Baggeroer, J. D. Zittel, and R. J. Williams, "Acoustic backscattering from the basin and margins of the Arctic Ocean," *J. Geophys. Res.* **87**, 9477–9488 (1982).
- <sup>23</sup>N. C. Makris, C. S. Chia, and L. T. Fialkowski, "The bi-azimuthal scattering distribution of an abyssal hill," *J. Acoust. Soc. Am.* (to be published).
- <sup>24</sup>For  $\theta_g$  large (group 1), the incident field at the bottom is dominated by sidelobe radiation from the source. Even though the incidence amplitude in this case is not well known, the reverberation exceeds the noise level significantly (Ref. 15), and therefore can be included in our analysis (with the forward HLA beam excluded because of ship-noise). For  $\theta_g$  small (group 2), radiation in the main lobe of the source is dominant at the bottom, and we thus have good estimates of the incident pressure field. In either case, AGC processing removes the mid-time average, and the higher-order statistical comparisons are thus not subject to uncertainties in the mean incident field.
- <sup>25</sup>F. G. Bass and I. M. Fuks, *Wave Scattering from Statistically Rough Surfaces* (Pergamon, New York, 1979).
- <sup>26</sup>The specular reflection direction from the X-axis in Fig. 9 is given by  $\alpha_{\text{spec}} = \theta_g + 2\gamma$ . Then the left boundary of a forward lobe is  $\alpha_{\text{left}} = \theta_g + 2\gamma + \alpha/2$ . The incident angle from X-axis is  $\theta = \pi - \theta_g$ . In our approximation, only facets with  $\theta < \alpha_{\text{left}}$  contribute to scattering, which defines the lower limit of the integral Eq. (3).
- <sup>27</sup>G. V. Trunk, "Radar properties of non-Rayleigh sea clutter," *IEEE Trans. Aerosp. Electron. Syst.* **2**, 196–204 (1972).



# The underwater sounds produced by impacting snowflakes

Lawrence A. Crum

*Applied Physics Laboratory, University of Washington, Seattle, Washington 98105*

Hugh C. Pumphrey

*Department of Meteorology, The University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, Scotland, United Kingdom*

Ronald A. Roy

*Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215*

Andrea Prosperetti

*Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland 21218*

(Received 11 August 1998; revised 2 June 1999; accepted 14 June 1999)

In 1985, Scrimger [Nature **318**, 647 (1985)] reported measurements of noise levels significantly above the ambient level for snow falling on a quiet freshwater lake. He examined only the time-averaged sound levels and did not report measurements of individual snowflake impacts. Subsequently, the noise produced by individual and multiple snowflake impacts was examined for a number of different snowfalls. The radiated acoustic signals generated by the impact of individual snowflakes upon a body of water have a remarkable similarity to each other and differ principally in the frequency of the emitted sound wave. The acoustic signal of a snowflake impact thus generates a characteristic signature for snowfall that is clearly distinct from other forms of precipitation noise. Various aspects of this signature suggest that the radiated acoustic waveform from a snowflake impacting with water is due to the entrainment of a gas bubble into the liquid, and the subsequent oscillation of this bubble as it establishes its equilibrium state. Various scenarios are presented for bubble entrainment and approximations to the amplitude of the radiated signal and the acoustic waveform are obtained. © 1999 Acoustical Society of America.

[S0001-4966(99)02710-1]

PACS numbers: 43.30.Lz, 43.30.Nb [DLB]

## INTRODUCTION

In 1985, while instruments were in place for the determination of the underwater sound produced in a fresh water lake by rain, Scrimger<sup>1</sup> observed significant noise above the background when it chanced to snow. Although he was unable to determine the source mechanism for the sound production, or even the complete frequency spectrum (his hydrophone had a high-frequency cutoff at about 50 kHz), these measurements seem to be the first reports of underwater acoustic emissions produced by falling snow.

Scrimger's initial reports of the noise produced by precipitation<sup>1,2</sup> have been followed by extensive investigations of rain noise by other investigators<sup>3-12</sup> and also some preliminary reports of noise produced by hail.<sup>2,13</sup> McConnell *et al.*<sup>14</sup> also have reported evidence of significant increases in the ambient noise level in an Alaskan fjord during snowfall. A brief review of precipitation noise has been given by Crum *et al.*<sup>15</sup>

We present in this paper some further, yet still preliminary, studies of the acoustic emissions associated with the impact of individual snowflakes on a quiescent water surface, provide evidence that these emissions are associated with gas bubble oscillations, and also present some estimates for the shape of the acoustic waveforms, and the amplitude of the radiated acoustic pressures. Some of the results de-

scribed in this report were presented at the 122nd meeting of the Acoustical Society of America.<sup>16</sup>

## I. RESULTS

After reading the intriguing reports by Scrimger<sup>1,2</sup> of the noise produced by precipitation, a chance snowfall in Mississippi in 1987 enabled us to obtain some measurements of the sound radiated by an individual "impacting" snowflake when it struck the surface of the water contained in a small container, and to obtain an average power spectrum of a number of these impacts. The initial traces were so unique and contrary to our intuitions and expectations that it has inspired us to accumulate data from a number of storms in a number of different locations over the last few years.

We wish to present at this time a brief description of our collected data and offer an explanation for the results that we have of this curious phenomenon. Because the acoustic signal produced by an impacting snowflake is of relatively high frequency and short duration, the apparatus required to obtain these measurements can be quite simple. A small container of water will suffice—the reverberation is insignificant, and one can easily detect reflections from the container walls. Furthermore, the duration of an acoustic pulse (on the order of a few tens of microseconds) is substantially shorter than the time between impacts (a few tenths of a second). Together with a sensitive hydrophone (we used a B&K



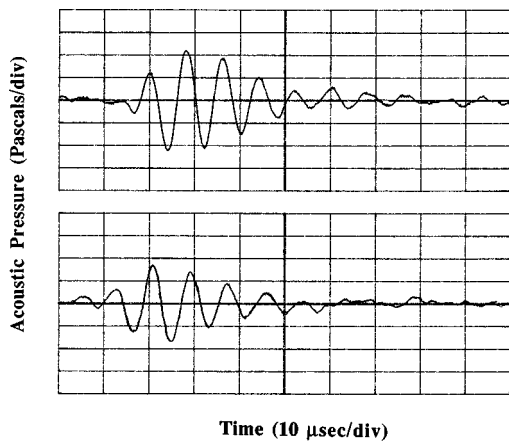


FIG. 1. Pressure-time traces of two individual snowflake impacts obtained for a snowfall in Mississippi. The two traces were obtained for fluffy flakes falling into a container of water; the calibrated hydrophone was positioned about 10 cm below the impact site.

8103), a modern storage oscilloscope with internal memory and mathematical function capability, and a chart recorder for hard output, one can obtain the requisite data. With such an apparatus, we have acquired data for four different snow storms in two different States of the U.S. In these measurements, we commonly looked for the pressure-time history of the event and then computed an average power spectrum from a number of events.

We show in Fig. 1 two pressure-time traces of “light, fluffy” snowflakes impacting a water surface during a snow shower in Mississippi. Note the similarity of the traces. Shown in Fig. 2 are ten pressure-time traces of impacting snowflakes, including those from a snow storm in Virginia, about 2 years later. Also shown is the average power spectrum obtained from 50 such traces. Note again in Fig. 2 the similarity of the traces. Although the various snow storms provided snowflakes of different sizes and shapes, we did not categorize individual pressure-time histories with snowflake morphology. Furthermore, we did not measure the temperature of the water on which the snowflakes fell, although it was significantly higher than 0 °C.

It will be seen from these and other data that an individual, impacting snowflake gives rise to a characteristic underwater signature. We now examine certain features of these traces in an attempt to understand the physical mechanism(s) that gives rise to these signatures.

One characteristic feature of these traces is the decay of the pressure with time. Because these decays resemble the sounds emitted by an oscillating gas bubble, we used these traces to obtain a measurement of the damping constant. Specifically, if the natural log of the ratio of successive maxima and minima is plotted versus cycle number, then the slope of this line is equal to  $(\beta/f)$ , where  $\beta$  is the damping constant and  $f$  is the frequency.<sup>7</sup> Likewise, this slope is equal to  $(\pi/Q)$ , where  $Q$  is the quality factor. Shown in Fig. 3 are measurements of  $Q$  as a function of frequency for a number of snowflake impacts. Plotted also on this figure are the calculated values of  $Q$  for a gas bubble, taken from the theory of Prosperetti.<sup>17</sup> It is seen from this figure that the pressure-time traces are probably the result of oscillating gas bubbles

somehow entrained by the snowflake impact. Since we have previously investigated gas bubble entrainment from rainfall, a review of that process serves to educate us to the possible scenario for the snowfall case.

Consider Fig. 4, which shows three pressure-time traces. The top trace is the pressure-time history of an impacting raindrop whose radius was 1.5 mm and whose impact velocity was 2.0 m/s. The first bump on the trace, occurring at a time of about 12 ms, corresponds to the impulse noise radiated from the water-hammer effect of the impact of the drop with the water surface. The second feature, commencing at about 32 ms, is the sound that results from gas bubble entrainment. This signal is expanded in time and shown in the middle trace of Fig. 4. Note that there is an initial, positive-pressure peak followed by a steady decay to background. The quality factor measured from this oscillation agrees closely with the calculated decay curve shown in Fig. 3.<sup>7</sup> The analysis of similar such traces enabled us to determine that a peak in the light-rainfall spectrum near 15 kHz was due to bubble entrainment.<sup>6</sup>

Also shown in Fig. 4, as the bottom trace, is a pressure-time history resulting from the impact of a snowflake. If one compares this bottom trace with the middle one, it is possible to demonstrate that they both have the characteristic decay of an oscillating gas bubble, and that the quality factors are nearly equal. Beyond this favorable comparison, however, the traces are quite dissimilar. We believe we can offer a plausible explanation for the characteristic shape of this pressure-time history, and thus of the noise produced by snowfall.

Suppose a slowly falling snowflake strikes the flat surface of a large volume of water. Unlike the case for falling raindrops or hailstones, there is little momentum delivered to the surface as a result of the impact. However, a close examination of the water surface indicates a radiating capillary wave subsequent to this impact. It seems likely that, rather than the snowflake creating a depression in the water surface, surface tension and capillary forces cause the water to rise up along the many surfaces of the snowflake, creating a small protuberance or bump. The subsequent relaxation of this slight elevation results in the radiating capillary wave. We suggest that the encapsulation of a gas bubble within this small protuberance is the reason the snowflake trace has its characteristic shape. If the oscillating bubble produced by the snowflake is contained in a projection extending above the surface, it becomes an inefficient radiator; however, when the surface tension and gravitation forces pull the bubble down below the surface, it can radiate much more efficiently. It is noted from the various pressure traces of snowflake impacts that the first peak in these traces may be either positive or negative. In the case of a raindrop impact that produces a bubble, the first peak is always positive. If the snowflake/bubble is “inserted,” rather than “produced,” underwater, the polarity of the first peak would depend upon the insertion time, and could be either positive or negative.

We can also offer some quantitative arguments for this hypothesis. Suppose that a bubble of radius  $R$  is encapsulated within a protuberance generated by snowflake impact. Consider the time required for the protuberance to recede. We

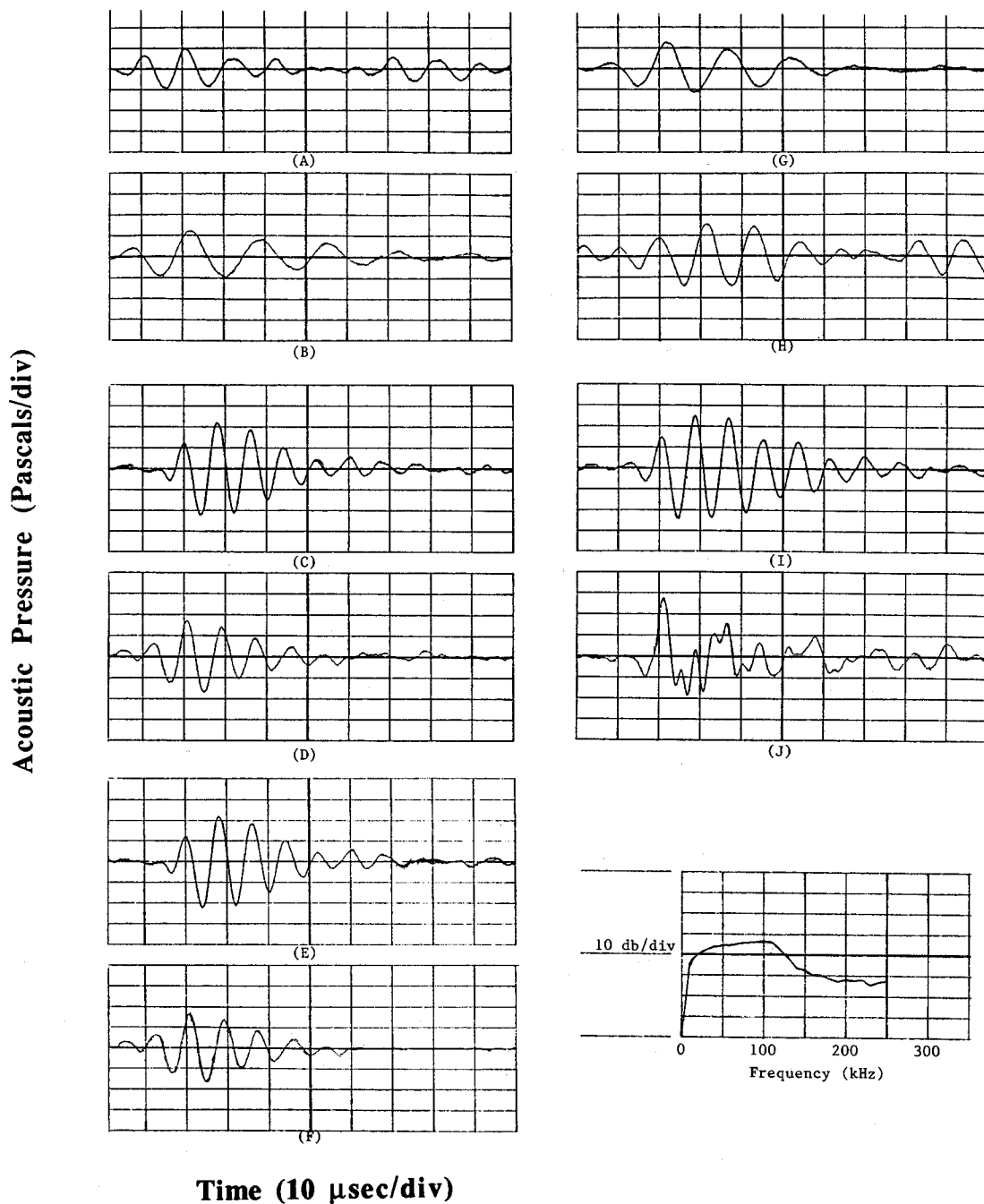


FIG. 2. A collection of some representative pressure-time traces obtained from different snow storms in different States of the U.S. Note the similarity of the curves, and that multiple events (trace J) are uncommon, although successive ones are not (traces A and H). The insert in the lower right is an average power spectrum for 50 traces during a single storm. The falloff in the spectrum above 100 kHz matches that for the hydrophone frequency response.

can estimate this time by examining the characteristic times associated with the propagation of a capillary wave. The velocity of a capillary wave is given by

$$v = (2\pi\sigma/\rho\lambda)^{1/2}, \quad (1)$$

where  $\sigma$  is the surface tension,  $\rho$  the density, and  $\lambda$  the wavelength, all of the liquid medium, such as water. The time required for the bump to be withdrawn completely below the liquid surface is one-half the period  $T$ . Thus, the bubble "insertion time,"  $t_i$ , is given approximately by

$$t_i = T/2 = [\rho\lambda^3/8\pi\sigma]^{1/2}. \quad (2a)$$

A reasonable estimate for the wavelength of the capillary wave would be twice the bubble radius,  $R$ . For frequencies in the range of 25–100 kHz, the relation between the bubble radius,  $R$ , and its natural frequency of oscillation,  $f$ , is given approximately by  $Rf = 310$  (cgs units). Using the values for water of  $\rho = 1.0 \text{ gm/cm}^3$  and  $\sigma = 72 \text{ dyn/cm}$ , we can rewrite Eq. (2a) as

$$t_i = 363/f^{3/2}. \quad (2b)$$

We thus find that for a frequency of 25 kHz, the insertion time should be approximately 92  $\mu\text{s}$ ; for a frequency of 100 kHz, it should be about 11  $\mu\text{s}$ . If we examine the pressure-

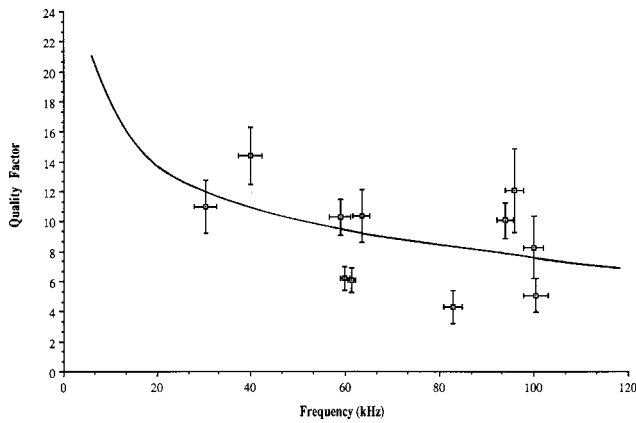


FIG. 3. Measurements of the quality factor associated with gas bubble oscillations produced by impacting snowflakes. The symbols refer to values of the quality factor obtained from pressure-time traces similar to those shown in Figs. 1 and 2. The solid line is the theoretical dependence of this factor on the frequency as obtained by Prosperetti (Ref. 17). This figure demonstrates that impacting snowflakes generate underwater signals that are probably associated with gas bubble oscillations.

time traces for snowflake impacts shown in Figs. 1, 2, and 4, we see that the time required for growth of the signal to a maximum amplitude, a measure of the insertion time, is on the order of 40  $\mu$ s. Thus, we see that our rough estimate of the insertion time is about right.

Consider next an examination of the absolute magnitude of the signal produced by the impacting snowflake. Later in this paper, we shall address the issue of the origin of the entrained gas bubble. Let us assume at this time that it exists and we wish to estimate the magnitude of the radiated pressure wave, which, of course, can be measured.

The bubble is radiating near the surface; we write as the expression for the pressure radiated by a dipole source of radius  $R_0$  at a distance  $d$  below the water surface,<sup>18</sup>

$$p(r, \theta, t) = \frac{2\rho c k^2 R_0^2 d}{r} U_0 e^{-\beta(t-r/c)} \cos \theta e^{i(\omega t - kr)}, \quad (3)$$

where  $r$ ,  $\theta$ ,  $t$  are the respective radial, angular, and temporal variables,  $c$  is the velocity of sound,  $k$  is the wave number, and  $\beta$  is the damping constant. The quantity  $U_0$  is the velocity amplitude, given by  $U = \omega_0 \Delta R$ , where  $\Delta R$  is the displacement amplitude and  $\omega_0$  is the natural angular resonance frequency of the bubble. It is assumed that when the bubble is created, it changes its equilibrium radius due to the force of surface tension, and the reestablishment of equilibrium results in radiated acoustic energy. We can find the magnitude of the source strength by considering the following equation that describes the work required to create the bubble:

$$(4\pi R_i^3/3)P_0 = (4\pi R_f^3/3)(P_0 + 2\sigma/R_f), \quad (4)$$

where  $R_i$  and  $R_f$  are initial and final bubble radii,  $P_0$  is the magnitude of the ambient pressure, and  $2\sigma/R_f$  is the ‘‘Laplace pressure’’ associated with surface tension. This relation leads to

$$\Delta R = R_i - R_f = 2\sigma/3P_0. \quad (5)$$

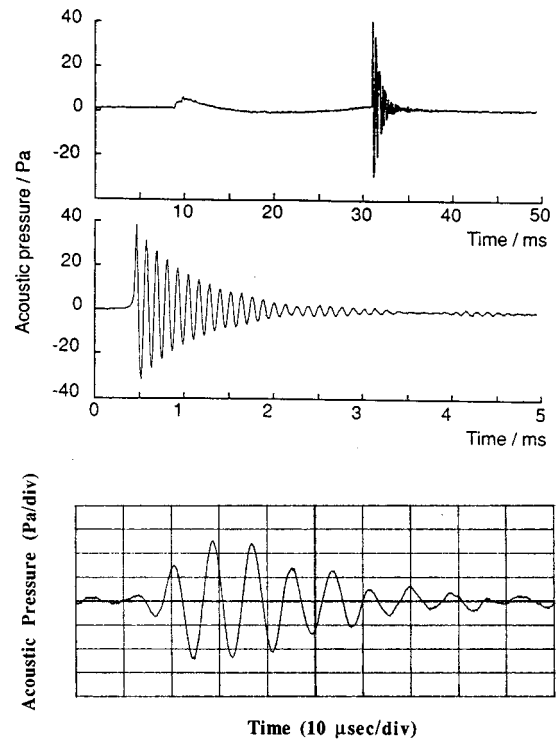


FIG. 4. Comparison of underwater acoustic emissions from raindrop and snowflake impacts. The top trace shows the underwater noise produced by an impacting raindrop. The first bump on the curve is associated with the direct impact of the drop with the surface and is mostly hydrodynamic in nature; the second bump, which is expanded in the second trace, shows a decaying sinusoid that is associated with the entrainment of a gas bubble by the impact process (see for example, Refs. 5 and 10). The bottom trace shows a pressure-time trace for a snowflake impact. Because the decay constant of these two traces are essentially identical and equal to that for freely oscillating gas bubbles, it can be presumed that gas bubbles are involved in both cases. The initial growth in the snowflake trace, which is absent in the raindrop case, indicates the existence of a physical mechanism that is not clearly understood. It is noted that similar geometries (size of container, position of hydrophone) were utilized in both the raindrop and snowflake impact measurements.

When Eq. (5) is used to express the source strength in Eq. (3), then the magnitude of the peak acoustic pressure amplitude,  $P_m$ , is given by

$$P_m = \frac{4}{3} \frac{(2\pi)^3 \rho f^3 R_0^2 d \sigma}{P_0 c r}. \quad (6)$$

If we use the following typical values of the relevant quantities,  $\rho = 1.0 \text{ gm/cm}^3$ ,  $f = 100 \text{ kHz}$ ,  $R_0 = 30 \mu\text{m}$ ,  $d = 1 \text{ mm}$ ,  $\sigma = 72 \text{ dyn/cm}$ ,  $P_0 = 1.0 \times 10^6 \text{ dyn/cm}^2$ , and  $c = 1.5 \times 10^5 \text{ cm/s}$ , then the peak pressure observed on the dipole axis at a distance of 5 cm below the surface is given by Eq. (6) to be about 2.8 Pa. It can be seen from Fig. 2 that some representative *measured* values were 2.3, 3.0, 2.2, and 2.0 Pa. We see that this comparison between estimated and measured absolute pressure values is rather good, recognizing the crudeness of the approximations. We also note that the amplitude of these acoustic pressures is of the same order of magnitude of those for individual raindrop impacts. However, because the frequency of the noise radiated by individual snowflakes is rarely lower than 50 kHz, one does not hear the ‘‘plunk’’ of an impacting snowflake.

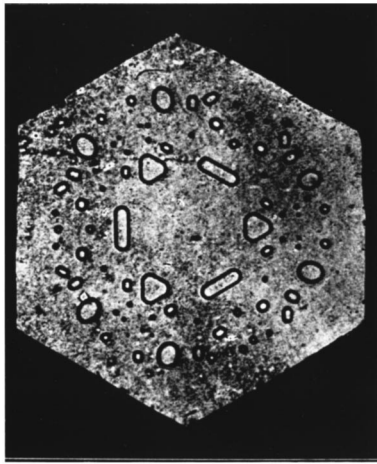


FIG. 5. Photograph of an ice crystal containing gas bubbles. If this crystal were to melt quickly, while immersed, the escaping gas could lead to acoustic emissions. The size of these crystals depends upon the local temperature at formation and may range from  $5\ \mu\text{m}$  to  $5\ \text{mm}$  (from Ref. 19).

## II. DISCUSSION

Our analysis of the data presented in this study suggests that a snowflake striking a body of water produces a sound most likely by entraining a gas bubble into the liquid, and by the subsequent oscillation of this bubble. We offer now some suggestions as to possible mechanisms for bubble entrainment:

### A. Air engulfment

Snowflakes are usually loose agglomerates of individual ice crystals with a significant air content as indicated by the fact that their mean density is usually only 10% of that of water.<sup>19</sup> When such a fluffy snowflake strikes a water surface, it almost immediately stops, having no momentum to depress the surface. It seems likely that water would rapidly move upward through the flake by capillary action and lead to a rapid melting of the individual ice crystals. The end result would be the formation of a small “foam patch” that, even though it consists of more than one bubble, would radiate as a single entity due to the near-field acoustic coupling of its constituents. An alternative scenario is that the individual small bubbles could quickly coalesce into a single one if the thin liquid membranes separating them burst. It would probably be difficult to distinguish between these two possibilities on the basis of their acoustic radiation properties and it appears that, in both cases, acoustic emissions of the type observed in this study would be likely. This emission process depends on a series of events whose precise sequence and timing may not occur with every flake. Indeed, our casual observation indicates that only a small fraction (say, 1 in 10) of falling snowflakes produces a pressure trace.

### B. Frozen bubble release

Snowflake experts tell us that many individual ice crystals contain small pockets of air trapped within the ice itself.<sup>19</sup> Figure 5 shows an example of such a flake. If this crystal were to melt rapidly, the bubble would be released from its ambient state, and oscillate—indeed, the gas within

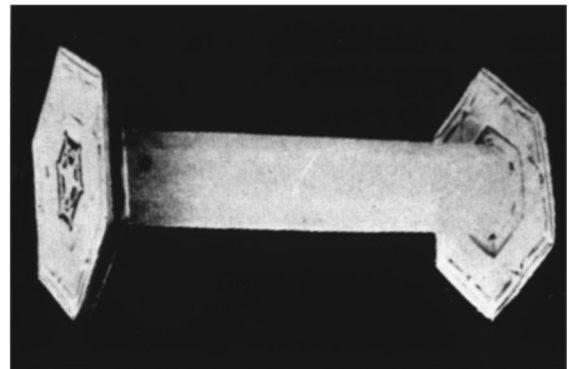
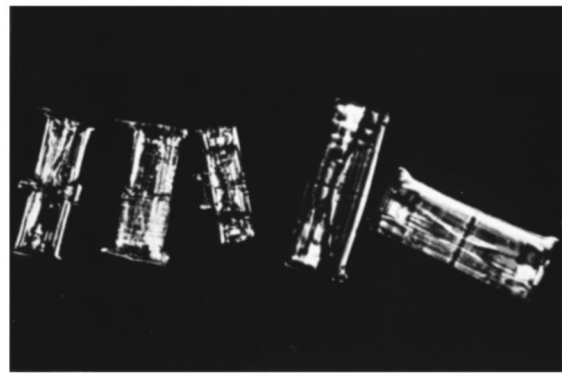


FIG. 6. Photograph of hollow ice crystals. If these crystals were to be quickly immersed in water, they could act as the source of the gas bubbles that apparently lead to acoustic emissions from impacting snowflakes from Ref. 20.

the bubble might be either compressed or rarefied. In this case, a different source mechanism for the acoustic energy is suggested, and the bubble oscillation amplitude should be significantly larger—the volume change during the freezing of water is on the order of 10%, and a bubble volume displacement of this magnitude (that would occur unless some of the entrapped air diffused out of the cavity) is relatively large.

### C. Hollow ice crystals

Although relatively rare, some snowflakes take the geometrical shape of hollow objects such as cylinders and rectangles. Figure 6 shows photographs of some hollow ice crystals.<sup>20</sup> For this scenario, one could expect that only the occasional snowflake would possess such a configuration. Similarly, one would expect this mechanism to provide easy encapsulation of air. Moreover, little energy would be stored in the air encapsulation and thus our previous arguments apply that suggest surface tension as the principal oscillation forcing mechanism.

## III. SUMMARY AND CONCLUSION

We have observed that snowflakes falling into a body of water produce noise levels significantly in excess of background. An analysis of individual flake impacts suggests that the principal noise source is an oscillating gas bubble that is somehow entrained within the water. We have speculated about mechanisms for bubble entrainment and have made



some order-of-magnitude calculations that tend to support our hypothesis. Additional experiments must be performed of the encapsulation process and the dynamics of bubble oscillation before an adequate explanation of this phenomenon can be given. In particular, high speed movies of the entrainment process coupled with simultaneous acoustic waveform measurements would be particularly revealing. Also, if ice crystal melting is important, then the temperature of the host liquid should have an effect on the pressure waveforms.

## ACKNOWLEDGMENTS

We wish to acknowledge the assistance of Ed Gardner with data acquisition for the Virginia snowfall and we offer a special thanks to Bob Apfel, who loaned us his car and some apparatus for this storm. We also wish to acknowledge that part of this work (a general study of precipitation noise) was supported by the Office of Naval Research.

- <sup>1</sup>J. A. Scrimger, "Underwater noise caused by precipitation," *Nature* (London) **318**, 647–649 (1985).
- <sup>2</sup>J. A. Scrimger, D. J. Evans, G. A. McBean, D. M. Farmer, and B. R. Kerman, "Underwater noise due to rain, hail, and snow," *J. Acoust. Soc. Am.* **81**, 79–85 (1987).
- <sup>3</sup>J. A. Nystuen, "Rainfall measurements using underwater ambient noise," *J. Acoust. Soc. Am.* **79**, 972–980 (1986).
- <sup>4</sup>H. C. Pumphrey and L. A. Crum, "Acoustic emissions associated with drop impacts," in *Sea Surface Sound*, edited by B. R. Kerman (Kluwer Academic, Norwell, Massachusetts, 1988), pp. 463–484.
- <sup>5</sup>H. C. Pumphrey, L. A. Crum, and L. Bjørnø, "Underwater sound produced by individual drop impacts and rainfall," *J. Acoust. Soc. Am.* **85**, 1518–1526 (1989).

- <sup>6</sup>A. Prosperetti, L. A. Crum, and H. C. Pumphrey, "The underwater noise of rain," *J. Geophys. Res.* **94**, 32–39 (1989).
- <sup>7</sup>H. C. Pumphrey and L. A. Crum, "Free oscillations of near-surface bubbles as a source of the underwater noise of rain," *J. Acoust. Soc. Am.* **87**, 142–150 (1990).
- <sup>8</sup>M. S. Longuet-Higgins, "An analytical model of sound production by raindrops," *J. Fluid Mech.* **214**, 395–402 (1990).
- <sup>9</sup>H. N. Oguz and A. Prosperetti, "Bubble entrainment by the impact of drops on liquid surfaces," *J. Fluid Mech.* **219**, 143–153 (1990).
- <sup>10</sup>H. N. Oguz and A. Prosperetti, "The underwater noise of rain," *J. Fluid Mech.* **228**, 417–431 (1991).
- <sup>11</sup>H. C. Pumphrey and P. A. Elmore, "The entrainment of bubbles by drop impacts," *J. Fluid Mech.* **220**, 539–549 (1990).
- <sup>12</sup>H. Medwin, J. A. Nystuen, P. W. Jacobus, D. E. Snyder, and L. H. Ostwald, "The anatomy of underwater rain noise," *J. Acoust. Soc. Am.* **92**, 1613–1624 (1992).
- <sup>13</sup>M. Nicholas, R. A. Roy, and L. A. Crum, "The underwater sound produced by the impact of solid objects," *J. Acoust. Soc. Am.* **89**, 1863–1864 (1991).
- <sup>14</sup>S. O. McConnell, M. P. Scott, and J. G. Dworski, "Ambient noise measurements from 100 Hz to 80 kHz in an Alaskan fjord," *J. Acoust. Soc. Am.* **91**, 1990–1997 (1992).
- <sup>15</sup>L. A. Crum, R. A. Roy, and A. Prosperetti, "The underwater noise of precipitation," *Nav. Res. Rev.* **24**, 2–11 (1992).
- <sup>16</sup>L. A. Crum, R. A. Roy, H. C. Pumphrey, and A. Prosperetti, "The underwater sounds of falling snow," *J. Acoust. Soc. Am.* **90**, 2318–2319 (1991).
- <sup>17</sup>A. Prosperetti, "Bubble phenomena in sound fields," *Ultrasonics* **22**, 69–82 (1984).
- <sup>18</sup>L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
- <sup>19</sup>B. J. Mason, *The Physics of Clouds* (Clarendon Press, Oxford, 1971), Chap. 5.
- <sup>20</sup>W. A. Bentley and W. J. Humphreys, *Snow Crystals* (Dover, New York, 1962).

# A Gaussian finite-element method for description of sound diffraction

Dehua Huang<sup>a)</sup> and M. A. Breazeale

National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

(Received 30 September 1997; revised 20 April 1999; accepted 2 June 1999)

A new method for solving diffraction problems is presented. It is based on the use of Gaussian diffraction theory. Comparison of the Gaussian beam expansion and Fourier series expansion reveals that the Gaussian expansion is a more general and more powerful technique. The new method combines the Gaussian beam superposition technique [Wen and Breazeale, *J. Acoust. Soc. Am.* **83**, 1752–1756 (1988)] and the numerical (finite-element) solution to the parabolic equation [Huang, *J. Acoust. Soc. Am.* **84**, 1405–1413 (1988)]. The new method is capable of solving for the sound field even in an inhomogeneous medium or in the presence of reflection, as shown by computer modeling. The source may be a Gaussian source or a distributed source. Calculated results are compared with experimental results in a laboratory water tank. © 1999 Acoustical Society of America. [S0001-4966(99)02509-6]

PACS numbers: 43.35.Bf [HEB]

## INTRODUCTION

In physics the propagation of waves in inhomogeneous media often is encountered. This is a basic problem in radar or in astronomy, and it also is a basic problem in acoustics. If, in addition, one introduces the effect of diffraction, then the solution becomes even more complicated. Even this complication, however, can be handled if one defines the solution in terms of Gaussian functions.<sup>1</sup>

The Rayleigh surface integral<sup>2</sup> based on Huygens' theorem is considered the first exact expression of the diffraction problem. Although this theory was developed for light diffraction, it can be used directly to describe diffraction of sound waves. The general diffraction at a circular aperture for acoustical problems becomes radiation from a piston radiator. A general review of the many integral expressions describing the piston radiator is presented by Harris.<sup>3</sup> Infinite series solutions for describing beam field problems appear more attractive than spatial Fourier analysis because of their better performance in mathematical analyses. Many series solutions using different basis function sets have been developed to represent the diffraction field of a piston transducer. Gaussian–Laguerre functions,<sup>4</sup> Bessel functions,<sup>5</sup> and products of spherical Bessel functions, Hankel functions, and Legendre polynomials<sup>6</sup> are some basis function sets employed in series solutions. However, because of the discontinuity at the edge of a piston, the number of terms needed to represent the radiated field usually is quite large. The mathematical analysis is correspondingly difficult with a large number of functions having different analytical form.

Besides the rigid piston transducer, planar ultrasonic sources with other surface distributions have been investigated. Greenspan<sup>7</sup> and Novikov *et al.*<sup>8</sup> studied source distributions with different degrees of discontinuity. Tjøtta and Tjøtta<sup>9</sup> reported studies of the Gaussian radiator function with a smooth edge.

Gaussian diffraction theory has been developed over the past several years. Du and Breazeale<sup>10–12</sup> showed that the diffraction field of a Gaussian transducer is readily described by Gaussian theory.

The Gaussian beam optimization representation was developed by Wen and Breazeale.<sup>1,13</sup> It uses the superposition of Gaussian functions to describe the diffraction by a piston source. In the Gaussian beam representation a set of Gaussian beams with different beam parameters is employed as the basic unit rather than infinite plane waves propagating in different directions, as is used in Fourier diffraction theory. Through computer determination of the coefficients and beam parameter sets imposed by the physical constraints, an axially symmetrical field can be expressed in a unified, analytical form—the superposition of a number of Gaussian beams. Using the optimization technique, a 10-term expansion of the field of a piston transducer in terms of Gaussian functions gives a better solution than that produced by a Gaussian–Laguerre series expansion technique with 65 terms.<sup>4</sup> To obtain an even better accuracy, we adopt the 15-term example from Ref. 13. The parameters of the 15-term Gaussian beam are listed in Table I. Using this Gaussian technique, any distributed planar source function can be represented as a summation of a series of Gaussian beam functions with different waists and phases.<sup>13</sup>

For a cylindrically symmetric treatment of an inhomogeneous medium such as water, the acoustic pressure  $p(r, z)$  satisfies the Helmholtz equation

$$\nabla^2 p + k_0^2 n^2 p = 0, \quad (1)$$

where  $k_0$  and  $n$  are the reference wave number and the refractive index (which can be a function of  $z$  or  $r$ ). We substitute

$$p(r, z) = u(r, z) H_0^1(kr) \quad (2)$$

into Eq. (1), and assume that  $u(r, z)$  is weakly dependent on  $r$ , so that  $u_{rr}$  is negligible. The result obtained is the Leontovich-Fock<sup>14</sup> parabolic wave equation

<sup>a)</sup>Electronic mail: DHHuang@AOL.com

TABLE I. Coefficients of a 10-term Gaussian beam solution for a rigid piston transducer.

$n$	$A_n$	$B_n$
1	$-2.9716+8.6187 i$	$4.1869-5.1560 i$
2	$-3.4811+0.9687 i$	$3.8398-10.8004 i$
3	$-1.3982-0.8128 i$	$3.4355-16.3582 i$
4	$0.0773-0.3303 i$	$2.4618-27.7134 i$
5	$2.8798+1.6109 i$	$5.4699+28.6319 i$
6	$0.1259-0.0957 i$	$1.9833-33.2885 i$
7	$-0.2641-0.6723 i$	$2.9335-22.0151 i$
8	$18.019+7.8291 i$	$6.3036+36.7772 i$
9	$0.0518+0.0182 i$	$1.3046-38.4650 i$
10	$-16.9438-9.9384 i$	$6.5889+37.0680 i$
11	$0.3708+5.4522 i$	$5.5518+22.4255 i$
12	$-6.6929+4.0722 i$	$5.4013+16.7326 i$
13	$-9.3638-4.9998 i$	$5.1498+11.1249 i$
14	$1.5872-15.4212 i$	$4.9665+5.6855 i$
15	$19.0024+3.6850 i$	$4.6296+0.3055 i$

$$2ik_0u_r + u_{zz} + k_0^2(n^2 - 1)u = 0. \quad (3)$$

Since this standard parabolic wave equation was introduced in underwater acoustics by Tappert,<sup>15</sup> many authors have studied the equation and have tried to solve it in different ways. Tappert and Hardin<sup>16</sup> solved this equation using the split-step Fourier transform method and this method also gave a successful prediction for the South China Sea experiment.<sup>17</sup> Lee and Papadakis<sup>18</sup> developed an ordinary differential equation method for solving the parabolic equation, and Lee and Botseas<sup>19</sup> solved it by an implicit finite difference method. To treat boundary conditions at interfaces with different sound speeds and densities, McDaniel and Lee<sup>20</sup> and Lee and McDaniel<sup>21</sup> introduced a technique based on a Taylor series expansion and incorporated that technique into their implicit finite difference model, so they can describe horizontal or irregular interface conditions between two media. Huang<sup>22</sup> applied the finite-element method to solve the underwater parabolic equation with interface boundary conditions. Continuity of both pressure and the normal component of particle velocity are satisfied at the horizontal or irregular interfaces. With the finite-element method one can change the number of nodes within a local depth range. This advantage of the finite-element method over the finite difference method leads to a more detailed calculation as well as a more rapid one.

In this paper the specialization is in the direction of defining conditions that would be encountered in sound propagation and communication. The effect of diffraction is simplified by assuming a Gaussian source, then the diffraction is generalized by using the superposition of Gaussian functions<sup>1</sup> to solve the general diffraction problem of apertures of arbitrary size and shape. The results then are used in a finite-element calculation capable of describing propagation through inhomogeneous media. In this paper the new approach is called the Gaussian finite-element method. It is described in detail and some of the numerical results are tested by comparison with experiments.

## I. THEORY

### A. Gaussian beam superposition procedure

For each source an optimization procedure is necessary to find coefficients of the Gaussian functions making up the series representation of the diffraction field. The extreme example would be the piston source because of the discontinuity at the edges. This source, however, has been solved.<sup>13</sup> Other sources can be handled as well. For a piston source of radius  $a$  the axisymmetric velocity distribution is expressed by

$$V(\xi) = 1 \quad \text{for } \xi < a \\ = 0 \quad \text{otherwise.} \quad (4)$$

This radiator can be expressed as a superposition of Gaussian functions:

$$V_0(\xi) = \sum_{n=1}^N A_n \exp[-B_n \xi^2]. \quad (5)$$

An optimized set of  $A_n$  and  $B_n$  valid throughout the radiation field of a piston transducer was found by Wen and Breazeale.<sup>13</sup> It is given in Table I.

For each Gaussian beam with coefficients  $A$  and  $B$ , one can express  $B$  in terms of the half-beam waist  $w$ :

$$B = \frac{1}{w^* w}. \quad (6)$$

The corresponding sound field in a homogeneous medium is

$$p(\hat{r}, \hat{z}) = \frac{\rho c w^2 A \exp[ik\hat{z}]}{w^2 + \frac{2i\hat{z}}{k}} \exp\left[-\frac{\hat{r}^2}{w^2 + \frac{2i\hat{z}}{k}}\right], \quad (7)$$

which can be derived from the parabolic approximation<sup>1,10-13</sup> or directly from the Rayleigh integral path.<sup>23</sup> Here,  $\hat{r}$  is the transducer radius direction, and the  $\hat{z}$ -axis is normal to the center of the transducer face.

### B. Comparison of Gaussian beam expansion and Fourier series expansion

The Gaussian beam expansion allows one to approximate a diffraction field by adjusting relative phases and amplitudes of the components. To understand how this takes place, it is useful to define the relationship between the Laurent expansion and the Taylor expansion.

The Taylor expansion for  $f(x)$  takes the form

$$f(x) = \sum_{n=0}^{\infty} \alpha_n (x - x_0)^n, \quad (8)$$

where  $x$  is real and  $\alpha_n = f^n(x_0)/n!$ . If one extends the Taylor expansion from the real plane to the complex plane, then one has a Laurent expansion. The Laurent series, then, for the complex function  $f(z)$  is

$$f(z) = \sum_{n=-\infty}^{\infty} b_n (z - z_0)^n \quad (9)$$

with

$$b_n = \frac{1}{2\pi i} \oint_{C_R} \frac{f(\xi)}{(\xi - z_0)^{n+1}} d\xi, \quad (10)$$

and  $C_R$  is the circle of integration on a plane with only a simple pole  $z_0$  inside.

In analogy to the Taylor Series, we consider the Fourier series expansion

$$f(x) = \sum_{n=0}^{\infty} (a_n \cos(k_n x) + b_n \sin(k_n x)); \quad (11)$$

here  $k_n = (n\pi)/L$  and  $f(x)$  is a function which satisfies the Fourier conditions

- (i)  $f(x + 2L) = f(x)$ ,
- (ii)  $f(x)$  is sectionally continuous in the interval  $(-L, +L)$ .

Since  $\cos(t) = (e^{it} + e^{-it})/2$  and  $\sin(t) = (e^{it} - e^{-it})/2i$ , Eq. (11) can be written as

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} C_n \exp[ik_n x] + \sum_{n=0}^{\infty} D_n \exp[-ik_n x] \\ &= \sum_{n=-\infty}^{\infty} A_n \exp[ik_n x]. \end{aligned} \quad (12)$$

Here both  $k_n$  and  $x$  are real, and  $C_n = (a_n - ib_n)/2$ ;  $D_n = (a_n + ib_n)/2$ . The  $\exp[ik_n x]$  can be considered as a set of orthogonal basis functions.

Now, to define the expansion in terms of Gaussian functions, we extend the real wave number  $k_n$  by making it complex. We have

$$k_n = K_n + i\alpha_n, \quad (13)$$

where we restrict  $\alpha_n x \geq 0$  in order to have a convergent series. Equation (12) becomes

$$f(x) = \sum_{n=-\infty}^{\infty} A_n \exp[-(\alpha_n - iK_n)x]. \quad (14)$$

Let  $B_n = \alpha_n - iK_n$  and  $x = \xi^2$ . Equation (14) then becomes

$$F(\xi) = f(\xi^2) = \sum_{n=-\infty}^{\infty} A_n \exp[-B_n \xi^2]. \quad (15)$$

This is the expansion in terms of Gaussian functions. If one wants to use a finite set of terms instead of an infinite series expansion for a source function  $F(\xi)$ , the best way is to use the optimization procedure.<sup>13</sup> Other expansions such as Gaussian-Laguerre expansion<sup>4</sup> or self-consistent Gaussian beam superposition via a biorthogonal function method<sup>24,23</sup> are mathematically rigorous, but with the rigor comes complications that can be circumvented by computer optimization.<sup>1</sup>

### C. Finite-element solution to the parabolic equation applied to underwater acoustics

For a Gaussian distributed source, Gaussian diffraction theory is mathematically analytic. However, it is valid only in a homogeneous medium. The optimization procedure allows one to express any axisymmetric beam field in an ana-

lytic form—the superposition of Gaussian beams about the same axis, but with beam waists of different size located at different positions along the axis. One is not limited to piston sources. Any distributed planar source function can be represented as a summation of a series of Gaussian beam functions with different waists and phases. The finite-element solution numerical scheme can handle either horizontal or irregular interface boundary conditions to the parabolic wave equation.<sup>22</sup>

A possibility then exists: Let Gaussian diffraction theory take care of a distributed source to generate a fast starting field, and the finite-element solution deal with the medium. Combining the two techniques results in a new method for rapid solution of underwater acoustics problems that has the strengths of the Gaussian beam representation and the facility of the finite-element solution.

Since the Gaussian source and the piston distributed source are representative of two extreme types of basic sources, they are used in this paper to illustrate the use of this new modeling technique and its experimental verification.

To use the new Gaussian finite-element method, we proceed as follows:

- (i) We express the source function as a single Gaussian beam or as the superposition of a number of Gaussian beams found by optimization in the transducer coordinates  $\hat{r}$  and  $\hat{z}$ .
- (ii) With the coefficients  $A_i$ ,  $B_i$ , of each Gaussian beam, the sound field  $p(R_0, z)$  expressed in the Leontovich-Fock parabolic equation (PE) coordinates, at a distance  $R_0$  away from the source is calculated. In so doing, to achieve rapid simulation we have approximated the medium within the distance  $R_0$  by a homogeneous one.
- (iii) This sound field profile  $p(R_0, z)$ , serves as the initial value starter for the finite-element calculation. The starter represents a distributed source beam profile (not a point source).
- (iv) The finite-element method is used to calculate the sound field at distances  $r > R_0$ . In so doing, inhomogeneity of the medium for  $r > R_0$  can be considered.

## II. NUMERICAL SIMULATION AND COMPUTER MODELING

The new mathematical method can be tested directly. With a single Gaussian beam in a homogeneous medium the new method should give the known solution predicted by Gaussian theory. With a piston beam field, the new method should also match the analytical results of using Gaussian beam superposition optimization representation as calculated by Wen and Breazeale<sup>1</sup> with 10 or 15 terms. As shown below, both of these situations are adequately described. This verifies the validity of the approach. The technique is further tested for a Gaussian source or piston distributed source in an inhomogeneous medium.

The numerical simulation continues with a beam propagating near a sloping bottom. A comparison is made between the fields of a Gaussian transducer and a piston transducer



approximated by 15 Gaussian functions. An improvement of the resolution enhances this comparison considerably.

### A. Single Gaussian source in a homogeneous medium

To demonstrate the capabilities of the new computational method, the first step in computer simulation is to introduce a single Gaussian beam into the calculation without surface reflection and with uniform velocity and density distribution. The Gaussian finite-element method then should give the known Gaussian solution of Eq. (15). We take one of the 15 optimized Gaussian beams (Table I) as an example. We choose the beam parameters:

$$A = -2.9716 + 8.6187 i, \quad B = 4.1869 - 5.1560 i.$$

The radiator is 1-in. diameter with 2.03 MHz. The medium has a sound velocity of 1500 m/s and a density of 1 gm/cm<sup>3</sup>. To simulate a homogeneous infinite half space numerically, we set the source depth at 1 m (1333 wavelengths), far from the boundaries. The source is assumed to be 1.3 m (1733 wavelengths) from the bottom boundary. We use Eq. (7) to generate an initial pressure profile at the distance  $R_0$  (chosen to be 3 cm). (We use transducer coordinates and a PE field coordinate transformation.) The range increment  $\Delta r$  is 0.0001 m (0.13 wavelength). The Gaussian finite-element method results calculated at 4 cm, 12 cm, and 32 cm are directly compared with Gaussian theory. The results at these ranges are shown in Fig. 1. Comparison of the results shows that the new calculation accurately represents the Gaussian function over the entire range as the beam diverges. In this case the beam waist is at  $-8$  cm (behind the radiator) because of the initial assumption. The half-width at the beam waist is 0.39 cm.

### B. Piston beam propagation in a homogeneous medium

The next step is to try a distributed source in a uniform medium without surface reflection. If one begins with geometries in which radiation patterns are known, then one can compare the results directly. The piston source generates a known diffraction field. It also has been expressed as a series of Gaussian beam functions<sup>1</sup> by use of the optimization technique. This procedure now is used along with the finite-element method to calculate the diffraction field of a piston source in a homogeneous medium.

We use the same parameters as in the previous example. The frequency is 2.03 MHz, while the diameter of the piston is 1 in. Instead of one beam, multiple Gaussian beams are used to calculate a piston diffraction field.

Figure 2 shows the 15-Gaussian beam expansion of the piston beam field and the Gaussian finite-element method comparison. The Gaussian analytic and Gaussian finite-element solutions agree very well. The reference distance used here is the Fresnel distance, which is 21.8 cm in this example.

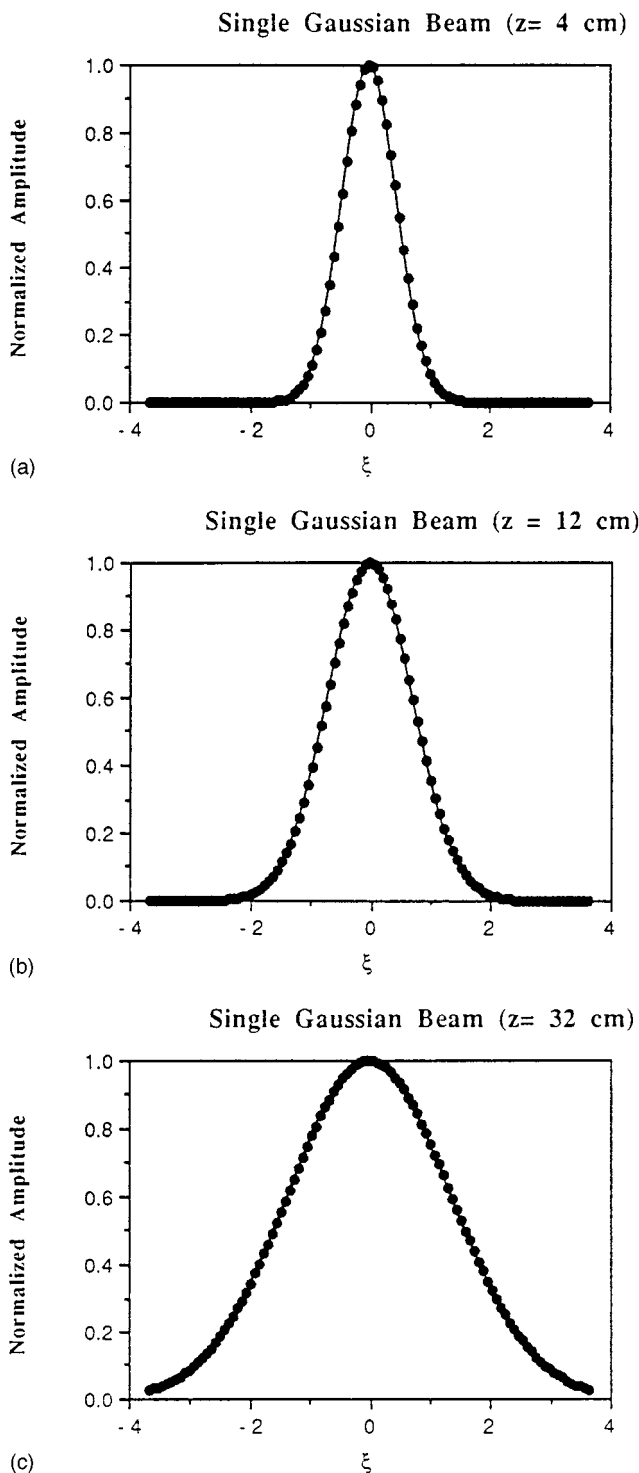


FIG. 1. Transverse beam pressure profiles at 2.03 MHz in a homogeneous medium, where  $\xi = r/a$ , and with  $B = 4.1869 - 5.1560 i$  and  $a = 1.27$  cm. Curves are Gaussian theory; dots are finite-element method. (a) 4 cm; (b) 12 cm; (c) 32 cm.

### C. Beam propagation in an inhomogeneous medium with an irregular interface

A medium with irregular interfaces is a more realistic geometry than horizontal regular interfaces. To simulate an irregular interface one must solve a more complicated interface boundary value problem.

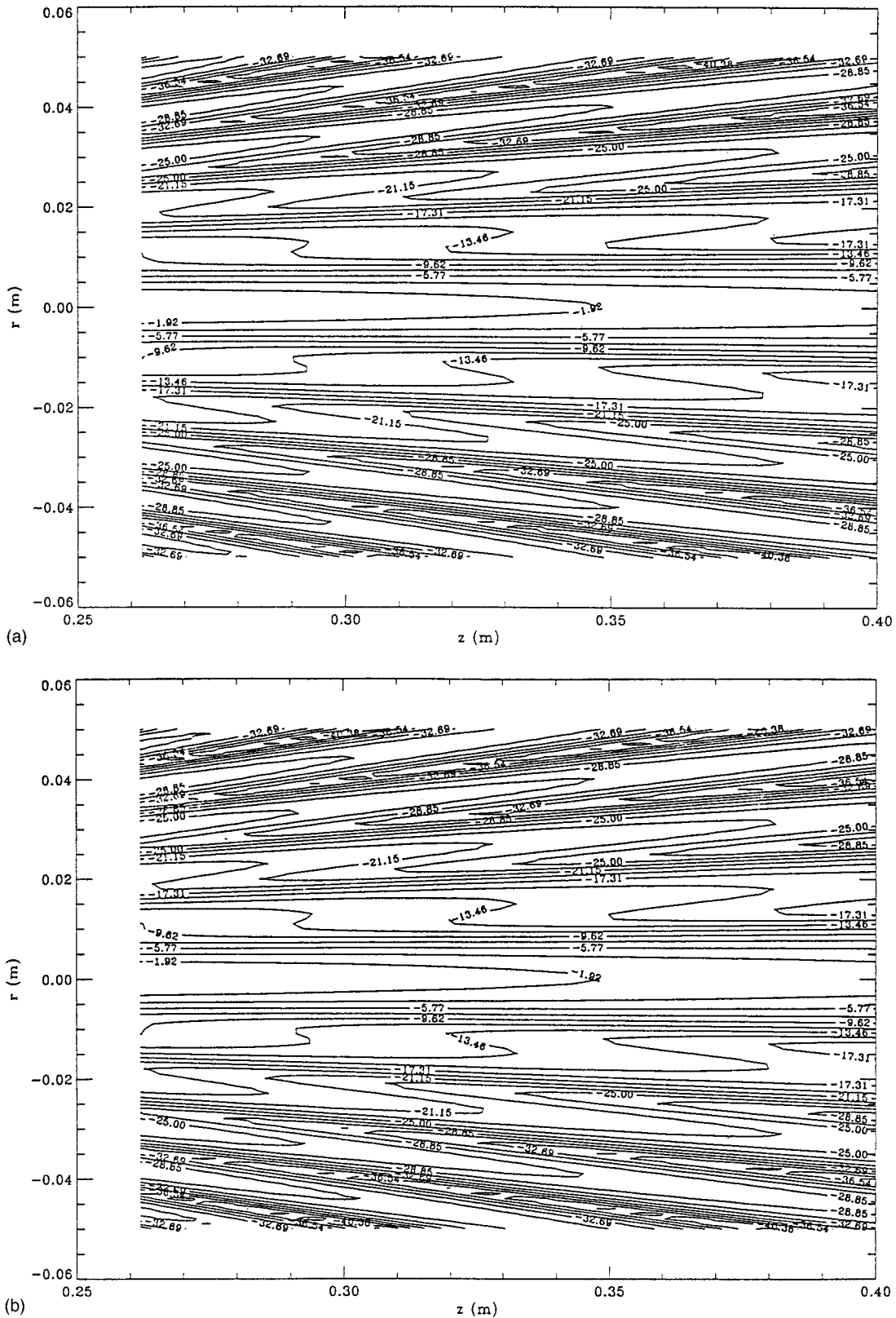


FIG. 2. Contour diagram of the field of a piston with 15 terms in the (a) analytic Gaussian representation; (b) Gaussian finite-element method. (Transmission loss normalized at the Fresnel distance.)

**1. Boundary conditions and numerical approach** and

Application of surface and irregular interface boundary conditions on pressure gives for  $u(r,z)$  in Eq. (2)

$$u_1|_{z=0}=0, \quad (16) \quad u_j|_{z=D_j(r)}=u_{j+1}|_{z=D_j(r)}, \text{ where } j=1,2,\dots,N-1. \quad (17)$$

The application of the boundary condition to the normal component of particle velocity gives

$$\left. \frac{1}{\rho_j} \frac{\partial p_j}{\partial \hat{n}} \right|_{z=D_j(r)} = \left. \frac{1}{\rho_{j+1}} \frac{\partial p_{j+1}}{\partial \hat{n}} \right|_{z=D_j(r)},$$

where  $j = 1, 2, \dots, N-1$ . (18)

Here  $\hat{n}$  is the unit vector along the normal direction at the interface. The finite-element approach<sup>22</sup> has been used, and the wide angle technique<sup>25</sup> is employed for the calculation.

## 2. Numerical example with low resolution

Let us consider propagation in deep-to-shallow water as shown in Fig. 3. This corresponds to the up-slope environment benchmark<sup>25</sup> case encountered in underwater acoustics. The region of propagation is bounded by a pressure release surface and an irregular bottom where the bottom remains flat at 350 m in depth for the first 40 m. At 40 m the bottom begins to slope upward to a range of 2040 m, then it remains flat at 50 m in depth. Propagation loss was calculated at a bottom slope angle of 8.5 degrees. The sound speed in the water is 1500 m/s; in the bottom it is 1600 m/s. In the bottom, a density of 1.5 g/cm<sup>3</sup> and an attenuation of 0.2 dB/wavelength are used. The source frequency is 25 Hz, and the source is placed at a depth of 25 m. A Gaussian source with Gaussian constant  $B = 4.39$  (half-beam waist 9.53 m) and a Piston distributed source are simulated. Both of them have a radius of 20 m. Contour diagrams of the sound fields for a Gaussian source, at grazing angle  $-12^\circ$ ,  $0^\circ$ , and  $+12^\circ$  are shown in Fig. 4 where for convenience the bottom is indicated by a line. The field below this line indicates bottom penetration. The sound pressure is reduced by transmission loss, and it is normalized at the reference distance 1 m from the source. The sound field for the Gaussian source is dependent on the grazing angle. However, the resolution is not sufficient to visualize the field in detail. The major differences between this example and the standard PE Gaussian starter<sup>25</sup> are the Gaussian constant  $B$  and the nonzero grazing angle.

Let us contrast the results of a single Gaussian source with the results of a 20-m diameter piston source approximated by 15 Gaussian beams under the conditions specified. The sound field for the piston at  $+12^\circ$  is shown in Fig. 5. The beam propagation is strongly dependent on grazing

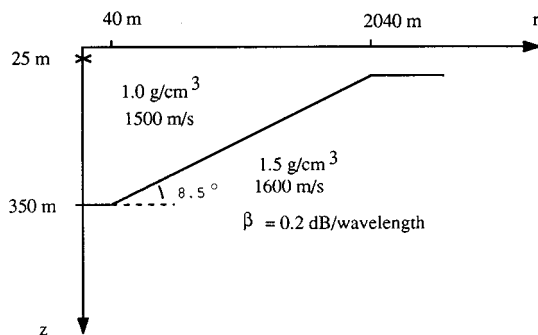


FIG. 3. Sloping bottom geometry.

angle also. Similarities exist when one compares corresponding fields, but similarities between the two sets of fields are not obvious under the low frequency condition (This means that the conditions are such that a distributed source appears almost like a point source.)

## 3. Numerical example with high resolution

If one changes the frequency from 25 Hz up to 250 Hz, and keeps the Gaussian radiator radius and beam waist the same, one can model increased resolution. Figure 6 shows a high resolution field when the Gaussian radiator is positioned at the same grazing angle of  $+12^\circ$ . The contour plot shows the slope of the bottom clearly. The beam fields are strongly dependent on the grazing angle.

The Gaussian finite-element method also can be applied to a piston source at 250 Hz for high resolution. Figure 7 shows the contour plot with a piston source simulated by a 15-term Gaussian expansion. The beam fields also are dependent on grazing angle. The slope of the bottom also can be seen clearly from these figures. Comparing the piston distributed beam fields with the single Gaussian beam fields, one can see sidelobes for the piston distributed source, whereas they are not observed with the Gaussian source. This comparison gives a ready demonstration of the value of the Gaussian beam in underwater acoustics. It also shows the value of the Gaussian finite-element method. More detailed simulation results are in Ref. 23.

## III. EXPERIMENTAL VERIFICATION

This section describes comparison of the Gaussian finite-element method results with experiment using both homogeneous and inhomogeneous media. The experiments are performed in a laboratory water tank. The inhomogeneous medium is created by floating oil on a water surface. More detailed test results are given in Ref. 23.

### A. Gaussian transducer with water-air interface

To examine a water-air interface, the 2.03-MHz 1-in. diameter Gaussian transducer<sup>24</sup> is used in the experiment. The transducer is 8 cm below the water surface. It is mounted on a frame such that the beam emitted by the Gaussian transducer has a grazing angle  $\varphi = -29.2$  degrees. The beam reaches the water-air surface, and then is reflected into the water as shown in Fig. 8. The beam pressure profile measured vertically at distances  $r = 26$  cm and 28 cm from the source are shown in Fig. 9. The dots are experimental results and the solid curve represents the solution made by the Gaussian finite-element method with  $B = 4.32$  (beam waist half-width 0.61 cm). The experimental data and the calculated results were normalized at the peak of the first curve; then the same normalization constant was used for successive curves. The starting profile is generated at a distance  $R_0 = 10$  cm. This means that the initial profile is evaluated before reflection and measurements are made after reflection. The effect of reflection is included in the Gaussian finite-element calculation. The measurements were performed by the pulse technique in a laboratory water tank

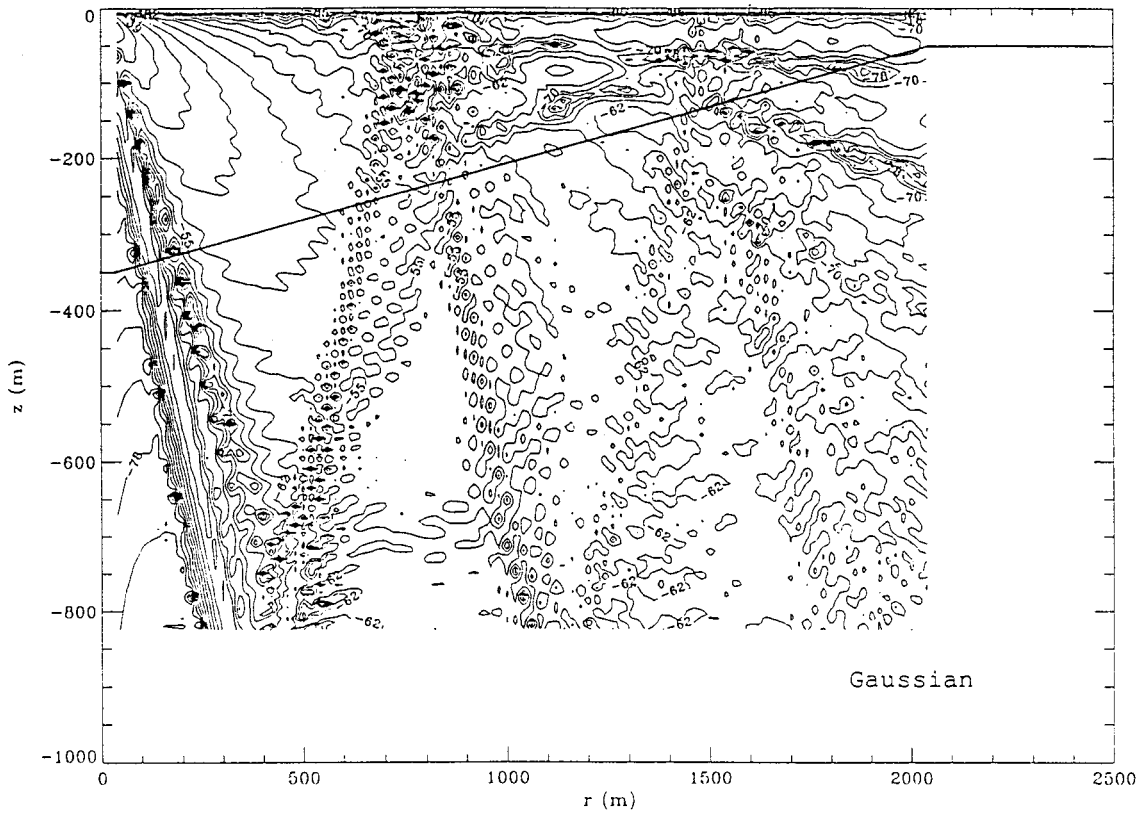


FIG. 4. Contour diagram of the field produced at a sloping interface by a 20-m diameter Gaussian transducer at 25 Hz (half-width of beam waist: 9.53 m at grazing angle of  $+12^\circ$ ).

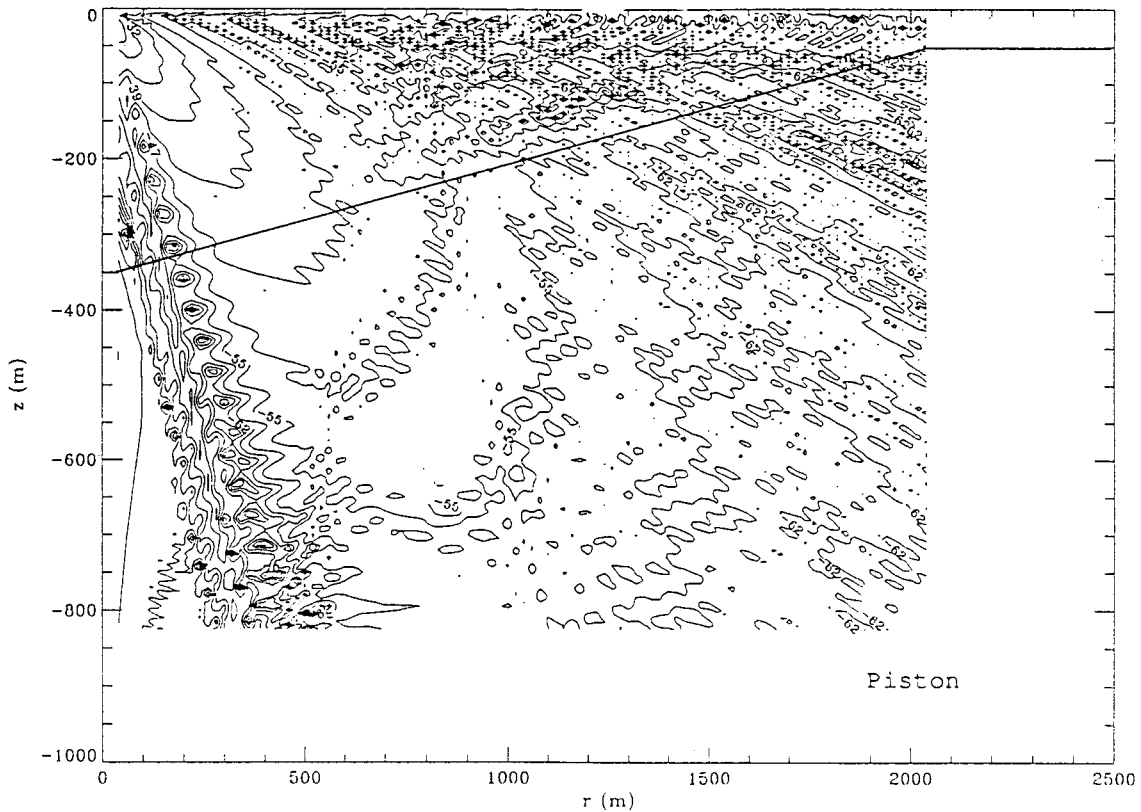


FIG. 5. Contour diagram of the field produced at a sloping interface by a 20-m diameter piston transducer at 25 Hz at grazing angle of  $+12^\circ$ .



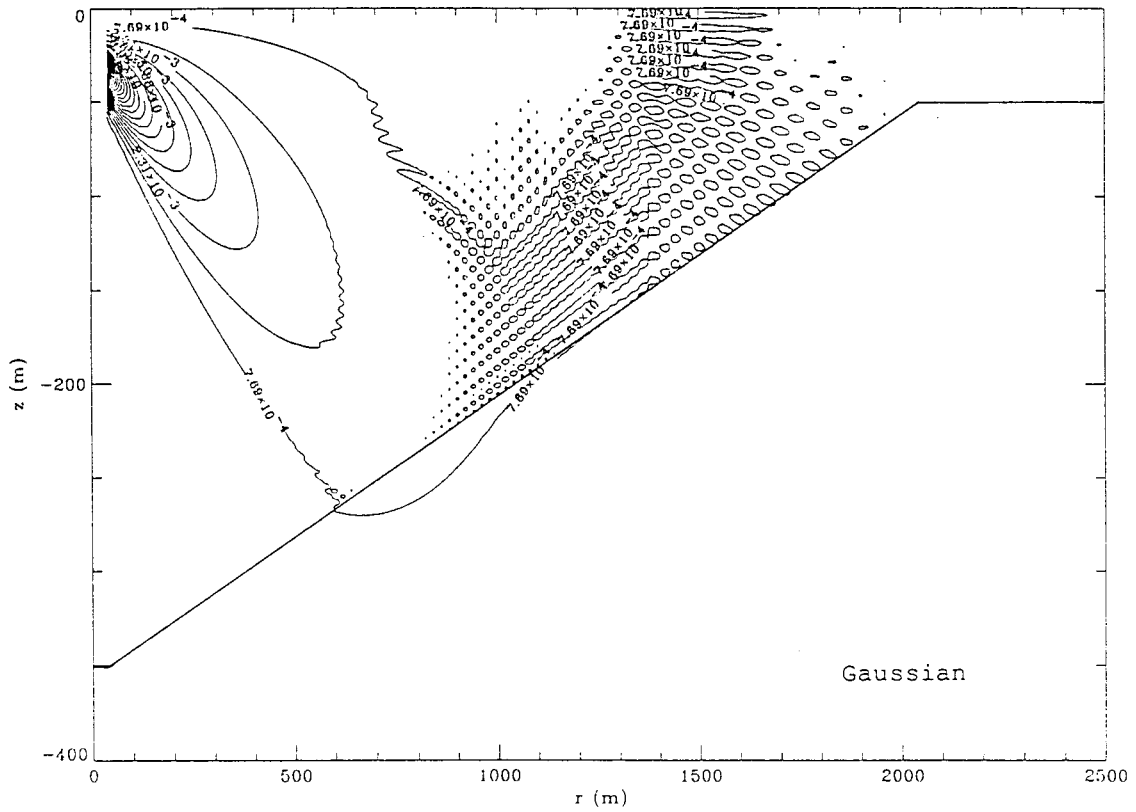


FIG. 6. High resolution plot of a single Gaussian beam at 250 Hz at a sloping interface (compare with Fig. 4) at grazing angle of  $+12^\circ$ .

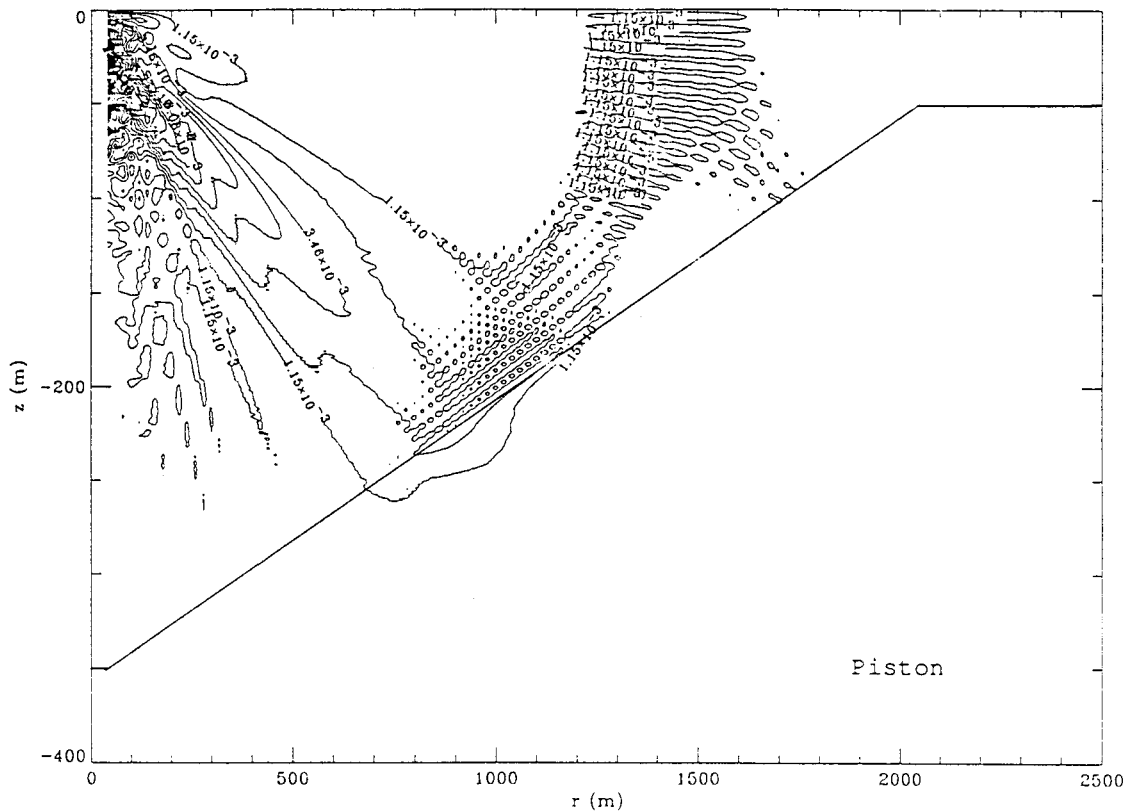


FIG. 7. High resolution plot of the field of a piston transducer (15 term Gaussian series) at 250 Hz at grazing angle of  $+12^\circ$ . (Compare with Fig. 5.)

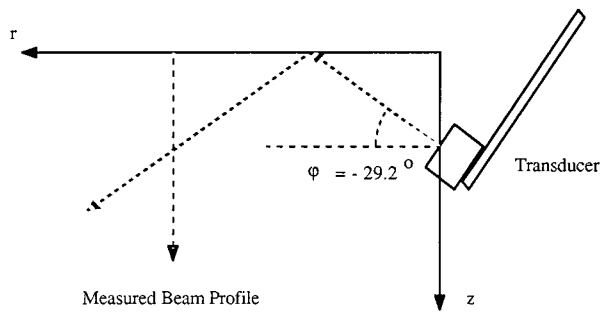


FIG. 8. Experimental configuration for measurement in water.

0.26 m wide, 0.30 m high, and 0.60 m long. The field distribution was measured with a hydrophone (made by MEDSONICS) which has a nominal resolution of 1 mm. The hydrophone is mounted on a three-dimensional scanning system to measure the relative sound field amplitude point-by-point in a water tank. The Gaussian transducer is driven by an ENI (model 350 L) rf amplifier with an input signal from an HP 8116A pulse generator. The output of the probe is amplified by a MEDISONICS preamplifier with +30-dB gain, and then goes to a LeCroy 9400A digital oscilloscope. If we fix the transducer and scan the field by moving the probe at a certain distance from the transducer, we measure

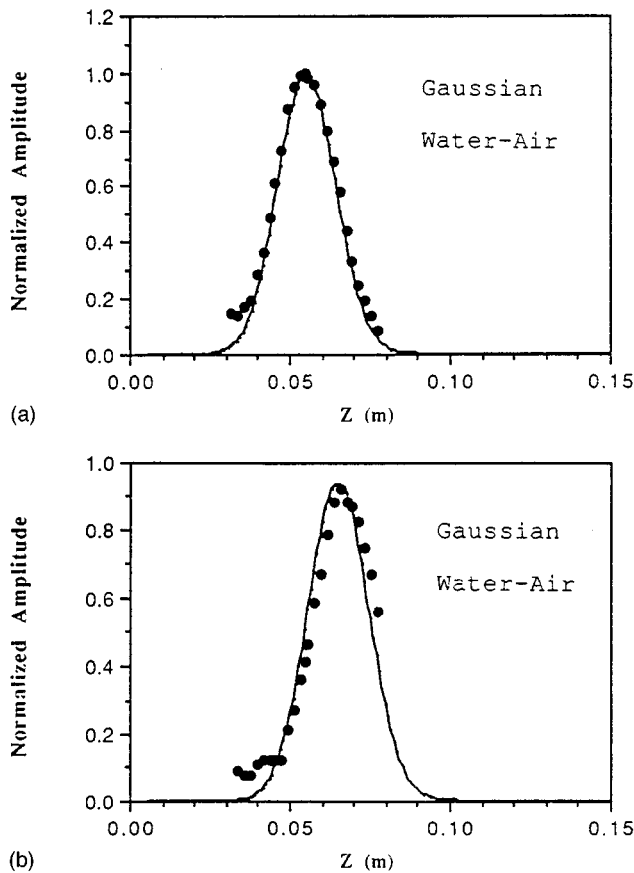


FIG. 9. Profile of the beam from a 2.03-MHz Gaussian transducer ( $B = 4.32$ ; half-width of beam waist 0.61 cm) after reflection from a water-air interface as shown in Fig. 8. (a) 26 cm; (b) 28 cm. Dots are measured results. Curves are Gaussian finite-element results.

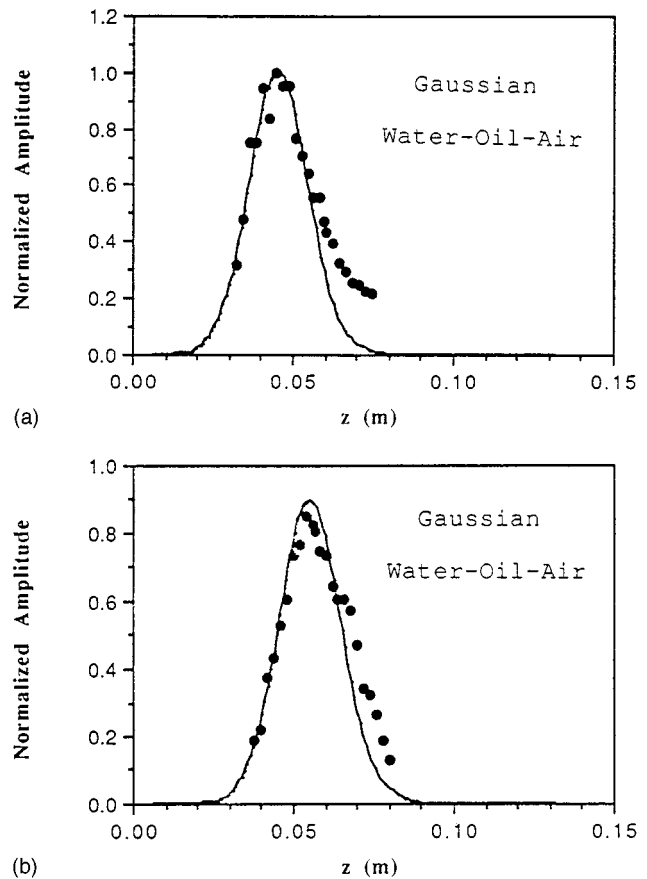
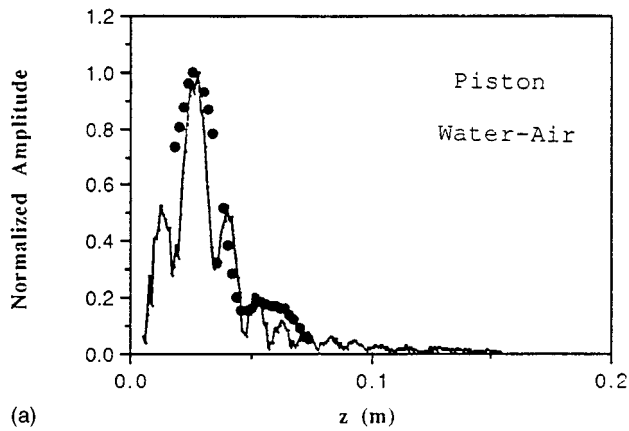


FIG. 10. Profile of 2.03-MHz Gaussian beam reflected from water-oil layer-air interfaces ( $B = 4.32$ ; half-width of beam waist 0.61 cm). (a) 26 cm; (b) 28 cm. Dots are measured results. Curves are Gaussian finite-element results.

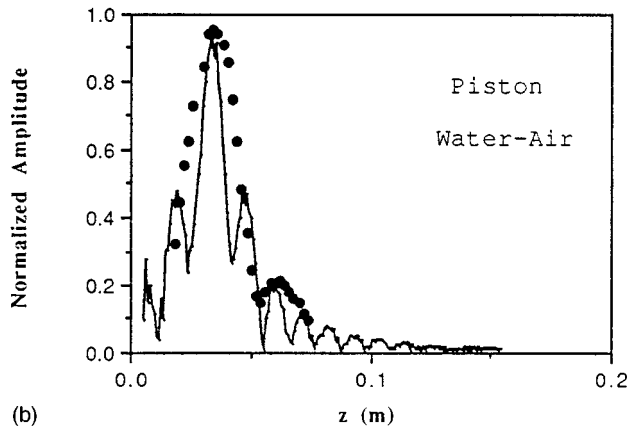
the sound pressure distribution across the beam. With a Gaussian coefficient  $B = 4.32$  ( $w = 0.61$  cm), the pressure profile is predicted at a distance from the transducer. There is a very good agreement between the prediction and experiments in the center and at the right edge. The left edge agrees less well. The possible measurement errors for this (or other experimental examples) could be the accuracy of the grazing angle measured and the precision of the 3-D mechanical system used.

### B. Gaussian transducer with oil layer on water

To create an inhomogeneous environment in a laboratory water tank, we introduce an oil layer. The oil used in the experiment is "SAE 30" nondetergent motor oil of 0.8 cm thickness. The transducer is mounted 8 cm below the water surface on the same frame such that the beam is emitted with a grazing angle  $\phi = -29.2$  degrees (see Fig. 8). The beam reaches the oil-air surface, then is reflected into the water through the oil layer. The experimental results measured vertically at distances of 26 cm and 28 cm from the source are shown in Fig. 10. All measurements were made in the water, below the oil layer. The test also shows no sidelobes. We see the agreement is good at the center and the left edge of the profile, but less good on the right edge. No observable sidelobes are found.



(a)



(b)

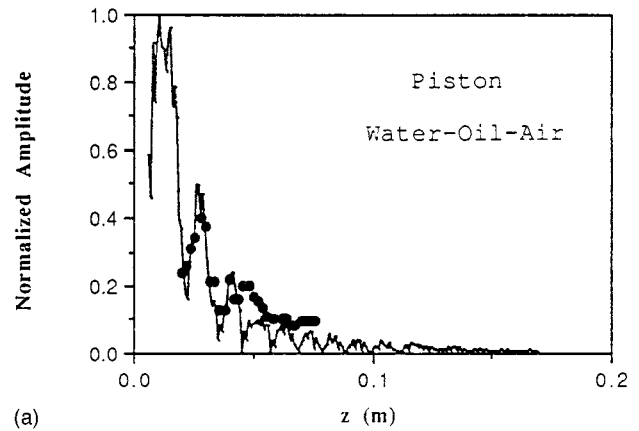
FIG. 11. Profile of beam from a piston transducer reflected from a water-air interface (a) 32 cm; (b) 34 cm. Dots are measured results. Curves are Gaussian finite-element results.

### C. Piston transducer with water-air interface

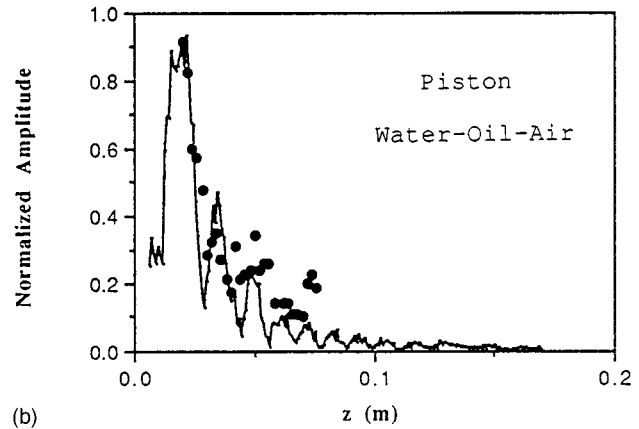
To see the effect of substituting a piston source, a 1.98-MHz 1-in. piston transducer is used in the water-air experiment. Here, the grazing angle  $\varphi$  is  $-19$  degrees (see Fig. 8). The piston transducer is 8 cm below the water surface. The same experimental device as mentioned in the previous section was used, and the hydrophone scanned the sound pressure profile vertically at distances of 32 cm and 34 cm from the piston source. The Gaussian finite-element starting profile is generated by Eq. (7), and  $R_0$  is 21 cm. In this case all measurements were made in the beam after reflection also. The experimental results are presented in Fig. 11. For comparison, numerical calculations with 15 terms in the Gaussian finite-element method are shown as a solid curve. The general agreement between the measured and calculated results is good. Some small sidelobes of the beam are hard to measure because of the resolution of the probe. We successfully measured the first sidelobe at 32 cm, however, and find agreement between experiment and calculation.

### D. Piston transducer with oil layer on water

To produce an inhomogeneous layer, a layer of oil (the same type as used in single Gaussian beam experiment) is gently poured into the water tank. The oil layer is 0.85 cm thick. The experimental results compared with the numerical simulation results at distances of 32 cm and 34 cm are shown



(a)



(b)

FIG. 12. Profile of beam from a piston transducer reflected from a water-oil layer-air interface. (a) 32 cm; (b) 34 cm. Dots are measured results. Curves are Gaussian finite-element results.

in Fig. 12. This time, the peak of the reflected beam shifted so much that only part of the beam could be covered by the three-dimensional scanning system where the hydrophone is mounted. At 32 cm the beam had not spread enough to be within the range of measurement. Only sidelobes could be measured. At 34 cm the beam was wide enough that we could measure half of the beam in addition to the sidelobes. The agreement between experiment and calculation is considered satisfactory.

## IV. SUMMARY AND CONCLUSIONS

A practical way to mathematically describe the sound field produced by a distributed source in inhomogeneous media is presented in this paper. The approach combines Gaussian beam superposition with the finite-element solution of the parabolic equation. With it one can describe an arbitrary sound field as it propagates in a medium in which both sound velocity and density can be functions of depth. The 3-D conical beam with a 2-D slice is demonstrated here, although the general concept is helpful in application of the 3-D parabolic equation. It is helpful for other numerical models and underwater communication as well, where directional beams often are encountered.

First, the background theory is described, then numerical simulation and computer modeling are presented. Under certain conditions, Gaussian beam expansion becomes similar

to a Fourier series expansion. The difference is that the Gaussian beam expansion is made with a complex argument. The theoretical study shows that the orthogonality relation that is so important to application of Fourier series to boundary value problems cannot be used to evaluate Gaussian coefficients.<sup>23</sup> A computer optimization procedure is used instead.

The Gaussian theory describes wave propagation in a homogeneous medium. The Gaussian beam optimization procedure effectively deals with a distributed source. Since the finite-element method also can describe inhomogeneous media, the two approaches are combined to take advantage of the strengths of each. This leads to the new Gaussian finite-element method.

The new method is verified experimentally by an improved Gaussian transducer<sup>24</sup> that produces a single ultrasonic Gaussian beam—the basic unit of the Gaussian beam representation. As a distributed source, a piston is employed in the experiment. Whether for a Gaussian source or a distributed piston source, predictions of the new method agree very well with the experimental results.

The most direct application of the results would be in describing an underwater sound field, although the general problem of beam diffraction in an inhomogeneous medium is of fundamental interest.

## ACKNOWLEDGMENT

The authors appreciate the financial support from the U.S. Office of Naval Technology for the Gaussian transducer development program.

<sup>1</sup>J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," *J. Acoust. Soc. Am.* **83**, 1752–1756 (1988).

<sup>2</sup>L. Rayleigh, *Theory of Sound* (Dover, New York, 1945), Vol. II, Sec. 278.

<sup>3</sup>G. R. Harris, "Transient field of a baffled planar piston having an arbitrary vibration amplitude distribution," *J. Acoust. Soc. Am.* **70**, 186–204 (1981).

<sup>4</sup>E. Cavanagh and B. D. Cook, "Gaussian-Laguerre description of ultrasonic fields-numerical sample: circular piston," *J. Acoust. Soc. Am.* **67**, 1136–1140 (1980).

<sup>5</sup>J. N. Tjøtta and S. Tjøtta, "An analytical model for the nearfield of a baffled piston transducer," *J. Acoust. Soc. Am.* **68**, 334–339 (1980).

<sup>6</sup>T. Hasegawa, N. Inoue, and K. Matsuzawa, "A new rigorous expansion

for the velocity potential of a circular piston source," *J. Acoust. Soc. Am.* **74**, 1044–1047 (1983).

<sup>7</sup>M. Greenspan, "Piston radiator: some extensions of the theory," *J. Acoust. Soc. Am.* **65**, 608–621 (1979).

<sup>8</sup>B. K. Novikov, O. V. Rudenko, and T. I. Timoshenko, *Nonlinear Underwater Acoustics* (American Institute of Physics, New York, 1987).

<sup>9</sup>J. N. Tjøtta and S. Tjøtta, "Nearfield and farfield of pulsed acoustic radiators," *J. Acoust. Soc. Am.* **71**, 824–833 (1982).

<sup>10</sup>G. Du and M. A. Breazeale, "The ultrasonic field of a Gaussian transducer," *J. Acoust. Soc. Am.* **78**, 2083–2086 (1985).

<sup>11</sup>G. Du and M. A. Breazeale, "Harmonic distortion of a finite amplitude Gaussian beam in fluid," *J. Acoust. Soc. Am.* **80**, 212–216 (1986).

<sup>12</sup>G. Du and M. A. Breazeale, "Theoretical description of a focused Gaussian ultrasonic beam in a nonlinear medium," *J. Acoust. Soc. Am.* **81**, 51–57 (1987).

<sup>13</sup>J. J. Wen and M. A. Breazeale, "Computer optimization of the Gaussian beam description of an ultrasonic field," in *Computational Acoustics*, Vol. 2, edited by D. Lee, A. Cakmak, and R. Vichnevetsky (Elsevier Science, Amsterdam, 1990).

<sup>14</sup>V. A. Fock, *Electromagnetic Diffraction and Propagation Problem* (Pergamon, New York, 1965).

<sup>15</sup>F. D. Tappert, "The parabolic approximation method," in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis, Lecture Notes in Physics, Vol. 70 (Springer, Heidelberg, 1977).

<sup>16</sup>F. R. DiNapoli and R. L. Deavenport, in *Topics in Current Physics*, edited by J. A. DeSanto (Springer, Heidelberg, 1979).

<sup>17</sup>Min Gong, X. Jinquan, W. Mengxin, W. Yingeng, and D. Huang, "An experimental investigation of turning-point convergence-zone in a deep sound channel in the South China Sea," *Chinese Acta Acoust.* **12**, 417–423 (1987).

<sup>18</sup>Ding Lee and J. S. Papadakis, "Numerical solutions of the parabolic wave equation: an ordinary-differential equation approach," *J. Acoust. Soc. Am.* **68**, 1482–1488 (1980).

<sup>19</sup>D. Lee, G. Botseas, and J. S. Papadakis, "Finite-difference solution to the parabolic wave equation," *J. Acoust. Soc. Am.* **70**, 795–800 (1981).

<sup>20</sup>S. T. McDaniel and D. Lee, "A finite-difference treatment of interface conditions for the parabolic wave equation: the horizontal interface," *J. Acoust. Soc. Am.* **71**, 855–858 (1982).

<sup>21</sup>Ding Lee and S. T. McDaniel, "A finite-difference treatment of interface conditions for the parabolic wave equation: the irregular interface," *J. Acoust. Soc. Am.* **73**, 1441–1447 (1983).

<sup>22</sup>Dehua Huang, "Finite element solution to the parabolic wave equation," *J. Acoust. Soc. Am.* **84**, 1405–1413 (1988).

<sup>23</sup>Dehua Huang, "Gaussian finite element method for description of underwater and sound diffraction," Ph.D. Dissertation, University of Mississippi (1992).

<sup>24</sup>Dehua Huang and M. A. Breazeale, in *IEEE Ultrasonics Symposium*, Lake Buena Vista, Florida, Vol. 1, p. 473, edited by B. R. McAvoy (IEEE, New York, 1991).

<sup>25</sup>Ding Lee, G. Botseas, and K. E. Gilbert, "IFD: Wide-Angle Capability," NUSC Technical Report 6905 (1983).



# Nonlinear propagation of narrow-band Rayleigh waves excited by a comb transducer

D. C. Hurley

National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80303

(Received 15 September 1998; revised 23 June 1999; accepted 25 June 1999)

Quantitative experiments have been performed to measure the harmonic generation of narrow-band Rayleigh waves in aluminum. Based on a comparison of several methods, the Sokolinskii comb transducer was selected as an excitation source. Results are presented for narrow-band Rayleigh waves (center frequency 9.85 MHz) generated by the comb and detected with a path-stabilized Michelson interferometer. The on-axis, out-of-plane displacement amplitude was measured as a function of distance from the source. Maximum out-of-plane displacements greater than 7 nm were observed. An existing quasilinear theory that included both diffraction and attenuation was used to interpret the data. Values of the nonlinearity parameter  $\beta_{11} = 0.21\text{--}0.23$  predicted displacements in good agreement with the data and were consistent with values for  $\beta_{11}$  calculated from literature values of the third-order elastic moduli of aluminum. [S0001-4966(99)03510-9]

PACS numbers: 43.35.Cg [HEB]

## INTRODUCTION

This paper describes experiments to measure the nonlinear propagation of finite-amplitude, narrow-band Rayleigh waves in aluminum using a comb transducer as the excitation source. The work is motivated by the desire to use nonlinear Rayleigh waves to characterize the mechanical and microstructural properties of surfaces. However, as discussed below, experimental progress is hindered by lack of an effective narrow-band Rayleigh-wave source for nonpiezoelectric substrates. Before nonlinear Rayleigh-wave techniques can be used for microstructural characterization, practical narrow-band sources must be developed. In addition, more quantitative measurements of nonlinear surface-wave propagation are needed for comparison to theoretical models. The experiments reported here represent a first step toward achieving these objectives.

Interest in nonlinear ultrasonics as a tool for materials characterization has grown markedly in recent years. One experimental technique for examining nonlinear effects is second harmonic generation of longitudinal waves. With this method, narrow-band waves at the fundamental angular frequency  $\omega_0$  are excited. The wave displacements at  $\omega_0$  as well as the second harmonic frequency  $2\omega_0$  are detected as the wave propagates away from the source and the relative amplitudes analyzed. Observation of second harmonic generation typically requires "finite-amplitude" waves, that is, waves with displacement amplitudes much larger (approximately 1–10 nm at 5–30 MHz) than those used in conventional, "linear" ultrasonic applications. Harmonic generation has been used to investigate relations between higher-order elastic properties and such microstructural features as fatigue, dislocations, thermal embrittlement, and microadhesion.<sup>1–4</sup>

Progress in understanding *bulk* properties leads naturally to the idea that harmonic generation of Rayleigh waves could provide valuable information about *surface* properties. Yet harmonic generation of Rayleigh waves to investigate higher-order elastic properties of surfaces remains largely

unexplored. Moreover, experimental work in this area lags far behind development of corresponding theoretical models. Why?

Historically, investigations have concentrated on applications involving surface-acoustic-wave devices, in which the excitation source was an interdigital transducer (IDT). Extensive studies of harmonic generation and nonlinear Rayleigh-wave propagation<sup>5–8</sup> have been made with IDTs on piezoelectric single crystals. Although an IDT can excite Rayleigh waves efficiently, its operation requires a piezoelectric substrate. Hence, its usefulness for materials characterization is severely limited.

In contrast, only a few experiments on surface-wave harmonic generation in nonpiezoelectric materials have been reported. Rischbieter<sup>9,10</sup> qualitatively examined the dependence of second-harmonic generation on applied stress in aluminum and steel. He used a wedge to convert longitudinal waves to Rayleigh waves and had to filter out spurious harmonics generated by the source using a series of grooves machined in the surface. Brysev *et al.*<sup>11</sup> demonstrated harmonic generation on a glass surface; they did not interpret the results in detail. Meegan<sup>12</sup> recently performed Rayleigh-wave experiments in aluminum and copper using comb sources, but was unable to observe harmonic generation unambiguously. Clearly, further work is needed to develop an effective, practical source and to evaluate quantitatively the Rayleigh waves it excites.

In this paper, experiments to investigate harmonic generation of Rayleigh waves in aluminum are reported. First, experimental techniques for the excitation of narrow-band, finite-amplitude Rayleigh waves and detection of broadband nonlinear Rayleigh waves are described. Next, measurements of the absolute out-of-plane components of the fundamental and second-harmonic displacements as a function of propagation distance are presented. The final section summarizes an existing quasilinear theory and uses it to predict the

measured displacements. Values of the Rayleigh-wave non-linearity parameter needed to obtain good agreement between theory and experiment are compared to literature values for the third-order elastic moduli of aluminum.

## I. EXPERIMENTAL TECHNIQUES

### A. Comb excitation of finite-amplitude Rayleigh waves

Given the lack of experimental literature in this area, the choice of technique to excite Rayleigh waves was not obvious. Several different methods for exciting narrow-band Rayleigh waves were evaluated experimentally under similar conditions; the results are summarized in the Appendix. Based on this comparison, a Sokolinskii comb device<sup>13</sup> was selected for harmonic-generation investigations. The Sokolinskii comb consists of a plate with alternating teeth and grooves, in which the groove spacing has been chosen equal to the Rayleigh wavelength in the substrate. A transducer is attached to the comb's flat side, while the other (tooth) side is bonded to the specimen. Exciting the transducer creates a surface wave in the specimen.

The comb used in the experiments described below was designed to operate at approximately 10 MHz. It was a thin aluminum disk with radius  $a_0 = 0.318 \pm 0.003$  cm. The comb was fabricated using electrical discharge machining techniques. It contained 21 grooves with a nominal center-to-center spacing of 300  $\mu\text{m}$  and 100  $\mu\text{m}$  deep. A single-crystal, compressional-mode  $\text{LiNbO}_3$  transducer of the same size was bonded to the comb. High-power, low-distortion toneburst electronics (25 cycles at 9.85 MHz, approximately 600 V peak-to-peak) were used to excite the transducer.

### B. Interferometric detection of nonlinear Rayleigh waves

The important characteristics of a detector for nonlinear Rayleigh waves differ from those of the source. Most significantly, the frequency response of the detector must be as broad band as possible, in order to detect wave components from  $\omega_0$  to  $2\omega_0$  or higher. Furthermore, it must be possible to calibrate the detector over the entire frequency range. That is, it must be possible to calculate the absolute ultrasonic displacement or velocity from the detector output. Additionally, the detector should be much smaller than the Rayleigh wavelength  $\lambda_R$ ; this means that for  $f_0 = \omega_0/2\pi = 10$  MHz ( $\lambda_R \sim 300 \mu\text{m}$ ), detector dimensions of 100  $\mu\text{m}$  or less are needed. Ideally, it should be possible to scan the detector to examine the spatial behavior of the waves. Furthermore, the detector should load the surface as little as possible in order to minimize its effect on wave propagation.

To meet these requirements, a path-stabilized Michelson interferometer incorporating an infrared laser was used to detect the ultrasonic waves. This apparatus has been described in detail previously,<sup>14</sup> so only its main features are mentioned here. The interferometer's frequency response was virtually flat from nearly dc to over 21 MHz (and can be extended to higher frequencies if needed). The interferometric technique directly measured the absolute out-of-plane ultrasonic displacement, using a straightforward calibration

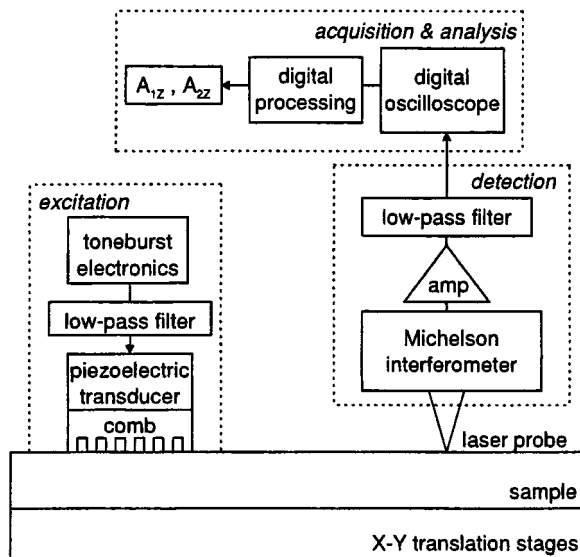


FIG. 1. Block diagram of apparatus for Rayleigh-wave harmonic-generation experiments. Key elements of each step in the experimental process (excitation, detection, and analysis) are indicated.

that relates the detector voltage to the well-known laser wavelength. A spatial resolution of approximately 10  $\mu\text{m}$  was achieved by focusing the laser beam onto the sample surface. The use of motorized translation stages provided two-dimensional scanning capabilities, so that the position of the laser probe beam relative to that of the source could be varied. Finally, this optical technique was noncontacting and thus did not load the surface.

Figure 1 contains a schematic diagram of the experimental apparatus. The figure indicates the main elements of the excitation and detection processes. It also shows the steps required to obtain the fundamental and second-harmonic out-of-plane displacement components  $A_{1z}$  and  $A_{2z}$ , respectively. First, an oscilloscope converted the time-varying voltage detected by the interferometer into a digital waveform. The waveform contained the detected signals at both  $\omega_0$  and  $2\omega_0$ . In order to calculate the displacement amplitudes  $A_{1z}$  and  $A_{2z}$  separately, the waveform was converted to the frequency domain and multiplied by simple notch filters centered at  $\omega_0$  and  $2\omega_0$ . The two bandpass-filtered signals obtained in this way were then converted back into the temporal domain, obtaining the unscaled out-of-plane displacements at  $\omega_0$  and  $2\omega_0$ . Finally, these frequency-filtered signals were scaled using information obtained in calibration measurements.<sup>14</sup> The absolute values for  $A_{1z}$  and  $A_{2z}$  were calculated over the central flat portion of the signal, where the amplitude reached a steady-state value.

## II. EXPERIMENTAL RESULTS

With the 10 MHz comb described above, harmonic-generation experiments were performed on a 2.5 cm thick plate of 6061-T651 aluminum alloy. One surface of the plate was polished by hand, ending with a mild chemical etch. This process produced a near-optical surface finish (estimated surface roughness  $< 2 \mu\text{m}$ ) which improved the reflected laser signal. The comb/transducer assembly was glued to the polished surface. In the experiments, the out-of-

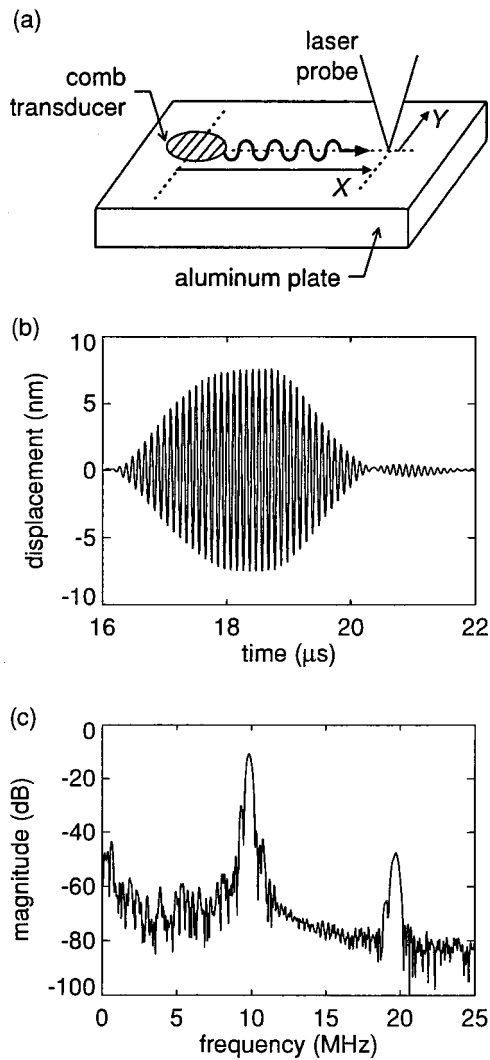


FIG. 2. (a) Schematic of the experimental geometry. (b) Ultrasonic waveform corresponding to the out-of-plane displacement of Rayleigh waves in aluminum, excited by a comb and detected with a Michelson interferometer. The laser probe beam was positioned at the point  $X=6.2$  cm,  $Y=0.0$  cm. (c) Corresponding frequency spectrum.

plane Rayleigh-wave displacement was measured at various positions of the laser probe beam relative to the fixed source. The source–detector position was specified by the coordinates  $X$  and  $Y$ . The axial coordinate  $X$  was the distance relative to the comb center, perpendicular to the long direction of the comb grooves. The transverse coordinate  $Y$  represented the distance relative to the comb center in the groove direction. The experimental uncertainty in  $X$  and  $Y$  was  $\pm 0.2$  cm for the absolute distance from the center of the comb, but only  $\pm 0.005$  cm for a series of relative measurements. The coordinate system and experimental geometry are depicted schematically in Fig. 2(a).

An example of the comb's performance is shown in Fig. 2(b). This plot shows the out-of-plane Rayleigh-wave displacement measured at the position  $X=6.2$  cm,  $Y=0.0$  cm. The plot reveals that the wave increased in amplitude for roughly the first third of its duration, reached a steady-state level during the middle third, and then decreased in the final third. If the number of excitation cycles or the number of

teeth were increased, the duration of the central flat portion would increase. (A detailed analysis of the buildup and decay of the waves as various portions were excited and attenuated by individual teeth is beyond the scope of this paper.) The displacement amplitudes mentioned below were calculated using several wave cycles in the central flat portion.

Figure 2(b) indicates that the maximum wave amplitude was larger than 7 nm—about the same as those used in bulk-wave harmonic-generation experiments.<sup>2</sup> (The overall displacement magnitude in a particular experiment depended on how well the comb had been bonded to the substrate. Out-of-plane displacements as large as 32 nm were obtained in some cases, but the adhesive bond could not be reproduced reliably. Efforts are being made to improve adhesion quality and reproducibility.) The corresponding frequency spectrum is shown in Fig. 2(c). The frequency content above  $\sim 21$  MHz is suppressed by a low-pass filter on the interferometer output. A second-harmonic component is clearly evident.

Further experiments were performed to investigate the spatial dependence of second-harmonic generation. In these experiments, the magnitude of  $A_{1z}$  and  $A_{2z}$  as a function of distance from the comb (that is, as a function of  $X$ ) was measured along the axis of the comb ( $Y=0$ ). Behavior at very small values of  $X$  (very near field) was not investigated. Here, it was likely that effects due to the nonideal (nonuniform) nature of a real comb would be much more pronounced. Off-axis ( $Y \neq 0$ ) behavior was not examined for similar reasons. At each axial position, a transverse scan in  $Y$  was performed. This process was necessary because the  $Y=0$  position did not always correspond to the position of maximum displacement, especially for small  $X$ . For each value of  $X$ , the transverse scan of  $A_{1z}$  was obtained and the  $Y=0$  position chosen based on symmetry. The squares in Fig. 3(a) and (b) represent the measured values of  $A_{1z}(X)$  and  $A_{2z}(X)$ , respectively. In these graphs, error bars for the amplitudes of  $A_{1z}$  and  $A_{2z}$  reflect instrumentation uncertainty only and do not include potential error due to uncertainty in the position of  $Y=0$ . The instrumentation uncertainty, which was primarily due to the uncertainty of the oscilloscope voltages,<sup>14</sup> was  $\pm 3\%$ .

Figure 3(a) shows that  $A_{1z}$  first increased and then decreased with increasing  $X$ . Such behavior indicates the influence of significant diffraction. This is not surprising, because the transducer radius  $a_0=0.318 \pm 0.003$  cm was only about ten times larger than the Rayleigh wavelength  $\lambda_R=298 \pm 3$   $\mu$ m. Increasing the transducer diameter would lessen the effects, but they cannot be eliminated totally for a practical-sized device. For  $X \gg x_0 = \frac{1}{2}\omega_0 a^2 / v_R$ , an asymptotic regime could be reached in which data interpretation would be simpler.<sup>15</sup> For the experiments described here,  $x_0 \approx 0.1$  m. Therefore, distances of  $X > 1$  m are necessary for an asymptotic expression to approximate the diffraction effects within a few percent. Such distances are usually impractical. Also, the cumulative effect of attenuation for such distances would significantly degrade the signal.

To correctly interpret the second-harmonic behavior in Fig. 3(b), therefore, it is necessary to consider diffraction effects in both the fundamental and second-harmonic components. This is discussed in the next section. However, two

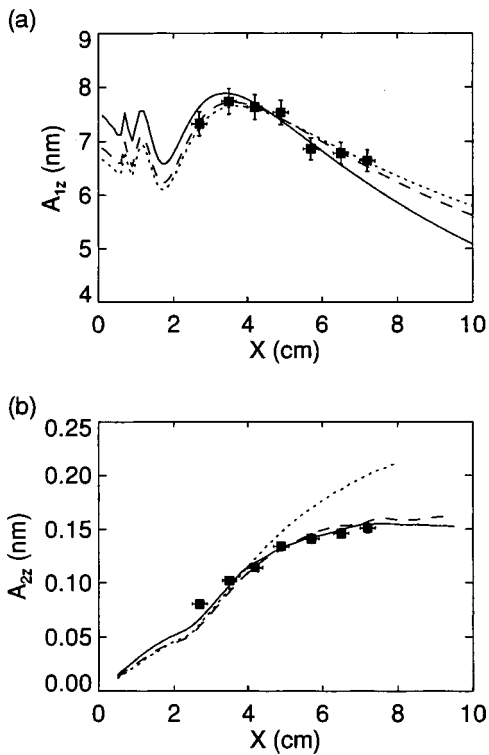


FIG. 3. Experimental data points and theoretical curves for the on-axis ( $Y = 0$ ) displacements generated in aluminum as a function of distance  $X$  from the center of the comb. (a) Out-of-plane fundamental displacement  $A_{1z}$ . The theoretical curves were calculated with  $v_0 = 0.95$  m/s and  $\alpha_1 = 0$  (dotted line);  $v_0 = 0.98$  m/s and  $\alpha_1 = 0.63$  Np/m (dashed line);  $v_0 = 1.07$  m/s and  $\alpha_1 = 2.50$  Np/m (solid line). (b) Corresponding out-of-plane second-harmonic displacement  $A_{2z}$ . The curves for  $A_{2z}$  were calculated using  $v_0 = 0.95$  m/s,  $\alpha_1 = \alpha_2 = 0$ , and  $\beta_{11} = 0.21$  (dotted line);  $v_0 = 0.97$  m/s,  $\alpha_1 = 0.63$  Np/m,  $\alpha_2 = 10.00$  Np/m, and  $\beta_{11} = 0.23$  (dashed line);  $v_0 = 1.07$  m/s,  $\alpha_1 = 2.50$  Np/m,  $\alpha_2 = 10.00$  Np/m, and  $\beta_{11} = 0.22$  (solid line).

immediate conclusions can be drawn from Fig. 3(b). First, the graph reveals that the comb excited Rayleigh waves large enough to produce measurable second-harmonic components. This in itself is significant: it was not known whether the comb (or indeed any method) would work in practice on nonpiezoelectric materials. Second, in spite of the decrease in  $A_{1z}$  by diffraction, Fig. 3(b) shows that  $A_{2z}$  increased with increasing  $X$ . This behavior reinforces physical intuition about the nature of the waves excited by the comb. For plane waves with zero attenuation, a linear increase in  $A_{2z}$  with  $x$  is required by quadratic nonlinearity.<sup>16</sup> In spite of diffraction, the comb-excited wave was expected to behave in reasonably consistent manner. This behavior also suggests that the second harmonics were generated as the wave propagated along the aluminum surface and not at the source.

### III. INTERPRETATION OF RESULTS

#### A. Model for diffraction and harmonic generation

The experimental results can be interpreted using an existing quasilinear theory<sup>15</sup> for Rayleigh-wave harmonic generation that includes diffraction effects. The theory develops expressions for the out-of-plane displacements excited by a various two-dimensional sources. Only the relevant results are summarized briefly here; for a full discussion see Ref. 15.

The on-axis solution for the magnitude of the fundamental component  $A_{1z}$  generated by a uniform line source of length  $2a$  as a function of perpendicular distance  $x$  from the line is given by

$$|A_{1z}(x, y = 0)| = \frac{v_0(1 + \xi_L \eta)}{\omega_0} e^{-\alpha_1 x} \left| \operatorname{erf} \sqrt{\frac{x}{ix_0}} \right|. \quad (3.1)$$

Here,  $v_0$  is the particle velocity,  $x_0 = \frac{1}{2} \omega_0 a^2 / v_R$  is the Rayleigh distance,  $v_R$  is the Rayleigh-wave velocity, and  $\alpha_1$  is the attenuation coefficient at  $\omega_0$ . The factor  $(1 + \xi_L \eta)$  comprises a combination of the longitudinal, transverse, and Rayleigh velocities for the material.<sup>17</sup> The on-axis expression for the magnitude of the second-harmonic component  $A_{2z}$  may be written as

$$|A_{2z}(x, y = 0)| = \frac{\beta_{11} v_0^2 (1 + \xi_L \eta) x_0}{8 v_R^2 \sqrt{\pi}} |\mathcal{I}(x, \alpha_1, \alpha_2, x_0)|, \quad (3.2)$$

where  $\mathcal{I}$  represents a Green's function integral<sup>15</sup> and depends on the position  $x$ , the Rayleigh distance  $x_0$ , and the attenuation coefficients  $\alpha_1$  and  $\alpha_2$  at  $\omega_0$  and  $2\omega_0$ . The dimensionless coefficient  $\beta_{11}$  is defined elsewhere.<sup>15,17</sup> The value of  $\beta_{11}$  is determined solely by the material's second- and third-order elastic moduli and may be interpreted as the nonlinearity parameter for Rayleigh waves.

Equations (3.1) and (3.2) predict the response of a single comb tooth. To use these equations to interpret this data, it was necessary to combine the contributions from each tooth in the following manner. First, the predicted response for each individual tooth was calculated for a range of detector positions  $X$  ( $Y = 0$ ). In the case of Eq. (3.2), this entailed numerical computation of the integral  $\mathcal{I}$ . The values of  $x$  and  $a$  varied slightly for each tooth depending on their specific length and position relative to the detector. The calculations produced a complex value for each tooth's response. The real and imaginary parts of these values were added together separately to obtain the complex predicted response for the entire extended comb source for each position  $X$ . Finally, the magnitude of the complex value was calculated to yield the predicted values of  $A_{1z}$  and  $A_{2z}$  vs  $X$ . This approach was not ideal since it was not a completely nonlinear solution. However, such predictions are likely to approximate comb behavior in its low-power limit at least.

#### B. Comparison of model and experiment

The curves in Fig. 3 show the predicted out-of-plane displacements obtained with the above model. For the fundamental component  $A_{1z}$ , the model prediction was fit to the data by adjusting only the values of  $v_0$  and  $\alpha_1$ ; all other parameters were known.<sup>18</sup> Adjusting  $v_0$  primarily affected the overall magnitude of the  $A_{1z}(X)$  curve; that is, the maximum value of  $A_{1z}$  increased as  $v_0$  increased. Changes in  $\alpha_1$  modified the curve's shape; mainly, the curve decreased more sharply with  $X$  as  $\alpha_1$  increased. With the assumption of no attenuation ( $\alpha_1 = 0$ ), a least-squares fit of the predicted  $A_{1z}$  to the data determined the value  $v_0 = 0.95$  m/s. These values are indicated by the dotted line in Fig. 3(a); they are in excellent agreement with the data. The effect of attenua-



tion on  $A_{1z}$  is shown by the other curves in Fig. 3(a). These curves were obtained by choosing a value for  $\alpha_1$  and then determining the value of  $v_0$  by a least-squares fit to the data. The dashed line indicates the predicted  $A_{1z}(X)$  for  $\alpha_1 = 0.63$  Np/m and  $v_0 = 0.98$  m/s. This combination of  $\alpha_1$  and  $v_0$  produced a curve that agreed with the data better than any other combination tried. To demonstrate how much  $v_0$  and  $\alpha_1$  could be adjusted and still achieve a reasonable fit to the data, a much larger value for the attenuation was chosen and the resulting  $v_0$  determined. The solid line corresponds to the predicted  $A_{1z}(X)$  for  $\alpha_1 = 2.50$  Np/m and  $v_0 = 1.07$  m/s. The predicted values agree reasonably well with the measured values, but diverge at large  $X$ . Note that by substantially increasing the attenuation, the best-fit value of  $v_0$  increased by only 13%. Also note that the relatively low values of  $v_0$  confirm the assumption of low-power operation and validate the use of a quasilinear model.

The agreement between model and experiment for  $A_{1z}(X)$  was not surprising, since diffraction and propagation of the fundamental are well understood. A more rigorous test of the model was its ability to predict the second-harmonic displacements  $A_{2z}(X)$ . For these calculations,  $v_0$  and  $\alpha_1$  were limited to the values obtained by fits to  $A_{1z}(X)$ . Thus only two adjustable parameters— $\alpha_2$  and  $\beta_{11}$ —were available to fit the predicted  $A_{2z}(X)$  to the data. Adjusting  $\alpha_2$  modified the curve's shape, particularly whether  $A_{2z}$  continued to increase with  $X$  (smaller  $\alpha_2$ ) or eventually reached a maximum (larger  $\alpha_2$ ). Larger values of  $\beta_{11}$  increased the overall magnitude of the curve. The dotted line in Fig. 3(b) shows the predicted values for  $\alpha_1 = \alpha_2 = 0$ ,  $v_0 = 0.95$  m/s, and  $\beta_{11} = 0.21$ . This value of  $\beta_{11}$  was chosen since it gave the best qualitative fit to the data at small values of  $X$ . However, the agreement lessens with increasing  $X$ .

Closer agreement between the model predictions and the measured values was obtained by including finite attenuation. In this case, a frequency dependence to the attenuation was assumed. The value of  $\alpha_2$  could thus be determined from the value of  $\alpha_1$  obtained in the fit to  $A_{1z}$ . The frequency dependences used in Fig. 3(b) represent the best qualitative match to the data points for the given  $\alpha_1$ . The value of  $\beta_{11}$  was then determined by a least-squares fit to the data points. For the case  $v_0 = 0.98$  m/s and  $\alpha_1 = 0.63$  Np/m, using the assumption  $\alpha \propto f^4$ ,  $\alpha_2 = 16\alpha_1 = 10.00$  Np/m. The least-squares fit to the data produced  $\beta_{11} = 0.23$ ; the corresponding curve for  $A_{2z}(X)$  is shown by the dashed line in Fig. 3(b). Also shown in Fig. 3(b) are the calculations for  $A_{2z}(X)$  using  $v_0 = 1.07$  m/s and  $\alpha_1 = 2.50$  Np/m. Here it was assumed that  $\alpha \propto f^2$ , so  $\alpha_2 = 4\alpha_1 = 10.00$  Np/m. The solid line shows the curve using the least-squares fit value  $\beta_{11} = 0.22$ . Both curves are in excellent agreement with the data except at the smallest values of  $X$ , which were in the near field of the transducer. In the near field, oscillations in the displacement occurred in the transverse ( $Y$ ) direction and the predicted on-axis amplitude of  $A_{2z}$  was much lower than its value slightly off-axis. Therefore, a small experimental uncertainty in the position of  $Y = 0$  could have resulted in too large a measured value  $A_{2z}(X)$ . Also, second-order effects such as interactions between the individual comb teeth may

have produced additional signals close to the source that were not accounted for in the model.

The numerical integration required by Eq. (3.2) was sufficiently cumbersome that a simultaneous multiparameter optimization of  $v_0$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $\beta_{11}$  proved impossible. Indeed, the objective of this work was not to determine precise values of the parameters from a fit to the data. The aim instead was to demonstrate the validity of the model in correctly predicting  $A_{1z}(X)$  and  $A_{2z}(X)$  using realistic estimates of the parameters. The graphs in Fig. 3 reveal that this aim has been achieved. Both the key qualitative features of the spatial evolution of  $A_{1z}$  and  $A_{2z}$  and their absolute amplitudes were correctly predicted by the model.

Moreover, the values for  $\beta_{11} = 0.21$ – $0.23$  obtained in this analysis agree with previous elastic-moduli measurements. Using literature values of the second- and third-order moduli of aluminum alloys,<sup>19,20</sup> calculated values of  $\beta_{11}$  fall between 0.13 and 0.45. (The large range in  $\beta_{11}$  is due partly to compositional differences in the alloys examined and partly to experimental uncertainty in measurements of third-order elastic moduli.) The values of the attenuation coefficient used in this analysis are also consistent with previous attenuation measurements<sup>21,22</sup> of megahertz-range Rayleigh waves in aluminum.

Before concluding, the origin of the harmonic generation is addressed. Harmonics could arise from physical sources such as dislocations or residual stress. However, good agreement was found between the data and the predictions of a model which included only the nonlinear elastic properties of aluminum. It is also possible that waves of frequency  $2\omega_0$  were excited by the comb itself. The predicted diffraction pattern versus  $X$  for such waves was like that in Fig. 3(a), but with the  $X$ -axis foreshortened. The diffraction pattern contained a region of increasing  $A_{2z}$  for  $X$  between the last minimum and last maximum (that is, at the transition from the near field to the far field). Thus one might wonder whether the observed increase in  $A_{2z}$  was due to this effect.

However, the last minimum in the predicted diffraction pattern occurs at  $X \approx 3.3$  cm. If comb excitation of  $2\omega_0$ -waves were significant, this implies that  $A_{2z}$  would increase and then decrease over the experimental range of  $X$ , rather than continually increasing as observed. Also, the measured increase in  $A_{2z}$  was larger than that predicted by diffraction calculations. If the  $2\omega_0$ -waves were generated only at the source, the diffraction pattern would show a relative increase of about 25% from their near-field minimum to their far-field maximum. However, an increase of about 90% was measured over the experimental range of  $X$ . (Introducing attenuation only reduces the relative increase.) This evidence supports the belief that most of the observed second harmonics were generated by the aluminum surface and not the comb.

#### IV. SUMMARY

In this paper, several methods for exciting megahertz-range, finite-amplitude, narrow-band Rayleigh waves in non-piezoelectric materials were compared. As a result of the comparison, a Sokolinskii comb device was chosen for

Rayleigh-wave harmonic-generation experiments in aluminum. The absolute out-of-plane components of the fundamental and second-harmonic displacements as a function of propagation distance were measured with a Michelson interferometer. The data were compared to the predictions of an existing quasilinear harmonic-generation model which included diffraction and attenuation. Good quantitative agreement between model and data was found using values of the nonlinearity parameter  $\beta_{11}=0.21-0.23$ . This is consistent with values for  $\beta_{11}$  calculated from literature values of the third-order elastic moduli for aluminum.

These experiments have demonstrated that comb devices can produce Rayleigh waves with sufficiently large amplitudes to observe second-harmonic generation. In addition, they are absolute measurements of harmonic displacements on a nonpiezoelectric substrate. Such quantitative measurements enable the detailed validation of theoretical models needed for a deeper understanding of nonlinear Rayleigh-wave propagation.

Plans for the future include experiments to better understand comb behavior and refinement of a theoretical model for comb transduction. This will enable the design and implementation of combs with greater efficiency that are optimized to the substrate under study. Devices with greater efficiency will also enable experiments to examine the second-harmonic amplitude as a function of fundamental amplitude. Such experiments could provide more information about  $\beta_{11}$  and thus verify the nonlinear effects observed here. In addition, a three-dimensional, completely nonlinear model for the comb is needed to replace the two-dimensional, linear superposition of nonlinear solutions used here. Much work is needed before Rayleigh-wave harmonic-generation techniques can be used for detailed materials characterization of surfaces. Nonetheless, these results demonstrate the potential of combs for nonlinear ultrasonic surface-wave experiments.

## ACKNOWLEDGMENTS

The author gratefully acknowledges useful discussions with E. J. Danicki, C. M. Fortunko (deceased), M. F. Hamilton, M. Hirao, and G. D. Meegan. M. C. Renken assisted with the initial transducer evaluation experiments. D. Gallagher fabricated the comb devices.

## APPENDIX: COMPARISON OF EXCITATION METHODS

A Rayleigh-wave transducer suitable for use in harmonic-generation experiments must possess the following features: ability to excite relatively large displacement amplitudes with high spectral purity and low harmonic distortion, operation at ultrasonic frequencies (typically, 1–30 MHz), and a material-independent excitation mechanism. What is the best way to meet these requirements? Several methods exist to excite narrowband Rayleigh waves, but information about their ability to create finite-amplitude waves is scant. To select the most promising approach for harmonic-generation experiments, several techniques were compared: an electromagnetic acoustic Rayleigh-wave trans-

ducer (EMAT);<sup>23</sup> longitudinal to Rayleigh-wave conversion by scattering from a series of grooves in the surface;<sup>24</sup> a longitudinal transducer tilted at 45° on the surface edge; a longitudinal transducer attached to a plastic wedge bonded to the surface; and a Sokolinskii comb. The edge, wedge, and comb techniques are discussed in detail by Viktorov.<sup>13</sup> Like the comb, the grooved-surface technique requires that the groove spacing correspond to the Rayleigh wavelength in the specimen.

Each of these methods was used to excite narrowband Rayleigh waves with a center frequency of 1 MHz in an aluminum plate. Identical experimental conditions were maintained when possible. For instance, all of the experiments except those with the EMAT used the same longitudinal transducer with the same applied voltage. The out-of-plane wave displacements at approximately 0.1 m from the source were measured using the Michelson interferometer described in Sec. I B. The signals were evaluated in both the time and frequency domains according to the characteristics discussed above. In addition, the practical convenience of each method was considered.

Qualitative comparison of the detected signals revealed that the wedge method produced the largest displacements. Unfortunately, the wedge assembly excited too many spurious harmonic components. The wave amplitudes excited by the EMAT and edge methods were judged too small for further consideration. On the other hand, the displacements created by both the comb and grooved-surface methods were at least half as large as those obtained by the wedge method and were roughly comparable in size. In addition, both the comb and grooved-surface methods displayed excellent spectral purity and low harmonic distortion. Since their excitation mechanisms are frequency selective, this was not surprising. The comb was judged preferable to the grooved-surface method from the standpoint of convenience, since it required neither surface modification nor two-sided access. Therefore, the experimental results indicated that the comb method was the most promising candidate for further experiments.

<sup>1</sup>J. H. Cantrell and W. T. Yost, "Acoustic harmonic generation from fatigue-induced dislocation dipoles," *Philos. Mag. A* **69**, 315–326 (1994).

<sup>2</sup>D. C. Hurley, D. Balzar, P. T. Purtscher, and K. W. Hollman, "Nonlinear ultrasonic parameter of quenched martensitic steels," *J. Appl. Phys.* **83**, 4584–4588 (1998).

<sup>3</sup>D. J. Barnard, G. E. Dace, and O. Buck, "Acoustic harmonic generation due to thermal embrittlement of Inconel 718," *J. Nondestruct. Eval.* **16**, 67–75 (1997).

<sup>4</sup>S. U. Faßbender, M. Kröning, and W. Arnold, "Measurement of adhesion strength using nonlinear acoustics," in *Proceedings of the Seventh International Symposium on the Nondestructive Characterization of Materials*, edited by A. L. Bartos, R. E. Green, Jr., and C. O. Ruud (Transtec Publications, Lebanon, NH, 1996), pp. 783–790.

<sup>5</sup>P. O. López, "Second-harmonic generation in an elastic surface wave in  $\alpha$ -quartz," *J. Appl. Phys.* **39**, 5400–5404 (1968).

<sup>6</sup>A. J. Slobodnik, Jr., "Nonlinear effects in microwave acoustic LiNbO<sub>3</sub> surface-wave delay lines," *J. Acoust. Soc. Am.* **48**, 203–210 (1970).

<sup>7</sup>E. G. H. Lean, C. C. Tseng, and C. G. Powell, "Optical probing of acoustic surface-wave harmonic generation," *Appl. Phys. Lett.* **16**, 32–35 (1970).

<sup>8</sup>E. L. Adler, E. Bridoux, G. Coussot, and E. Dieulesaint, "Harmonic generation of acoustic surface waves in Bi<sub>12</sub>GeO<sub>20</sub> and LiNbO<sub>3</sub>," *IEEE Trans. Sonics Ultrason.* **SU-20**, 13–16 (1973).

- <sup>9</sup>F. Rischbieter, "Messungen an Oberflächenwellen in festen Körpern," *Acustica* **16**, 76–83 (1965).
- <sup>10</sup>F. Rischbieter, "Messung des nichtlinearen Schallverhaltes von Aluminium mit Hilfe von Rayleighwellen," *Acustica* **18**, 109–112 (1967).
- <sup>11</sup>A. P. Brysev, V. A. Krasil'nikov, A. A. Podgornov, and I. Yu. Solodov, "Direct observation of the profile of an elastic wave of finite amplitude on the surface of a solid," *Sov. Phys. Solid State* **26**, 1275–1276 (1984).
- <sup>12</sup>G. D. Meegan, Jr., "Theory for the spectral evolution of nonlinear Stoneley and Scholte waves, and related experiments," Ph. D. thesis, University of Texas at Austin, 1996.
- <sup>13</sup>I. A. Viktorov, *Rayleigh and Lamb Waves: Physical Theory and Applications* (Plenum, New York, 1967), Chap. 2.
- <sup>14</sup>D. C. Hurley and C. M. Fortunko, "Determination of the nonlinear ultrasonic parameter  $\beta$  using a Michelson interferometer," *Meas. Sci. Technol.* **8**, 634–642 (1997).
- <sup>15</sup>D. J. Shull, E. E. Kim, M. F. Hamilton, and E. A. Zabolotskaya, "Diffraction effects in nonlinear Rayleigh wave beams," *J. Acoust. Soc. Am.* **97**, 2126–2137 (1995).
- <sup>16</sup>D. J. Shull, M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, "Harmonic generation in plane and cylindrical nonlinear Rayleigh waves," *J. Acoust. Soc. Am.* **94**, 418–427 (1993).
- <sup>17</sup>E. A. Zabolotskaya, "Nonlinear propagation of plane and circular Rayleigh waves in isotropic solids," *J. Acoust. Soc. Am.* **91**, 2569–2575 (1992).
- <sup>18</sup>Using contact transducers and standard pulse-echo techniques, the following values were obtained for the longitudinal and transverse velocities  $v_L$  and  $v_T$  of the aluminum specimen:  $v_L = 6.35 \pm 0.03$  km/s,  $v_T = 3.17 \pm 0.02$  km/s. The second-order elastic moduli were calculated from these values assuming the density  $\rho = 2.7 \times 10^3$  kg/m<sup>3</sup>. The Rayleigh velocity  $v_R = 2.94 \pm 0.03$  km/s was determined by measuring the time of arrival of low-amplitude Rayleigh waves for a range of source–detector separations. The slope of the line formed by plotting arrival time against propagation distance yielded  $v_R$ .
- <sup>19</sup>R. T. Smith, R. Sterns, and R. W. B. Stephens, "Third-order elastic moduli of polycrystalline metals from ultrasonic velocity measurements," *J. Acoust. Soc. Am.* **40**, 1002–1008 (1966).
- <sup>20</sup>V. K. Raju and P. J. Reddy, "Third-order elastic moduli of polycrystalline Al-Mg and Al-Cu alloys," *J. Phys. D* **14**, 65–70 (1981).
- <sup>21</sup>K. N. Vinogradov and G. K. Ul'yanov, "Measurement of the velocity and attenuation of ultrasonic surface waves in hard materials," *Sov. Phys. Acoust.* **5**, 296–299 (1959).
- <sup>22</sup>M. de Billy, G. Quentin, and E. Baron, "Attenuation measurements of an ultrasonic Rayleigh wave propagating along rough surfaces," *J. Appl. Phys.* **61**, 2140–2145 (1987).
- <sup>23</sup>R. B. Thompson, "Physical principles of measurements with EMAT transducers," in *Physical Acoustics XIX*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1990), Chap. 4.
- <sup>24</sup>R. F. Humphryes and E. A. Ash, "Acoustic bulk-surface-wave transducer," *Electron. Lett.* **5**, 175–176 (1969).

# An experimental verification of Rayleigh's interpretation of the Sondhauss tube

Guadalupe Huelsz and Eduardo Ramos

*Centro de Investigación en Energía, UNAM, AP 34, 62580 Temixco, Mor., Mexico*

(Received 13 January 1999; accepted for publication 29 April 1999)

The interaction of a sound wave in a compressible fluid in contact with solid boundaries produces thermoacoustic effects. A century ago, Lord Rayleigh gave a general qualitative explanation of the acoustic power production by unsteady heat transfer to compressible fluid in terms of the relative phase of the pressure wave and the heat transferred to the fluid. In quantitative terms his interpretation is equivalent to saying that a wave is promoted if the absolute value of this relative phase is less than  $\pi/2$ ; otherwise the wave is attenuated. In this paper, an experimental investigation of the attenuation or promotion of an acoustic wave as a function of an axial temperature gradient at the wall of the waveguide is presented. Using the techniques previously developed [J. Acoust. Soc. Am. **103**, 1532–1537 (1998)], acoustic temperature and pressure oscillations measurements in the thermal boundary layer of an acoustic standing wave in air were made. The results reported here experimentally confirm the validity of Rayleigh's interpretation of the power production for this case. This confirmation is more direct than any previous evidence. © 1999 Acoustical Society of America. [S0001-4966(99)02708-3]

PACS numbers: 43.35.Ud, 43.58.Vb [SLE]

## INTRODUCTION

A sound wave in a compressible fluid consists of coupled velocity, density, pressure, and temperature oscillations. Their combined interaction with solid boundaries produces thermoacoustic effects. Since the eighteenth century, glassblowers noticed that when a hot glass bulb was attached to a cold glass tubular stem, the stem tip sometimes emitted sound.<sup>1</sup> Among the first accounts of thermoacoustic oscillations is that of Sondhauss<sup>2</sup> published in 1850. The discovery of the inverse effect, i.e., that acoustic oscillations on a fluid in contact with a solid boundary can produce a heat pumping effect, is more recent.<sup>3</sup> In the last 30 years many authors have been interested in thermoacoustic effects in the context of its applications to engines and refrigerators.<sup>4–8</sup>

Rayleigh's interpretation to the generation of acoustic oscillations in a fluid due to heat transfer<sup>9</sup> stated that if a gas receives heat at the moment of greatest compression and gives it up when it is most rarefied, the oscillatory motion is promoted or amplified, that is, there is a production of acoustic power. Contrarily, an acoustic wave is attenuated if heat is taken from the gas at the moment of maximum compression and supplied to it when it is most rarefied; in this case there is absorption of acoustic power. Therefore, the relative phase of the pressure wave and the heat communication determines the production or absorption of acoustic power. Defining  $\theta$  as the relative phase of the pressure wave and the heat flux to the fluid, Rayleigh's interpretation gives the following criterion: if the absolute value of  $\theta$  is less than  $\pi/2$  there is production of acoustic power, while if it is larger than  $\pi/2$  there is absorption of acoustic power. Rayleigh's criterion can also be expressed in an integral form: if the time average of the product of pressure wave and the heat flux to the fluid ( $\langle p^* q^* \rangle \equiv (1/\tau) \int_0^\tau p^* q^* dt$ ) is positive, then there is production of acoustic power, while if it is negative,

there is absorption of acoustic power.  $p^*$  and  $q^*$  are pressure and the heat transferred to the fluid.

Putman and Dennis<sup>10</sup> calculated the conditions for wave attenuation or excitation in terms of the relative phase of pressure and heat transferred oscillations for an inviscid, ideal gas, and unspecified heat transfer mechanism. Their calculation agreed with Rayleigh's criterion. Theoretical studies that use Rayleigh's criterion in integral form for the driving mechanism in different systems have been developed.<sup>11–15</sup> Heckl<sup>13</sup> implicitly used Rayleigh's criterion in integral form in the equation for the energy gain at the gauze for a nonlinear theory of the Rijke tube, and sound pressure levels predicted by this theory were close to that experimentally measured. Experimental investigations where the Rayleigh criterion is directly or indirectly involved are numerous, mostly in the fields of combustion and thermoacoustic applications.<sup>6</sup> Here are only reviewed those that are more relevant in the present context, i.e., those where specific phase measurements have been made in order to use Rayleigh's criterion for interpreting results. Kaskan<sup>16</sup> used Rayleigh's criterion to experimentally determine the relative importance of two possible causes of the oscillatory heat release in premixed flames burning in tubes. Langhorne<sup>17</sup> studied the reheat buzz, a combustion instability that occurs in the afterburners of jet aeroengines. He made  $C_2$  or CH emission intensity measurements which he proved to be a measure of the heat release rate at a given fuel-to-air ratio. The author correlated the global behavior of the system with pressure band levels. Experiments where the emission and pressure were in phase along the afterburner showed higher pressure band levels than those where he found combined regions where emissions and pressure were in and out of phase. His results are consistent with Rayleigh's criterion. Macquisten and Dowling<sup>18</sup> made similar experiments for higher inlet Mach numbers. They reported experiments with



high pressure band levels where  $C_2$  emission and pressure along the afterburner were in phase; this again is consistent with Rayleigh's criterion. The experimental results quoted above are in agreement with Rayleigh's criterion, and therefore can be considered indirect experimental verifications.

The unsteady heat transfer in combustion systems is volumetric, whereas in the thermoacoustic effect studied here, the heat injection or withdrawal to or from the fluid takes place at the solid boundary; thus in this case the relative phase of the pressure wave and the transverse heat flux to the fluid at the boundary determines whether there is production or absorption of acoustic power. In a previous theoretical work<sup>19</sup> the authors demonstrated the agreement between the physical interpretation given by Rayleigh and the modern linear theory used to study the thermoacoustic effect when the heat is transferred through the wall. The experimental confirmation of Rayleigh's interpretation in this case is the objective of the present work.

## BACKGROUND THEORY

In order to help understanding the experimental results reported here, in the following a brief description of theoretical work is presented. The linear theory used is based in that developed by Rott.<sup>20</sup> It assumes the existence of a one-dimensional harmonic acoustic wave in a compressible fluid propagating parallel to a solid boundary. The solid boundary has a time-average temperature gradient parallel to the wave propagation direction and that extends through the transverse direction measured from the wall ( $y^*$ ). The theory assumes that the resulting oscillatory thermal boundary layer depth ( $\delta_k$ ) is much smaller than the wavelength, thus the boundary layer approximation is applicable.  $\delta_k$  is equal to  $\sqrt{2k/\rho_m^* C_p \omega}$ , where  $k$ ,  $\rho_m^*$ , and  $C_p$  are, respectively, the thermal conductivity, the density, and the isobaric specific heat of the fluid, and  $\omega$  is the wave angular frequency. Using the governing balance equations, expressions for temperature, density, and velocity oscillations can be obtained in terms of time-average values of these variables, fluid and solid properties, and the pressure oscillation. These expressions are used to calculate the relative phase of the pressure wave and the transverse heat flux to the fluid at the boundary  $|\theta| = \arccos(2\langle p^* q_w^* \rangle / \|p^*\| \|q_w^*\|)$ .  $q_w^*$  is the heat supplied to the fluid from the solid boundary at a given axial position  $x^*$ , and is given by  $q_w^* = -k(\partial T^* / \partial y^*)|_{y^*=0}$ , where  $T^*$  is the temperature. The time-average acoustic power production per unit area at  $x^*$  was also calculated. The resulting expressions for these quantities satisfy the criterion for acoustic power production based in Rayleigh's interpretation.

In acoustic waves propagating parallel to a wall, if the attenuation or excitation is a small enough local effect that it does not affect the pressure oscillation externally imposed, the short plate approximation<sup>6</sup> can be used. Using this approximation, the authors<sup>21</sup> calculated the phase of the temperature oscillation relative to pressure ( $\alpha$ ) and demonstrated that in standing waves there is a one-to-one relation between this phase in the limit when the distance to the solid boundary goes to zero [ $\lim_{y \rightarrow 0}(\alpha)$ ] and the local time-average acoustic power production. Thus in standing waves, in the

short plate approximation, Rayleigh's criterion for acoustic power production can be translated from the relative phase between pressure and transverse heat flux ( $\theta$ ) to the relative phase of the pressure and the temperature oscillations in the limit when the distance to the solid boundary goes to zero [ $\lim_{y \rightarrow 0}(\alpha)$ ]. If the absolute value of  $\lim_{y \rightarrow 0}(\alpha)$  is larger than  $\pi/2$  there is local acoustic power production, while if the absolute value of it is smaller than  $\pi/2$  there is local acoustic power absorption. In the present work, the authors used this criterion to provide experimental evidence for Rayleigh's interpretation of the acoustic power production.

It is worth emphasizing the difference between the two relative phases introduced.  $\theta$  is the relative phase of the pressure wave and the transverse heat flux to the fluid at the boundary and evidently is only defined at the boundary, while  $\alpha$  is the relative phase of the pressure wave and the temperature and it is defined in all the fluid. In standing waves,  $\lim_{y \rightarrow 0}(\alpha)$  is directly related with  $\theta$ ; in fact  $\lim_{y \rightarrow 0}(\alpha) = \theta \pm \pi/2$  (+ or - is taken such that the value is in the interval where  $\alpha$  is defined).

In this work all results will be presented in a nondimensional form. The axial coordinate origin is placed at the pressure node and is scaled by  $x = \omega x^* / c$ , where  $c$  is the velocity of sound in the fluid at the time-average temperature of a reference point. The transverse coordinate is scaled with  $y = y^* / \delta_k$ , where  $\delta_k$  is evaluated at the time-average temperature of the reference point. Time-average quantities are scaled with their corresponding value at a reference point, while oscillatory components are scaled using the corresponding local adiabatic amplitude. This scaling leads to the following expression for the nondimensional axial temperature gradient:  $\nabla T_m = c \nabla T_m^* / \omega T_{mr}^*$ , where  $\nabla T_m^*$  is the dimensional axial temperature gradient and  $T_{mr}^*$  the time-average temperature at the reference point (subindex  $r$ ).

Short plate approximation theory<sup>6</sup> predicts that for a given axial position, local acoustic power production occurs in a standing wave, provided that there is a time-average axial temperature gradient in such a way that the region at the pressure antinode is hotter than that at the node and this axial temperature gradient is larger than a critical value. For gradients smaller than the critical value the wave will be attenuated, whereas for larger gradients the wave will be promoted and there will be acoustic power production. The theory also predicts that for  $\nabla T_m > \nabla T_{mc}$  there is an optimum position of the plate ( $x_{ro}$ ) where acoustic power production achieves a maximum. This position is slightly dependent on the temperature gradient, but is close to  $0.25\pi$ . The nondimensional critical temperature gradient obtained by this theory is

$$\nabla T_{mc} = \frac{(\gamma - 1)(1 + \sqrt{\text{Pr}})}{\beta_{mr} T_{mr}^*} \tan(x_r),$$

where  $\gamma$  is the specific heats ratio and Pr is the Prandtl number; all properties of the fluid are evaluated at the temperature  $T_{mr}^*$  of the center of the plate which is the reference point. If there is an axial temperature gradient, the temperature oscillations in the fluid are caused not only by the pressure oscillations but also by the combined action of heat transfer through the wall and the displacement oscillation.

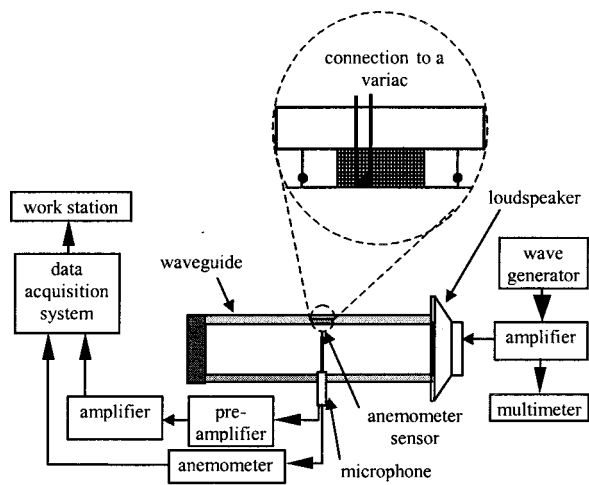


FIG. 1. Experimental setup.

The critical temperature gradient is such that the physical conditions lead to cancellation of the two effects, thus for  $\nabla T_{mc}$  the temperature oscillation near the wall vanishes, and the unsteady heat transferred from the wall is zero.

## EXPERIMENTS

The objective of this work is to experimentally verify Rayleigh's criterion in the case where heat is transferred through a wall which has a time-average axial temperature gradient. Therefore the aim is to measure  $\lim_{y \rightarrow 0}(\alpha)$  as a function of  $\nabla T_m$  in a standing wave. To do this, an experimental setup capable to establish a one-dimensional harmonic acoustic wave in air propagating parallel to a solid boundary was designed and constructed. This setup (Fig. 1) consists of a semi-closed waveguide filled with air at atmospheric pressure, the waveguide is a duct with a rectangular internal cross section of  $0.098 \text{ m} \times 0.054 \text{ m}$  and  $0.60\text{-m}$  internal length. One end is closed with a massive coldroll cap; in the other end there is a loudspeaker connected to a wave generator via an amplifier. Using this system, a standing wave with frequency  $130 \text{ Hz}$  and pressure amplitude up to  $200 \text{ Pa}$  ( $140 \text{ dB}$ ) can be established. The waveguide dimensions satisfy the requirements for the generation of a one-dimensional wave.<sup>22</sup> More details of the experimental setup are given in Ref. 23. In the central zone ( $x = 0.26\pi$ ) of the upper wall, there is a flush mounted piece of ceramic (from an automobile catalytic converter) where an axial temperature gradient is generated. This is done by an appropriate threading of an electric resistance wire that passes through some of the linear narrow channels of the ceramic (see inset in Fig. 1). Due to the geometry and ceramic properties, it is possible to produce and maintain an axial temperature gradient up to  $100 \text{ }^\circ\text{C/cm}$  in this zone of the wall. In the following, this zone will be simply denominated as the plate. The axial position of this plate ( $x_r = 0.26\pi$ ) was chosen considering the theoretical optimum position for thermoacoustic power production. The plate was placed at the top wall of the waveguide in order to reduce natural convection in the air in the region adjacent to the plate. The front and back walls of the waveguide are made of transparent plexiglass to permit vi-

ualization of the position of the cold wire anemometer sensor which is used to measure the temperature oscillation. The pressure oscillation is measured with a condenser microphone. Both instruments are located at the same axial position of the ceramic plate; the microphone is flush mounted in the bottom wall and the anemometer sensor can be displaced vertically to vary its distance to the plate. A fast data acquisition system ( $10^5$  samples/s) and a workstation are used to register the microphone and cold wire anemometer output signals. The analysis output signals program includes a band pass filter ( $130 \pm 2.5 \text{ Hz}$ ) in the frequency domain which is applied to the microphone and cold wire anemometer output signals. In a previous work the authors<sup>23</sup> developed an experimental technique for measuring the relative phase between temperature and pressure oscillations in acoustic waves. This technique proved very reliable for measurements inside the oscillatory boundary layer produced by a standing acoustic wave with isothermal conditions in the solid boundary.

For the experimental conditions reported here  $\delta_k$  is about  $2.5 \times 10^{-4} \text{ m}$  and the theoretical critical nondimensional axial temperature gradient of the plate ( $\nabla T_m$ ) is  $0.737$ . The axial temperature gradient at the wall was varied from  $0$  to  $4.5$ . This range allows one to monitor the behavior of acoustic waves from attenuation to excitation.

The procedure followed to obtain the limit value of the relative phase between pressure and temperature oscillations when the distance to the solid boundary goes to zero [ $\lim_{y \rightarrow 0}(\alpha)$ ] is:

(a) Given an axial temperature gradient, the relative phase measurements were made within the thermal boundary layer at several distances from the plate. The uncertainty associated with the definition of the zero position of the temperature sensor is mainly due to two factors: one of them is the irregularity of the plate surface; the other one is the actual position of the wire sensor. The platinum sensor is held in position by welding it to the two prongs of the holder. In principle, the two prongs are the same length and the wire is attached at the very end of the prongs, but in fact the prongs may have a slightly different length. Although every effort was made to minimize the errors incurred due to these factors, it is estimated that the total uncertainty in the nondimensional position of the sensor is  $\pm 0.03$  ( $\sim 7.5 \text{ } \mu\text{m}$ ).

(b) The limit of  $\alpha$  as  $y \rightarrow 0$  is estimated by extrapolating the best fit curve to the data points at distances smaller than  $0.25$  from the plate. For axial temperature gradients in the range  $0.25 < \nabla T_m < 1.0$  the amplitude of the temperature oscillation for points near the plate is very much reduced. This can be expected from the physical meaning of  $\nabla T_{mc}$  (see Background theory section). As a result, the signal-to-noise ratio is reduced and for distances smaller than  $0.25$  no reliable data can be obtained. Thus for these  $\nabla T_m$  the limit of  $\alpha$  as  $y \rightarrow 0$  is estimated by extrapolating the best fit curve to the data points at distances from the plate in the range of  $0.25 < y < 0.60$ .

## RESULTS

The main results of this work are the measurements of the relative phase of the pressure and temperature oscilla-

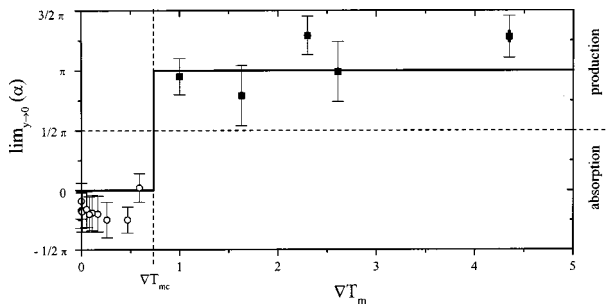


FIG. 2. Relative phase of the pressure and temperature oscillations in the limit when the distance to the solid boundary goes to zero [ $\lim_{y \rightarrow 0}(\alpha)$ ] plotted as a function of the nondimensional axial temperature gradient of the plate ( $\nabla T_m$ ). Symbols represent experimental results:  $\circ$  attenuation and  $\blacksquare$  amplification. The continuous line is the linear theory prediction.

tions in the limit when the distance to the solid boundary goes to zero [ $\lim_{y \rightarrow 0}(\alpha)$ ] as a function of the nondimensional axial temperature gradient of the plate ( $\nabla T_m$ ). These are plotted in Fig. 2. Symbols represent experimental results and the continuous line is the linear theory prediction. The experimental points with  $\nabla T_m$  smaller than the theoretical critical value  $\nabla T_{mc} = 0.737$  correspond to attenuation of the acoustic wave, and their absolute value of  $\lim_{y \rightarrow 0}(\alpha)$  is smaller than  $\pi/2$ .  $\nabla T_m$  larger than  $\nabla T_{mc}$  corresponds to amplification and their absolute value of  $\lim_{y \rightarrow 0}(\alpha)$  is larger than  $\pi/2$ . These results are the experimental confirmation of the validity of Rayleigh's interpretation of the power production in the thermoacoustic effect.

Quantitative differences between experimental and theoretical results (less than 15%) are best understood by examining the features of the relative phase of the pressure and temperature oscillations as a function of the distance to the solid boundary. Depending on the axial temperature gradient, the theory predicts three qualitatively different trends of this function: (a) for  $0 < \nabla T_m < \gamma - 1$  both  $\lim_{y \rightarrow 0}(\alpha)$  and  $\lim_{y \rightarrow \infty}(\alpha)$  are zero, for the experimental conditions reported here  $\gamma - 1 = 0.40$ ; (b) for  $\gamma - 1 < \nabla T_m < \nabla T_{mc}$ ,  $\lim_{y \rightarrow 0}(\alpha)$  is zero, but  $\lim_{y \rightarrow \infty}(\alpha) = \pi$ , and (c) for  $\nabla T_{mc} < \nabla T_m$ , both  $\lim_{y \rightarrow 0}(\alpha)$  and  $\lim_{y \rightarrow \infty}(\alpha)$  are  $\pi$ . These three cases were experimentally confirmed. One example of each is plotted in Fig. 3. In all cases, there is a large gradient near the solid boundary, and therefore the experimental difficulties to obtain measurements close to the solid boundary prevent an accurate determination of the limit value. The experimental observations for  $\nabla T_m = 0.26$  coincide quantitatively within the experimental error with the linear theory up to  $y = 0.04$ . For  $\nabla T_m = 0.59$ , the linear theory and the experimental measurements are in agreement for  $0.25 < y < 0.60$ . At larger distances, the experimental observations are located systematically under the linear theory curve. This effect is due to the fact that the experimental axial temperature gradient is not maintained at distances  $y > 0.60$ . Also, as explained before, in this case the amplitude of the temperature oscillations at distances  $y < 0.25$  were not large enough to be considered reliable. In spite of these technical difficulties, reliable data could be obtained to extrapolate the limiting value of  $\alpha$ . The experimental data for axial temperature gradients larger than the critical value ( $\nabla T_m = 1.00$ ) have a qualitative difference with the theoretical prediction. The experimental observa-

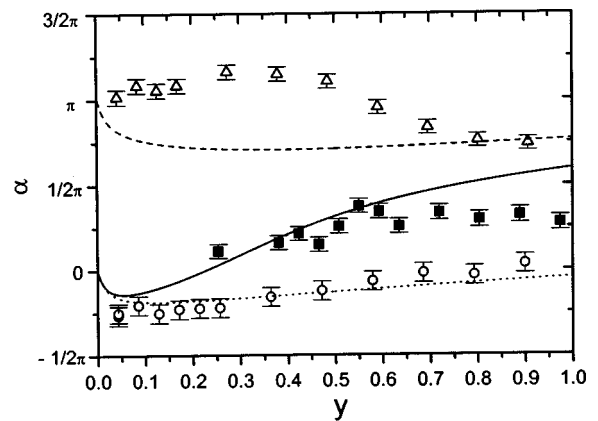


FIG. 3. Three qualitatively different trends of the relative phase of the pressure and temperature oscillations ( $\alpha$ ) as a function of the nondimensional distance to the solid boundary ( $y$ ). Symbols represent experimental results and lines are linear theory predictions, corresponding to different nondimensional axial temperature gradients:  $\circ$  . . . .  $\nabla T_m = 0.26$ ,  $\blacksquare$  —  $\nabla T_m = 0.59$ , and  $\triangle$  ---  $\nabla T_m = 1.00$ .

tions display a local maximum at  $y \sim 0.30$  which is not predicted by the linear theory. This feature was observed in all experiments with  $\nabla T_m > \nabla T_{mc}$ . This difference can be due to simplifications made in the theory that become more important as the axial temperature gradient increases, such as neglect second order terms in variables, assume that the axial temperature gradient is constant through the transverse direction, and that this gradient does not affect the standing wave pressure oscillation pattern.

As further comparison between theoretical and experimental results, the nondimensional amplitude of the temperature oscillation, away from the wall ( $T_0$ ), is plotted in Fig. 4 as a function of the nondimensional axial temperature gradient ( $\nabla T_m$ ). For clarity in the presentation, when temperature and pressure oscillations are in phase, the amplitudes  $T_0$  are defined as negative. The theoretical value of  $T_0(\nabla T_{mc}) = \sqrt{\text{Pr}}$ . For the experimental conditions  $\sqrt{\text{Pr}} = 0.84$ . Experi-

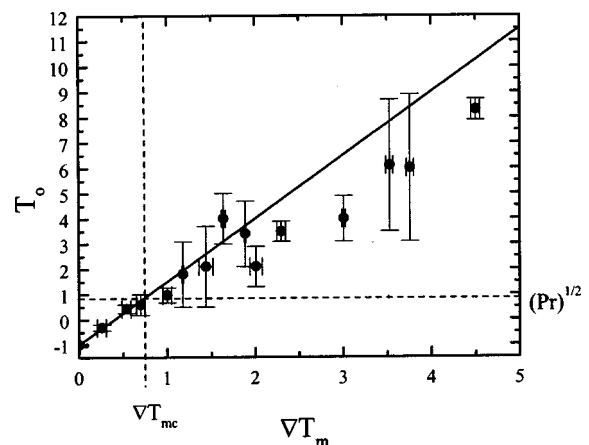


FIG. 4. Nondimensional amplitude of the temperature oscillation, away from the wall ( $T_0$ ) as a function of the nondimensional axial temperature gradient ( $\nabla T_m$ ). When temperature and pressure oscillations are in phase, the amplitudes  $T_0$  are defined as negative. Symbols represent experimental results and line is the linear theory prediction.

mental results are the average of the nondimensional amplitude of the temperature oscillation for  $3 < \gamma < 8$ . Theory predicts that when  $\nabla T_m < \nabla T_{mc}$  then  $T_0 < \sqrt{\text{Pr}}$ , while if  $\nabla T_m > \nabla T_{mc}$  then  $T_0 > \sqrt{\text{Pr}}$ . As can be appreciated from Fig. 4, this is in agreement with experimental results. Points with  $\nabla T_m > 2$  fall consistently below the theoretical curve. This is an indication that, as commented before, the thermal gradient is not maintained away from the wall.

In summary, Rayleigh's interpretation of the acoustic power production in thermoacoustic phenomenon has been experimentally verified for the case where the unsteady heat transfer is given through a wall with a time-average axial temperature gradient.

## ACKNOWLEDGMENTS

Useful discussions with Professor Mihir Sen of Notre Dame University are gratefully acknowledged. This work has been partially supported by DGAPA-UNAM IN103197 and CONACYT G0044E projects.

<sup>1</sup>Castberg, "Ueber die Fata Morgana und ähnliche Phänomene," Gilbert's Annalen der Physik **17**, 183–199 (1804).

<sup>2</sup>C. Sondhauss, "Ueber die Schallschwingungen der Luft in erhitzten Glasröhren und in gedeckten Pfiefen von ungleicher Weite," Ann. Phys. (Leipzig) **79**, 1–34 (1850).

<sup>3</sup>P. Merkli and H. Thomann, "Thermoacoustic effects in a resonant tube," J. Fluid Mech. **70**, 161–177 (1975).

<sup>4</sup>J. Wheatley, T. Hofler, G. W. Swift, and A. Migliori, "An intrinsically irreversible thermoacoustic heat engine," J. Acoust. Soc. Am. **74**, 153–170 (1983).

<sup>5</sup>G. W. Swift, A. Migliori, T. Hofler, and J. Wheatley, "Theory and calculations for an intrinsically irreversible acoustic prime mover using liquid sodium as primary working fluid," J. Acoust. Soc. Am. **78**, 767–781 (1985).

<sup>6</sup>G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1988).

<sup>7</sup>A. A. Atchley, "Standing wave analysis of a thermoacoustic prime mover below onset of self-oscillation," J. Acoust. Soc. Am. **92**, 2907–2914 (1990).

<sup>8</sup>G. W. Swift, "Thermoacoustic engines and refrigerators," Phys. Today **48**, 22–28 (1995).

<sup>9</sup>J. W. S. Rayleigh, *The Theory of Sound*, Vol. 2, Sec. 322 (1886) (Dover, New York, 1945).

<sup>10</sup>A. A. Putman and W. R. Dennis, "A study of burner oscillations of the organ-pipe type," Trans. ASME **75**, 15–28 (1953).

<sup>11</sup>J. J. Bailey, "A type of flame-excited oscillation in a tube," J. Appl. Mech. **24**, 333–339 (1957).

<sup>12</sup>F. E. C. Culick, "A note on Rayleigh's criterion," Combust. Sci. Technol. **56**, 159–166 (1987).

<sup>13</sup>M. A. Heckl, "Nonlinear acoustic effects in the Rijke tube," Acustica **72**, 63–71 (1990).

<sup>14</sup>F. E. C. Culick, "Combustion instabilities and Rayleigh's criterion," in *Modern Research Topics in Aerospace Propulsion (In honor of Corrado Casci)*, edited by G. Angelino, L. De Luca, and W. A. Sirignano (Springer-Verlag, New York, 1992), pp. 135–151.

<sup>15</sup>T.-S. Roh, I.-S. Tseng, and V. Yang, "Effects of acoustic oscillations on flame dynamics of homogeneous propellants in rocket motors," J. Propul. Power **11**, 640–650 (1995).

<sup>16</sup>W. E. Kaskan, "An investigation of vibrating flames," *4th Int. Symp. Combust* (Williams and Wilkins, Baltimore, 1953), pp. 575–591.

<sup>17</sup>P. J. Langhorne, "Reheat buzz: An acoustically coupled combustion instability. Part 1. Experiment," J. Fluid Mech. **193**, 417–443 (1988).

<sup>18</sup>M. A. Macquisten and A. P. Dowling, "Low-frequency combustion oscillations in a model afterburner," Combust. Flame **94**, 253–264 (1993).

<sup>19</sup>G. Huelsz and E. Ramos, "A physical interpretation of the thermoacoustic effect," J. Non-Equilib. Thermodyn. **21**, 278–284 (1996).

<sup>20</sup>N. Rott, "Damped and thermally driven acoustic oscillations in wide and narrow tubes," Z. Angew. Math. Phys. **20**, 230–243 (1969).

<sup>21</sup>G. Huelsz and E. Ramos, "On the phase difference of the temperature and pressure waves in the thermoacoustic effect," Int. Comm. Heat Mass Transfer **22**, 71–80 (1995).

<sup>22</sup>L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982), pp. 200.

<sup>23</sup>G. Huelsz and E. Ramos, "Temperature measurements inside the oscillatory boundary layer produced by acoustic waves," J. Acoust. Soc. Am. **103**, 1532–1537 (1998).



# Thermal diffusion and mixture separation in the acoustic boundary layer

G. W. Swift and P. S. Spoor

Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos,  
New Mexico 87545

(Received 11 February 1999; revised 25 June 1999; accepted 25 June 1999)

Oscillating thermal diffusion in a sound wave in a mixture of two gases is remarkably effective for separating the components of the mixture. We consider this separation process in boundary-layer approximation, with zero temperature gradient and zero concentration gradient along the direction of sound propagation. In the boundary layer, the combination of thermal diffusion with the oscillating temperature gradient and oscillating velocity gradient leads to second-order time-averaged fluxes of the two components of the mixture in opposite directions, parallel to the wave-propagation direction. The oscillating thermal diffusion also adds to the dissipation of acoustic power in the boundary layer, modifying thermal-relaxation dissipation but leaving viscous dissipation unchanged. © 1999 Acoustical Society of America. [S0001-4966(99)03110-0]

PACS numbers: 43.35.Ud [HEB]

## INTRODUCTION

In experiments on mode locking in acoustically coupled acoustic resonators,<sup>1</sup> we observed an anomalous difference in the resonance frequencies of the two resonators when using a He–Xe mixture. This frequency difference, up to 3% of the resonance frequency, was too large to be explained by any known difference between the resonators, such as temperature or geometry. We concluded that the sound wave in the acoustic coupler was separating the helium and xenon, thereby enriching one resonator with helium and the other with xenon.

We realized that this mass separation could be due to the mechanism illustrated in Fig. 1, which shows how the mixture can be separated by a combination of three effects in the boundary layers: oscillating temperature gradients in the thermal boundary layer, thermal diffusion, and oscillating velocity gradients in the viscous boundary layer. In a typical mixture of helium and xenon, the Prandtl number  $\sigma$  is about 1/4, so the viscous penetration depth  $\delta_\nu$  is about half the thermal penetration depth  $\delta_\kappa$ , as shown in Fig. 1. For standing-wave phasing in a channel whose diameter is much larger than these penetration depths, we might think of the wave as consisting of the four steps equally spaced in time shown in Fig. 1(a)–(d). In the first step, while the pressure is high, the time-dependent part of the temperature has a steep gradient within  $\delta_\kappa$  of the boundary as shown at the bottom of Fig. 1(a), due to the adiabatic temperature rise in the gas far from the wall and the large solid heat capacity of the wall itself. During this time, thermal diffusion drives the heavy component down the temperature gradient toward the boundary and the light component up the temperature gradient away from the boundary. (Signs may differ for different gases.) Hence, at the end of this time the gas near the solid boundary is enriched in the heavy component and depleted of the light component, while the gas approximately  $\delta_\kappa$  from the solid boundary is enriched in the light component and depleted of the heavy component. In the second step, the gas moves upward, with a steep gradient of velocity within  $\delta_\nu$  of

the solid boundary due to viscosity, as shown at the bottom of Fig. 1(b). During this time, the heavy-enriched gas is relatively immobilized in the viscous boundary layer, while the light-enriched gas, just outside of the viscous boundary layer, moves easily upward. In the third step, shown in Fig. 1(c), low pressure reverses the sign of the temperature gradient, so the thermal diffusion reverses direction, forcing the heavy component away from the boundary and the light component toward the boundary. Thus in the fourth step, shown in Fig. 1(d), light-enriched gas is relatively immobilized while heavy-enriched gas moves easily downward. The net effect of these four steps is that some of the heavy component moves downward while some of the light component moves upward, as shown in Fig. 1(e).

Separation of gas mixtures using a sound wave has been observed earlier<sup>2</sup> and was attributed both to barodiffusion and thermal diffusion,<sup>3</sup> but with the latter involving the interaction of the second-order, time-averaged boundary-layer temperature gradient with the ordinary, second-order Rayleigh streaming.<sup>4</sup> We suspect that this fourth-order separation mechanism is generally weaker than our second-order process. Acoustic separation in mixtures of three gases using the differences between the ordinary mass diffusivities of two of the components through the third component has also been described<sup>5</sup> by Howell. This mechanism works independently of the boundary-layer temperature gradients and thermal diffusion of our process. Thermoacoustic condensation and evaporation of one component in a gas mixture have also been studied<sup>6</sup> by Raspert *et al.*, with oscillating mass diffusion of the condensing component.

Here, we present the first theoretical steps in support of the process illustrated in Fig. 1. To investigate a simple case, we restrict our analysis to the boundary-layer approximation in a two-component gas, and we assume that time-averaged temperature gradients and concentration gradients parallel to the direction of wave propagation are negligible.

After a brief introduction to the relevant length scales in the problem, we derive the oscillating temperature and oscil-

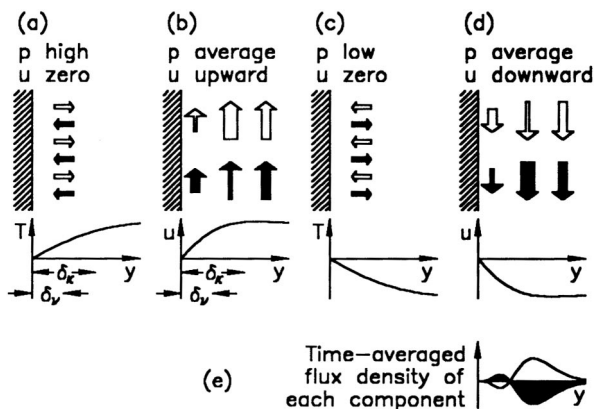


FIG. 1. Schematic of a possible separation process near a solid boundary in a standing wave, in a gas mixture with Prandtl number  $\sigma \sim 1/4$ . The solid arrows show motion of the heavy component, and the light arrows show motion of the light component. The lengths of the arrows represent velocity, and the widths of the arrows represent the local concentration of the components. As described in the text, (a) through (d) show processes occurring at time intervals separated by  $1/4$  of the period of the sound wave. The net result, shown in (e), is flux of the light component upward and flux of the heavy component downward.

lating concentration in a binary ideal gas mixture in boundary-layer approximation. We find that significant concentration oscillations exist, but with a different phasing than that suggested by Fig. 1. Next, we derive an expression for the second-order separation flux. The molar separation rate is surprisingly large for He–Xe and He–Ar mixtures, of order  $10^{-3}M^2a$ , where  $M$  is the acoustic Mach number and  $a$  is the sound speed, suggesting that this process might be practical for deliberate separation of gas mixtures. The spectrum of practical applications requiring separation of mixtures is broad, including large-scale industrial processes such as petroleum refining, air separation, and beverage processing, and smaller-scale processes such as isotope separation and chemical analysis. A large number of “physical” mixture-separation techniques<sup>7</sup> are well understood and in widespread use, including time-independent thermal diffusion, ordinary diffusion, fractional distillation, centrifugation, electromagnetic separation, chromatography, and superfluid heat flush. Much work remains in order to evaluate whether thermoacoustic mixture separation might find a useful niche in this vast industry. If so, we anticipate apparatus resembling thermoacoustic refrigerators,<sup>8</sup> with the large surface area of the stack (in which pore dimensions are a few times the thermal penetration depth) providing a large separation rate, and with feedstock entering and products leaving the acoustic system at standing-wave nodes.<sup>9</sup>

Next, we derive an expression for dissipation of acoustic power, showing that thermal diffusion adds slightly to the well-known viscous and thermal-relaxation dissipation mechanisms. Although the contribution of thermal diffusion to *bulk* attenuation of sound in gas mixtures is well known,<sup>10</sup> we have found only one published discussion<sup>2</sup> of this effect for boundary-layer attenuation.

It will be important to extend this work beyond the restrictive assumptions we have used here, in order to evaluate the phenomenon’s usefulness for practical mixture separation and its effect on acoustic power dissipation/production

and enthalpy flux in the stacks of thermoacoustic engines and refrigerators<sup>8</sup> using gas mixtures.<sup>11,12</sup> These devices do not operate in the boundary-layer regime, and axial time-averaged temperature gradients always exist and concentration gradients may in fact exist. Until the present work is suitably extended, thermoacoustics calculations (using codes such as DeltaE<sup>13</sup> versions 1 through 4) for gas mixtures cannot be trusted.

## I. IMPORTANT LENGTH SCALES

We will consider oscillations in a channel, in which all variables oscillate sinusoidally with time at frequency  $f$ . The wavelength  $\lambda = a/f$ , where  $a$  is the sound speed, is an important length scale, especially in the direction  $x$  of the gas displacement oscillations, and is much larger than all other length scales in the problem. The amplitude of the gas displacement oscillations in the  $x$  direction is a second important length scale, which typically is much smaller than both the wavelength and the length of the channel.

In the direction  $y$  perpendicular to the gas displacement oscillations, one key length scale is the thermal penetration depth

$$\delta_\kappa = \sqrt{2k/\omega\rho c_p} = \sqrt{2\kappa/\omega}, \quad (1)$$

where  $k$  is the thermal conductivity of the gas,  $\rho$  is its density,  $c_p$  is its isobaric heat capacity per unit mass,  $\kappa$  is its thermal diffusivity, and  $\omega = 2\pi f$  is the angular frequency. The thermal penetration depth is approximately the distance that heat diffuses through the gas in a time  $1/\pi f$ . Gas much farther than this from the nearest solid surface experiences adiabatic oscillations, and will not participate in thermoacoustic effects. Closely related to the thermal penetration depth is the viscous penetration depth

$$\delta_\nu = \sqrt{2\mu/\omega\rho} = \sqrt{2\nu/\omega}, \quad (2)$$

where  $\mu$  is viscosity and  $\nu$  is kinematic viscosity. Within  $\delta_\nu$  of solid surfaces, viscous shear forces cause gradients in the oscillating velocity and displacement. In gas mixtures, a third key length scale is the mass-diffusion penetration depth

$$\delta_D = \sqrt{2D/\omega}, \quad (3)$$

where  $D$  is the binary mass diffusion coefficient (called  $D_{12}$  in some literature).

The Prandtl number

$$\sigma = \mu c_p/k = (\delta_\nu/\delta_\kappa)^2 \quad (4)$$

is a dimensionless measure of the ratio of viscous to thermal effects, which is near  $2/3$  for pure monatomic ideal gases and significantly smaller for some gas mixtures. A second dimensionless number,

$$L = k/\rho c_p D = (\delta_\kappa/\delta_D)^2, \quad (5)$$

is a measure of the ratio of thermal to mass-diffusion effects, and is also of order one. Simple ideal-gas kinetic theory predicts  $\sigma$  and  $L$  independent of pressure and temperature, which is close to experimental observation.<sup>14</sup>

Since  $\delta_\nu$ ,  $\delta_\kappa$ , and  $\delta_D$  are all of comparable size in gases, we can expect that viscous effects and mass-diffusion effects may be important whenever thermoacoustic effects are important.

## II. THE OSCILLATING VARIABLES

We consider sound propagating in the  $x$  direction in a uniformly mixed two-component ideal gas within a channel with constant cross sectional area  $A$  and hydraulic<sup>15</sup> radius  $r_h$  much larger than the viscous, thermal, and diffusion penetration depths but much smaller than the acoustic wavelength. We adopt the common<sup>8</sup> complex notation for time-oscillating quantities (pressure  $p$ , temperature  $T$ , vector velocity  $\mathbf{v}$  with component  $u$  parallel to  $x$  and component  $v$  perpendicular to  $x$ , density  $\rho$ , mass fraction  $c$ , entropy per unit mass  $s$ ):

$$p = p_m + \text{Re}[p_1(x)e^{i\omega t}] + \dots, \quad (6)$$

$$u = \text{Re}[u_1(x,y)e^{i\omega t}] + \dots, \quad (7)$$

$$T = T_m + \text{Re}[T_1(x,y)e^{i\omega t}] + \dots, \quad (8)$$

$$\rho, c, s, \text{ etc.} = \text{similar to } T, \quad (9)$$

$$\mathbf{v}, v = \text{similar to } u. \quad (10)$$

In this monofrequency, steady-state acoustic approximation, all the time dependence appears in the factor  $e^{i\omega t}$ . The mean values (subscript  $m$ ) are real, but the small amplitudes (subscript 1) are in general complex to account for the time phasing of the oscillating quantities. The coordinate  $y$  measures the distance from the wall.

To establish notation and method, we begin by deriving the well-known  $y$  dependence of the gas velocity,<sup>16</sup> using the  $x$ -component of the momentum equation, for which our acoustic approximation is

$$i\omega\rho_m u_1 = -\frac{dp_1}{dx} + \frac{\mu}{\rho_m} \frac{\partial^2 u_1}{\partial y^2}. \quad (11)$$

The momentum equation for a gas mixture is identical to that of a pure gas. The  $x$  derivatives of  $u_1$  have been neglected because they are of order  $u_1/\lambda$ , and hence are much smaller than the  $y$  derivatives, of order  $u_1/\delta_\nu$ . Equation (11) is an ordinary differential equation for  $u_1(y)$ . With boundary condition  $u_1(0) = 0$  at the solid surface, its boundary-layer solution is

$$u_1 = \frac{i}{\omega\rho_m} [1 - e^{-(1+i)y/\delta_\nu}] \frac{dp_1}{dx}. \quad (12)$$

Later, we will need the spatial average of Eq. (12) over the cross section  $A$  of the channel:

$$\langle u_1 \rangle = \frac{i}{\omega\rho_m} (1 - f_\nu) \frac{dp_1}{dx}, \quad (13)$$

where  $\langle \rangle$  denotes the spatial average over  $A$  and

$$f_\nu = (1 - i)\delta_\nu/2r_h \quad (14)$$

is the spatial average of the exponential. Combining Eqs. (12) and (13) gives another useful expression for the velocity,

$$u_1 = \frac{\langle u_1 \rangle}{1 - f_\nu} [1 - e^{-(1+i)y/\delta_\nu}]. \quad (15)$$

To find the dependence of the oscillating temperature  $T$  on  $y$  is complicated in a gas mixture, because thermal diffusion couples the oscillations of temperature and concentration. Following Landau and Lifshitz,<sup>17</sup> let the concentration  $c$  be the local mass fraction of the lighter component; i.e.,  $c$  is the ratio of the mass of the component with the lower molecular weight to the total mass of gas, per unit volume. Then the convective mass flux density of this component is  $\rho c \mathbf{v}$ , and the diffusive mass flux density of this component is

$$\mathbf{i} = -\rho D[\nabla c + (k'_T/T)\nabla T], \quad (16)$$

with barodiffusion neglected. The diffusion coefficient  $D$  gives diffusion in response to a concentration gradient, and the thermal diffusion ratio  $k'_T$  gives the diffusion in response to a temperature gradient. Using Eq. (16) with Eq. (57.3) of Landau and Lifshitz,

$$\rho(\partial c/\partial t + \mathbf{v} \cdot \nabla c) = -\nabla \cdot \mathbf{i}, \quad (17)$$

to eliminate  $\mathbf{i}$  yields

$$(\partial c/\partial t + \mathbf{v} \cdot \nabla c) = \nabla \cdot [D\nabla c + (Dk'_T/T)\nabla T]. \quad (18)$$

This equation shows that the concentration at a point changes in time due to convection of a concentration gradient past that point plus diffusion caused by both a concentration gradient and a temperature gradient. Using Eqs. (6)–(10) for all variables, keeping terms to first order, and realizing that  $dc_m/dx = 0$  for a well-mixed gas, Eq. (18) becomes simply

$$c_1 = \frac{\delta_D^2}{2i} \left[ \frac{\partial^2 c_1}{\partial y^2} + \frac{k'_T}{T_m} \frac{\partial^2 T_1}{\partial y^2} \right]. \quad (19)$$

To examine oscillating heat transfer in the mixture, we begin by combining Eqs. (57.6) and (58.12) of Landau and Lifshitz, eliminating  $\mathbf{q} - g\mathbf{i}$ , substituting Eqs. (6)–(10), and keeping terms to first order:

$$\begin{aligned} \rho_m T_m \left( i\omega s_1 + u_1 \frac{ds_m}{dx} \right) \\ = k \frac{\partial^2 T_1}{\partial y^2} - \left[ k'_T \left( \frac{\partial g}{\partial c} \right)_{p,T} - T_m \left( \frac{\partial g}{\partial T} \right)_{p,c} \right] \nabla \cdot \mathbf{i}_1, \end{aligned} \quad (20)$$

where  $\mathbf{q}$  is the heat flux density and  $g$  is the chemical potential per unit mass. Equation (17) shows that  $\nabla \cdot \mathbf{i}_1 = -i\omega\rho_m c_1$ . We have  $ds_m/dx = 0$  in the present simple problem, although this will not be the case when  $dT_m/dx \neq 0$  in the stacks of thermoacoustic engines and refrigerators, nor when  $dc_m/dx \neq 0$  in apparatus with substantial net mixture separation. We eliminate  $s_1$  using

$$ds = \left( \frac{\partial s}{\partial T} \right)_{p,c} dT + \left( \frac{\partial s}{\partial c} \right)_{p,T} dc + \left( \frac{\partial s}{\partial p} \right)_{T,c} dp \quad (21)$$

$$= \frac{c_p}{T} dT - \left( \frac{\partial g}{\partial T} \right)_{p,c} dc - \frac{1}{\rho T} dp, \quad (22)$$

where we have used two Maxwell relations and the ideal-gas equation of state. With these substitutions, Eq. (20) becomes

$$T_1 = \frac{p_1}{\rho_m c_p} + \frac{\varepsilon T_m c_1}{k'_T} + \frac{\delta_\kappa^2}{2i} \frac{\partial^2 T_1}{\partial y^2}, \quad (23)$$

using the definition

$$\varepsilon = \frac{(k'_T)^2}{T_m c_p} \left( \frac{\partial g}{\partial c} \right)_{p,T} \quad (24)$$

for future simplicity.

Equations (19) and (23) comprise two coupled differential equations in the unknown functions  $c_1(y)$  and  $T_1(y)$ . In general, the solid has sufficient heat capacity and thermal conductivity to enforce

$$T_1(0) = 0 \quad (25)$$

on the gas at the solid surface, so this provides one boundary condition for the solution. The other boundary condition is obtained from the fact that, absent condensation and evaporation at the solid, the first-order concentration flux density perpendicular to the wall must be zero, which yields

$$\left. \frac{\partial c_1}{\partial y} \right|_0 + \frac{k'_T}{T_m} \left. \frac{\partial T_1}{\partial y} \right|_0 = 0. \quad (26)$$

Equations (19) and (23) are very similar to Eqs. (58.14) and (58.15) of Landau and Lifshitz, and are simplified versions of Eqs. (5) and (3) of Raspet *et al.*<sup>6</sup> if we also use our results for  $\varepsilon$  and  $k'_T$  (see next section). However, the present problem differs significantly from that of Raspet *et al.* in two ways. First, our boundary condition Eq. (26) allows no flux of either component into the wall, while their “wet” boundary condition allows flux of their condensing component into the wall by ensuring that the partial pressure of the condensing component is constant at the wall. Second, we keep the  $k'_T$  term in Eq. (19) while they neglect it in their Eq. (28).

To solve Eqs. (19) and (23), subject to the boundary conditions given by Eqs. (25) and (26), we can use Eq. (23) to eliminate  $c_1$  from the other equations, obtaining a fourth-order differential equation for  $T_1$  with two boundary conditions:

$$T_1 = \frac{p_1}{\rho_m c_p} + \frac{1}{2i} [\delta_\kappa^2 + (1 + \varepsilon) \delta_D^2] \frac{\partial^2 T_1}{\partial y^2} + \frac{\delta_\kappa^2 \delta_D^2}{4} \frac{\partial^4 T_1}{\partial y^4}, \quad (27)$$

$$T_1(0) = 0, \quad (28)$$

$$(1 + \varepsilon) \left. \frac{\partial T_1}{\partial y} \right|_0 - \frac{\delta_\kappa^2}{2i} \left. \frac{\partial^3 T_1}{\partial y^3} \right|_0 = 0. \quad (29)$$

An additional boundary condition is simply that  $T_1$  must remain finite as  $y \rightarrow \infty$ . The solution is

$$T_1 = \frac{p_1}{\rho_m c_p} [1 - C e^{-(1+i)y/\delta_{\kappa D}} - (1-C) e^{-(1+i)y/\delta_{D\kappa}}], \quad (30)$$

where

$$\delta_{\kappa D}^2 = \frac{1}{2} \delta_\kappa^2 [1 + (1 + \varepsilon)/L + \sqrt{[1 + (1 + \varepsilon)/L]^2 - 4/L}], \quad (31)$$

$$\delta_{D\kappa}^2 = \frac{1}{2} \delta_\kappa^2 [1 + (1 + \varepsilon)/L - \sqrt{[1 + (1 + \varepsilon)/L]^2 - 4/L}], \quad (32)$$

$$C = \frac{\sqrt{L} \delta_{\kappa D} - \delta_{D\kappa}}{(1 + \sqrt{L})(\delta_{\kappa D} - \delta_{D\kappa})}, \quad (33)$$

which can be verified with modest difficulty by direct substitution into Eqs. (27)–(29). [The identities

$$\frac{\delta_{\kappa D}^2}{\delta_\kappa^2} + \frac{\delta_{D\kappa}^2}{\delta_\kappa^2} = 1 + \frac{1 + \varepsilon}{L}, \quad (34)$$

$$L \delta_{\kappa D}^2 \delta_{D\kappa}^2 = \delta_\kappa^4, \quad (35)$$

obtained by manipulating Eqs. (31) and (32), and the algebra identity

$$\delta_{\kappa D}^3 \pm \delta_{D\kappa}^3 = (\delta_{\kappa D} \pm \delta_{D\kappa})(\delta_{\kappa D}^2 + \delta_{D\kappa}^2 \mp \delta_{\kappa D} \delta_{D\kappa}), \quad (36)$$

are useful when working through some of the tedious steps in the derivations in this paper.] Note that  $\varepsilon \rightarrow 0$  recovers the usual thermoacoustic solution: When  $L \geq 1$ ,  $\delta_{\kappa D} \rightarrow \delta_\kappa$ ,  $\delta_{D\kappa} \rightarrow \delta_D$ , and  $C \rightarrow 1$ ; or when  $L < 1$ ,  $\delta_{\kappa D} \rightarrow \delta_D$ ,  $\delta_{D\kappa} \rightarrow \delta_\kappa$ , and  $C \rightarrow 0$ . The spatial average of the temperature over the cross-sectional area of the channel is

$$\langle T_1 \rangle = \frac{p_1}{\rho_m c_p} [1 - C f_{\kappa D} - (1 - C) f_{D\kappa}], \quad (37)$$

where

$$f_{\kappa D} = (1 - i) \delta_{\kappa D} / 2r_h \quad (38)$$

and similarly for  $f_{D\kappa}$ .

### III. TYPICAL VALUES

To present some typical numerical values, we consider He–Ar and He–Xe mixtures, which are of interest in thermoacoustic refrigerators.<sup>11,12</sup> Although the derivation elsewhere in this paper follows Landau and Lifshitz’s notational preference for mass fraction  $c$ , most data are tabulated in terms of mole fraction  $n$ ; the two are related by

$$c = \frac{n_L m_L}{n_L m_L + (1 - n_L) m_H}, \quad (39)$$

where  $m$  is molar mass and the subscripts refer to the lighter and heavier species. We use viscosity and thermal conductivity calculations from Giacobbe<sup>18</sup> at 20 °C, and mass diffusion coefficients interpolated to 20 °C from the measurements of Srivastava<sup>19</sup>, which are in good agreement with calculations. We also include the weak concentration dependence of  $D$  according to the recommendations of Chapman and Cowling.<sup>20</sup> Based on these data,  $\sigma$  and  $L$  are shown in Fig. 2(a).

Next we need  $\varepsilon$ , which requires evaluation of  $(\partial g / \partial c)_{p,T}$ . Landau and Lifshitz suggest how to proceed. The chemical potential  $g$  (per unit mass) is

$$g = \hat{g}_L / m_L - \hat{g}_H / m_H, \quad (40)$$

where the caret indicates a molar chemical potential. For ideal gases, we have

$$\hat{g}_L = \hat{g}_{L,\text{pure}} + RT \ln n_L, \quad (41)$$

$$\hat{g}_H = \hat{g}_{H,\text{pure}} + RT \ln(1 - n_L), \quad (42)$$



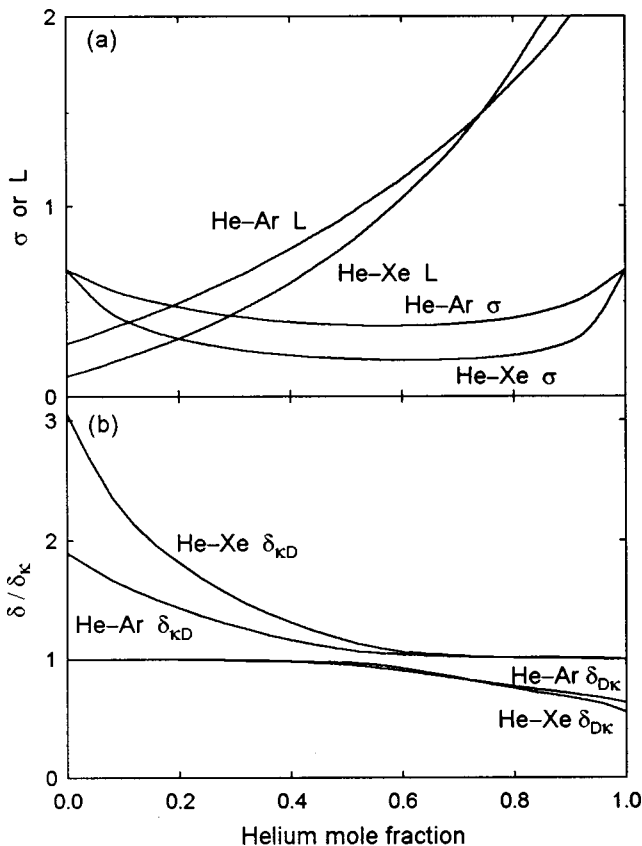


FIG. 2. Some relevant properties of He-Ar and He-Xe mixtures. (a) Values of  $\sigma$  and  $L$ , giving the ratios  $\delta_v^2/\delta_\kappa^2$  and  $\delta_\kappa^2/\delta_D^2$ , respectively. (b) The ratios  $\delta_{\kappa D}/\delta_\kappa$  and  $\delta_{D\kappa}/\delta_\kappa$ , which appear throughout our calculations. Each of these approaches 1 and  $1/\sqrt{L}$  in the two pure-gas limits.

with  $R$  the universal gas constant. Using Eqs. (39)–(42), it is straightforward to compute

$$\left(\frac{\partial g}{\partial c}\right)_{p,T} = \left(\frac{\partial g}{\partial n_L}\right)_{p,T} / \left(\frac{\partial c}{\partial n_L}\right)_{p,T} = \frac{RT}{c(1-c)[m_L + c(m_H - m_L)]}. \quad (43)$$

We must also convert from Landau and Lifshitz's thermal diffusion ratio  $k'_T$  to the thermal diffusion ratio  $k_T$  used in most other treatments. These differ by a factor of  $dc/dn_L$ , so that

$$k'_T = k_T \frac{m_L m_H}{[n_L m_L + (1 - n_L) m_H]^2} = k_T \frac{[m_L + (m_H - m_L)c]^2}{m_L m_H}. \quad (44)$$

Combining Eqs. (24), (43), and (44) we have finally

$$\varepsilon = \frac{\gamma - 1}{\gamma} \frac{k_T^2}{n_L(1 - n_L)}, \quad (45)$$

where  $\gamma$  is the ratio of isobaric to isochoric specific heats.

For  $k_T$ , we use the experimental data of Atkins *et al.*,<sup>21</sup> ranging over helium mole fractions  $0.1 \leq n_L \leq 0.5$ . To interpolate and extrapolate elsewhere, we fit their data with  $k_T = 0.38 n_L^{1.2} (1 - n_L)^{0.8}$  for He-Ar and  $k_T = 0.40 n_L^{1.3} (1 - n_L)^{0.7}$  for He-Xe, which give better fits to the data than

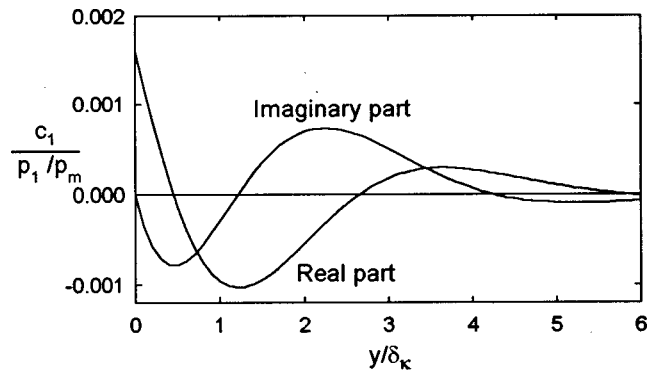


FIG. 3. Real and imaginary parts of  $c_1$ , normalized by  $p_1/p_m$ , for a He-Xe mixture with  $n_L = 0.50$ . For this mixture,  $\delta_\mu/\delta_\kappa = 0.445$ ,  $\delta_{\kappa D}/\delta_\kappa = 1.16$ , and  $\delta_{D\kappa}/\delta_\kappa = 0.97$ .

the functional form  $n_L(1 - n_L)$  suggested by the simplest kinetic theory. Figure 2(b) shows the resulting values of  $\delta_{\kappa D}$  and  $\delta_{D\kappa}$ . Clearly Fig. 1 was unrealistically naive. Heat and mass diffusion are so intimately linked that the length scales  $\delta_{\kappa D}$  and  $\delta_{D\kappa}$  appearing in  $T_1(y)$  are very different from the familiar  $\delta_\kappa$ . Hence, it is also clear that correct calculations of enthalpy flux, proportional to  $\text{Re}[T_1 \tilde{u}_1]$ , must include thermal diffusion. (The tilde denotes complex conjugation.)

We can use Eqs. (23) and (30) to obtain  $c_1(y)$ , which is plotted in Fig. 3 for a 50–50 He-Xe mixture. The imaginary part is negative for  $y \leq \delta_\kappa$  and is positive for  $\delta_\kappa \leq y \leq 4\delta_\kappa$ , which is qualitatively consistent with the phenomena we described in Fig. 1, where the thermal diffusion occurs during pressure extrema of Fig. 1(a) and (c) so that the mixture is most separated in Fig. 1(b) and (d). However, the large  $\text{Re}[c_1]$  in Fig. 3 shows again that this oscillating thermal diffusion is much richer than we had anticipated in Fig. 1.

#### IV. BOUNDARY-LAYER MASS SEPARATION

The phased oscillating phenomena described above cause a time-averaged mass separation, whose origin is easy to understand qualitatively. The oscillating temperature gradient  $\partial T_1/\partial y$  near the wall causes nonzero concentration oscillations near the wall. At this distance from the wall, comparable to  $\delta_v$ , the oscillating velocity  $u_1$  in the  $x$  direction depends on  $y$ . If the time phasing is favorable, this produces a time-averaged mass flux density  $\overline{\rho c u}$  of the lighter component and  $\overline{\rho(1-c)u}$  of the heavier component in the  $x$  direction, while the total mass flux density  $\overline{\rho u}$  remains zero. (The overbar signifies time average.) Exaggerating the magnitudes of the effects, we could imagine that during one half of the cycle the light component would be stuck deep in the viscous boundary layer and the heavy component would be free to move outside the viscous boundary layer; during the other half of the cycle the roles would be reversed, with the heavy component immobilized by viscosity and the lighter component free to move. If the velocity were phased correctly with respect to this oscillating concentration, the time-averaged result would be flux of the light component in one direction along  $x$  and flux of the heavy component in the opposite direction.

The time-averaged second-order mass flux of the lighter component is

$$A\langle\overline{\rho cu}\rangle_2 = \frac{A\rho_m}{2} \operatorname{Re}[\langle c_1\tilde{u}_1\rangle]. \quad (46)$$

To evaluate the right-hand side, we use Eq. (23) for  $c_1(y)$ , Eq. (30) for  $T_1(y)$ , and Eq. (15) for  $u_1(y)$ . The spatial average  $\langle c_1\tilde{u}_1\rangle$  involves boundary-layer-approximation integrals of the form

$$\frac{1}{r_h} \int_0^{r_h} e^{-(1+i)y/\delta} [1 - e^{-(1-i)y/\delta}] dy = \frac{(\delta - \delta_v) - i(\delta + \delta_v)}{2r_h(1 + \delta_v^2/\delta^2)}, \quad (47)$$

where  $\delta$  is either  $\delta_{\kappa D}$  or  $\delta_{D\kappa}$ . After tedious algebra leading to an intermediate result,

$$\begin{aligned} \langle\overline{\rho cu}\rangle_2 &= \frac{\delta_\kappa}{4r_h} \frac{k'_T/\varepsilon}{c_p T_m} \operatorname{Re} \left[ \frac{p_1\langle\tilde{u}_1\rangle}{(1-\tilde{f}_v)} \left\{ C \left( 1 - \frac{\delta_{\kappa D}^2}{\delta_\kappa^2} \right) \right. \right. \\ &\quad \times \frac{(\delta_{\kappa D}/\delta_\kappa - \sqrt{\sigma}) - i(\delta_{\kappa D}/\delta_\kappa + \sqrt{\sigma})}{\delta_{\kappa D}^2/\delta_\kappa^2 + \sigma} + (1-C) \\ &\quad \left. \left. \times \left( 1 - \frac{\delta_{D\kappa}^2}{\delta_\kappa^2} \right) \frac{(\delta_{D\kappa}/\delta_\kappa - \sqrt{\sigma}) - i(\delta_{D\kappa}/\delta_\kappa + \sqrt{\sigma})}{\delta_{D\kappa}^2/\delta_\kappa^2 + \sigma} \right\} \right], \quad (48) \end{aligned}$$

the final result is

$$A\langle\overline{\rho cu}\rangle_2 = \frac{\delta_\kappa}{4r_h} \frac{k'_T}{c_p T_m} (F_{\text{trav}} \operatorname{Re}[p_1\tilde{U}_1] + F_{\text{stand}} \operatorname{Im}[p_1\tilde{U}_1]) \quad (49)$$

to lowest order in  $\delta/r_h$ , where  $U_1 = A\langle u_1\rangle$  is the volume flow rate. The traveling-wave and standing-wave factors are given by

$$\begin{aligned} F_{\text{trav}} &= \frac{\sigma\sqrt{\sigma L} - \sqrt{\sigma} - \sigma\sqrt{L}(\delta_{\kappa D}/\delta_\kappa + \delta_{D\kappa}/\delta_\kappa)}{(1 + \sqrt{L})[1 + \sigma(1 + L + \varepsilon + \sigma L)]}, \quad (50) \\ F_{\text{stand}} &= \frac{-\sigma\sqrt{\sigma L} + \sqrt{\sigma} - \sigma\sqrt{L}(\delta_{\kappa D}/\delta_\kappa + \delta_{D\kappa}/\delta_\kappa)}{(1 + \sqrt{L})[1 + \sigma(1 + L + \varepsilon + \sigma L)]}. \quad (51) \end{aligned}$$

Equation (49) can be rewritten in terms of molar quantities:

$$\dot{N}_{L,2} = \frac{\delta_\kappa}{4r_h} \frac{\gamma - 1}{\gamma} \frac{k'_T}{RT_m} (F_{\text{trav}} \operatorname{Re}[p_1\tilde{U}_1] + F_{\text{stand}} \operatorname{Im}[p_1\tilde{U}_1]), \quad (52)$$

where  $\dot{N}_{L,2}$  is the rate at which moles of the light component move in the  $x$  direction. [Note  $k'_T$  in Eq. (49);  $k_T$  in Eq. (52).]

Figure 4 shows  $k_T F$  for He–Ar and He–Xe mixtures. The factors in Eq. (52) can be grouped into dimensionless ratios, showing that the average molar separation flux density in the channel scales like the product of  $p_1/p_m$ ,  $\langle u_1\rangle/a$ ,  $a$ ,  $\delta_\kappa/r_h$ ,  $(\gamma - 1)/4\gamma$ , and the molar density  $N/V$ . The details, captured in  $k_T F_{\text{trav}}$  and  $k_T F_{\text{stand}}$ , reduce the magnitude of the effect by roughly  $10^{-2}$ , as shown in Fig. 4. Hence, with  $(\gamma - 1)/4\gamma \sim 10^{-1}$ , the molar separation flux in a short porous medium having  $r_h \sim \delta_\kappa$  could be of order  $10^{-3} M^2 (N/V) a A$

(where  $M \sim |p_1|/p_m \sim |\langle u_1\rangle|/a$  is the acoustic Mach number). If such a short porous medium filled the cross section of a chamber whose length was of the order of the acoustic wavelength, and if nonzero  $dc_m/dx$  did little to change this mass separation rate, then a substantial concentration difference could establish itself in a time of order  $1/10^{-3} M^2 f$ , where  $f$  is the frequency of the wave. Hence, it seems likely that thermoacoustic refrigerators using gas mixtures might have concentration differences across their stacks, if bulk gas motion such as convection or streaming does not re-mix the gases with sufficient vigor.

Figure 4 also shows that traveling-wave phasing should be more effective than standing-wave phasing at separating the mixture, reconfirming that our initial view of this process, illustrated in Fig. 1, was much too naive.

## V. BOUNDARY-LAYER ACOUSTIC POWER DISSIPATION

To find the time-averaged acoustic power  $d\dot{E}_2$  dissipated in a length  $dx$  of the channel, we write

$$\frac{d\dot{E}}{dx} = -A \frac{d\langle pu\rangle}{dx}. \quad (53)$$

Expressing Eq. (53) in complex notation and expanding the derivative gives

$$\frac{d\dot{E}_2}{dx} = -\frac{1}{2} A \operatorname{Re} \left[ \langle \tilde{u}_1 \rangle \frac{dp_1}{dx} + \tilde{p}_1 \frac{d\langle u_1 \rangle}{dx} \right]. \quad (54)$$

We can obtain  $dp_1/dx$  from Eq. (13) above. To find  $d\langle u_1\rangle/dx$ , we use the continuity equation  $\partial\rho/\partial t + \nabla \cdot (\rho\mathbf{v}) = 0$ , which can be averaged with respect to  $y$  in our acoustic approximation to obtain

$$i\omega\langle\rho_1\rangle + \rho_m d\langle u_1\rangle/dx = 0. \quad (55)$$

Using  $d\rho = -(\rho/T) dT + (\gamma/a^2) dp$ , we can express the spatially averaged density as  $\langle\rho_1\rangle = -(\rho_m/T_m)\langle T_1\rangle + (\gamma/a^2)p_1$ . Substituting this into Eq. (55), using Eq. (37) for  $\langle T_1\rangle$ , and eliminating  $c_p$  by means of the thermodynamic identity  $\gamma - 1 = a^2/Tc_p$  yields

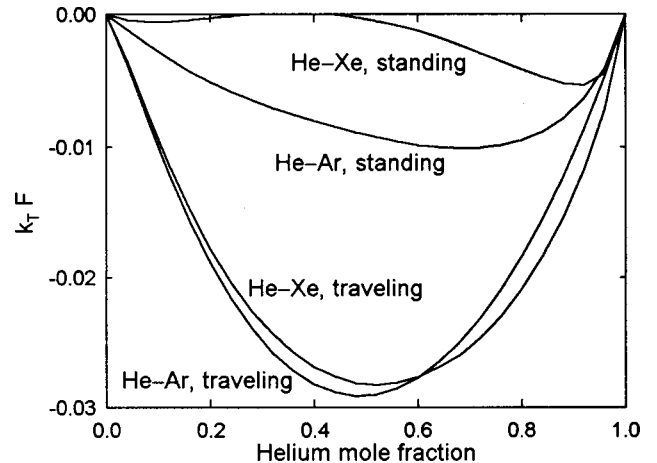


FIG. 4. Dimensionless molar-separation-flux parameters  $k_T F_{\text{trav}}$  and  $k_T F_{\text{stand}}$  for He–Ar and He–Xe mixtures.

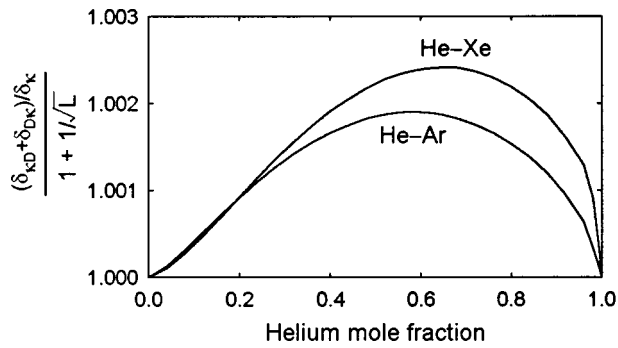


FIG. 5. The factor by which thermal diffusion multiplies boundary-layer thermal-relaxation dissipation of acoustic power, for He–Ar and He–Xe mixtures.

$$i\omega\{1+(\gamma-1)[Cf_{\kappa D}+(1-C)f_{D\kappa}]\}p_1 + \rho_m a^2 d\langle u_1 \rangle / dx = 0 \quad (56)$$

as an acoustic expression of the continuity equation which can be solved for  $d\langle u_1 \rangle / dx$ .

Finally, substituting Eqs. (56) and (13) into Eq. (54), we obtain

$$\frac{d\dot{E}_2}{dx} = \frac{A\omega}{2} \left[ \frac{\rho_m |\langle u_1 \rangle|^2}{|1-f_\nu|^2} \text{Im}[-f_\nu] + \frac{(\gamma-1)|p_1|^2}{\rho_m a^2} \text{Im}[-Cf_{\kappa D} - (1-C)f_{D\kappa}] \right]. \quad (57)$$

The first term gives the viscous damping of sound and the second term gives the complicated thermal damping in the presence of thermal diffusion. Using the definition<sup>15</sup> of hydraulic radius, Eq. (57) reduces to

$$\frac{d\dot{E}_2}{dS} = \frac{1}{4} \rho_m |\langle u_1 \rangle|^2 \omega \delta_\nu + \frac{1}{4} \frac{|p_1|^2}{\rho_m a^2} (\gamma-1) \omega \frac{(\delta_{\kappa D} + \delta_{D\kappa})}{1+1/\sqrt{L}}, \quad (58)$$

to lowest order in the  $\delta$ 's, where  $S$  is the surface area of the channel. The first term is the familiar<sup>8</sup> boundary-layer viscous dissipation per unit surface area; it is unchanged by thermal diffusion. The second term represents the combined dissipative effects of heat and mass diffusion. The limit  $\epsilon \rightarrow 0$  recovers the usual<sup>8</sup> thermoacoustic solution, with the final fraction in the second term of Eq. (58) reducing simply to  $\delta_\kappa$ . (The similar-looking result in Ref. 2 does not recover the usual thermoacoustic solution when  $k_T \rightarrow 0$ .)

Figure 5 displays this final fraction in the second term of Eq. (58) for He–Ar and He–Xe mixtures, divided by  $\delta_\kappa$ , so this figure shows the ratio of the present  $p^2$  dissipation term to the pure-gas  $p^2$  term involving only  $\delta_\kappa$ . The extra dissipation is less than one percent for these mixtures, so it is unlikely that this effect would have been noticed in measurements to date with thermoacoustic refrigerators. However, such refrigerators operate with a nonzero  $dT_m/dx$ , and presumably also with a nonzero  $dc_m/dx$ ; these effects might increase the mass-diffusion dissipation.

## ACKNOWLEDGMENTS

This work has been supported by the Office of Basic Energy Sciences in the U.S. Department of Energy. We thank Bill Ward and Mike Hayden for encouraging us to take this idea seriously. We are very grateful to Rich Raspet for quickly and kindly identifying a serious error in an earlier version of this paper.

- <sup>1</sup>P. S. Spoor and G. W. Swift, "Mode locking of acoustic resonators and its application to vibration cancellation in acoustic heat engines," *J. Acoust. Soc. Am.* **106**, 1353–1362 (1999).
- <sup>2</sup>A. M. Dykhne, A. F. Pal', V. D. Pis'mennyi, V. V. Starostin, and M. D. Taran, "Dynamics of gas-mixtures separation in the field of a sound wave," *Zh. Eksp. Teor. Fiz.* **88**, 1976–1983 (1985). [English translation: *Sov. Phys. JETP* **61**, 1171–1175 (1985).]
- <sup>3</sup>V. B. Bozhdankevich, A. M. Dykhne, A. F. Pal', V. D. Pis'mennyi, V. V. Pichugin, A. N. Starostin, and M. D. Taran, "Thermofusion separation of gas mixtures in the field of a sound wave," *Dokl. Akad. Nauk SSSR* **288**, 605–608 (1986). [English translation: *Sov. Phys. Dokl.* **31**, 428–430 (1986).]
- <sup>4</sup>W. L. M. Nyborg, "Acoustic streaming," in *Physical Acoustics, Volume IIB*, edited by W. P. Mason (Academic, New York, 1965), pp. 265–331.
- <sup>5</sup>G. W. Howell, "Separation of isotopes by oscillatory flow," *Phys. Fluids* **31**, 1803–1805 (1988).
- <sup>6</sup>R. Raspet, C. J. Hickey, and J. M. Sabatier, "The effect of mass transfer on sound propagation in cylindrical tubes using the low reduced frequency approximation," *J. Acoust. Soc. Am.* **105**, 65–73 (1999).
- <sup>7</sup>D. M. Ruthven, Ed., *Encyclopedia of Separation Technology* (Wiley, New York, 1997).
- <sup>8</sup>G. W. Swift, "Thermoacoustic engines," *J. Acoust. Soc. Am.* **84**, 1145–1180 (1988).
- <sup>9</sup>R. S. Reid, W. C. Ward, and G. W. Swift, "Cyclic thermodynamics with open flow," *Phys. Rev. Lett.* **80**, 4617–4620 (1998).
- <sup>10</sup>K. F. Herzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves* (Academic, New York, 1959).
- <sup>11</sup>S. L. Garrett, J. A. Adeff, and T. J. Hoffer, "Thermoacoustic refrigerator for space applications," *AAIA J. Thermophys. Heat Trans.* **7**, 595–599 (1993).
- <sup>12</sup>M. E. Poese, "Performance measurements on a thermoacoustic refrigerator driven at high amplitudes, Master's thesis, The Pennsylvania State University, 1998. Applied Research Laboratory.
- <sup>13</sup>W. C. Ward and G. W. Swift, "Design environment for low amplitude thermoacoustic engines (DeltaE)," *J. Acoust. Soc. Am.* **95**, 3671–3672 (1994). Fully tested software and user's guide available from Energy Science and Technology Software Center, U.S. Department of Energy, Oak Ridge, Tennessee. To review DeltaE's capabilities, visit the Los Alamos thermoacoustics web site at [www.lanl.gov/thermoacoustics/](http://www.lanl.gov/thermoacoustics/). For a beta-test version, contact [ww@lanl.gov](mailto:ww@lanl.gov) (Bill Ward) or [swift@lanl.gov](mailto:swift@lanl.gov) (Greg Swift) by email.
- <sup>14</sup>J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).
- <sup>15</sup>The hydraulic radius is the ratio of gas volume to gas-solid contact surface area.
- <sup>16</sup>N. Rott, "Damped and thermally driven acoustic oscillations in wide and narrow tubes," *Z. Angew. Math. Phys.* **20**, 230–243 (1969).
- <sup>17</sup>L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1982).
- <sup>18</sup>F. W. Giacobbe, "Estimation of Prandtl numbers in binary mixtures of helium and other noble gases," *J. Acoust. Soc. Am.* **96**, 3568–3580 (1994).
- <sup>19</sup>K. P. Srivastava, "Mutual diffusion of binary mixtures of helium, argon, and xenon at different temperatures," *Physica (Utrecht)* **25**, 571–578 (1959).
- <sup>20</sup>S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, Cambridge, 1939).
- <sup>21</sup>B. E. Atkins, R. E. Bastick, and T. L. Ibbs, "Thermal diffusion in mixtures of inert gases," *Proc. R. Soc. London, Ser. A* **172**, 142–158 (1939).

# A theoretical study of duct noise control by flexible panels

Lixi Huang

*Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong*

(Received 5 January 1999; accepted for publication 12 June 1999)

Theoretical exploration is undertaken for passive noise control by flush-mounted panels in an otherwise rigid duct. For a plane sound wave traveling in the flexible segment, the wall compliance renders a wave speed less than the isentropic speed of sound in air. Scattering and reflection occur at the upstream edge of the panel while the energy flux of the transmitted wave is partitioned between the wall flexural waves and the sound in air. For a lossless panel these waves are scattered and reflected again by the downstream edge forming standing waves responsible for the undesirable passbands. For panels with substantial structural damping, however, both flexural and sound waves diminish with distance, eliminating the passbands. It is estimated that the wave dissipation by panel materials like rubber could outperform typical fibrous duct lining. The combination of wave reflection, dissipation, and slowing down allows broadband, low-frequency noise reduction over a short distance. © 1999 Acoustical Society of America. [S0001-4966(99)00310-0]

PACS numbers: 43.50.Gf, 43.20.Mv, 43.20.Tb [MRS]

## INTRODUCTION

We are interested in devising a passive duct noise control which may work effectively for a broad frequency band and yet poses minimal intrusion for the mean flow inside the duct. In essence, traditional passive noise control takes two forms. One is wave reflection by discontinuity of impedance (defined as the ratio of perturbation pressure to the acoustic volume flow), such as the expansion-chamber-type muffler in the vehicle exhaust pipe, and the other is viscous dissipation, such as fibrous duct lining in air distribution systems. In practice these two often function simultaneously and conditions of resonance are exploited to achieve optimal results. But for the purpose of introducing our idea of noise control by flexible panels, it suffices to highlight the drawbacks of the two techniques in their basic settings.

For the expansion-chamber-type muffler, there are at least three problems: the existence of passbands, the bulkiness, and the aerodynamic loss associated with the flow separation across the duct junctions. To rectify the passband problem one may try a combination of chambers of unequal lengths (Munjal, 1987), but such a technique may aggravate the other two problems. For fibrous duct lining the noise reduction is broadband, but its effectiveness deteriorates rapidly towards low frequencies. In this investigation we seek to devise appropriate interactions between sound and flexible panels sandwiched in the rigid duct walls, as shown in Fig. 1. Noise reduction is achieved by the combination of two basic mechanisms: wave reflection at the upstream edge of the panel and the damping of excited flexural waves by the panel viscosity. First, when waves travel in a passage with considerable cross section distensibility, fluid compressibility gives way to wall compliance as the main mechanism controlling the pace of wave propagation (Lighthill, 1978), and a wave speed ( $c$ ) much less than the isentropic speed of sound in free space ( $c_0$ ) ensues. The change of wave speed from  $c_0$  to  $c \ll c_0$  leads to substantial wave scattering and reflection at the panel edges where the duct cross sections remain constant. The transmitted waves are partitioned between flexural

waves on the panel and sound waves in fluid, and they propagate downstream at the same phase speed. For a very compliant panel, as will be shown in the next section, more energy is imparted to the flexural waves than to the fluid. Damping of structural waves can be much more effective than in fluid, which creates an avenue of using good vibration-absorbing materials. It will also be shown that the coupling between sound and structure is such that there is constant pumping of energy from fluid to structure for damping. The final residual waves in both media can be quite small, minimizing further reflection by the downstream panel edge. This is crucial for avoiding passbands. A similar exchange of wave energy was studied by Ffowcs Williams and Hill (1987) in semi-infinite space, but the dynamics of the wall vibration was decoupled from sound.

We begin by examining the sound propagation in a flexible duct of infinite length, and the resulting axial wave speed will be compared with that in a more realistic configuration of finite panels. The effect of a mean flow on wave propagation and reflection is left to future studies. However, flow-induced vibration on the flexible panel is a potential issue which has to be avoided. In a recent study of biological flow in flexible tubes simulating the upper airway, Huang (1998) examined shear flow instability analytically and came to the conclusion that flexural instability may occur when the maximum flow speed at the duct center exceeds the characteristic wave speed on the wall. In other words, flexural wave instability may be avoided if the flow is everywhere slower than the flexural wave speed. Since in most applications of our primary interest, such as the pipework in a central air conditioning system, the maximum flow speed is probably within 30 m/s. Taking this value as the panel wave speed one may achieve approximately 16 dB noise reduction by reflection at the upstream edge alone, which is an attractive performance. When the stabilizing factor of panel friction is taken into account, the problem of panel instability is expected to diminish while the capacity of noise reduction enhances.



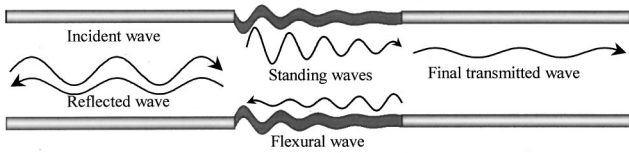


FIG. 1. Illustration of passive noise control by flush-mounted panels through a combination of wave reflection at the upstream edge of the panel and damping of excited flexural waves.

## I. FLEXURAL WAVES IN AN INFINITE CHANNEL

To illustrate the acoustic principle we first limit ourselves to the configuration of a two-dimensional channel with a single lower flexible wall. The word “duct” will also be used in the text, but the three-dimensional effect is not considered. The flexible wall vibrates and radiates sound externally and the so-called “breakout” noise will be important (e.g., Cummings, 1994). However, for the moment we will focus on the wave propagation inside the channel by ignoring the external pressure perturbation. We also assume that the flexible wall is very thin and the only mechanical restoring force against displacement is provided by a tensile stress.

Let the rigid upper wall be located at  $y^* = h^*$  and the flexible lower wall at  $y^* = 0$ . The asterisks denote dimensional variables while the corresponding dimensionless ones will be introduced below without asterisks. The fluid density and isentropic speed of sound are, respectively,  $\rho_0^*$  and  $c_0^*$ . The flexible wall has a mass per unit length  $M^*$ , a damping coefficient  $D^*$ , and a tensile force  $T^*$  is applied. The flexible wall is subject to a traveling wave perturbation of displacement

$$\eta^* = \eta_0^* e^{i(\omega^* t^* - k^* x^*)}.$$

The question is what is the characteristic flexural wave speed  $c^* = \omega^*/k^*$ ?

We first normalize all variables by three basic quantities,  $\rho_0^*$ ,  $h^*$ , and  $c_0^*$ :

$$\begin{aligned} x &= \frac{x^*}{h^*}, & y &= \frac{y^*}{h^*}, & t &= \frac{c_0^* t^*}{h^*}, & \eta_0 &= \frac{\eta_0^*}{h^*}, \\ c &= \frac{c^*}{c_0^*}, & \omega &= \frac{\omega^* h^*}{c_0^*} = k_0, & k &= \frac{\omega}{c}, \\ M &= \frac{M^*}{\rho_0^* h^*}, & c_T &= \frac{\sqrt{T^*/M^*}}{c_0^*}, & D &= \frac{D^*}{\rho_0^* c_0^*}, \\ \phi &= \frac{\phi^*}{c_0^* h^*}, & p &= \frac{p^*}{\rho_0^* (c_0^*)^2}. \end{aligned} \quad (1)$$

Here  $M$ ,  $D$ , and  $c_T$  are, respectively, the mass ratio, the dimensionless damping coefficient, and the *in vacuo* panel wave speed;  $\phi$  and  $p$  are, respectively, the dimensionless perturbation velocity potential and pressure inside the channel. Note that the dimensionless angular frequency  $\omega$  is equal to the dimensionless wave number  $k_0$  and the two will be used interchangeably. Also, they are related to the ratio of

the driving frequency to the first cut-on frequency of the rigid channel by a factor of  $\pi$ . Thus

$$\omega = k_0 = \pi f,$$

where  $f = f^*/f_1^*$  and  $f_1^* = c_0^*/2h^*$ .

The wave equation in fluid becomes

$$\left( -\frac{\partial^2}{c_0^2 \partial t^2} + \nabla^2 \right) \phi = 0,$$

which for harmonic excitation is rewritten as

$$\frac{\partial^2 \phi}{\partial y^2} = k_a^2 \phi,$$

where

$$k_a^2 = k^2 - k_0^2, \quad k_a = \begin{cases} \sqrt{k^2 - k_0^2}, & c < 1, \quad k > k_0, \\ i\sqrt{k^2 - k_0^2}, & c > 1, \quad k < k_0. \end{cases} \quad (2)$$

Here  $k_a$  is real for subsonic waves but imaginary for supersonic waves. The solution satisfying the following rigid wall condition and the kinematic condition on the flexible wall,

$$\left. \frac{\partial \phi}{\partial y} \right|_{y=1} = 0, \quad \left. \frac{\partial \phi}{\partial y} \right|_{y=0} = \frac{\partial \eta}{\partial t} = i\omega \eta,$$

is

$$\phi = -\frac{i\omega \eta}{k_a} \frac{e^{k_a(1-y)} + e^{-k_a(1-y)}}{e^{k_a} - e^{-k_a}}. \quad (3)$$

The pressure perturbation on the upper side of the flexible wall is

$$p_+ = -\left. \frac{\partial \phi}{\partial t} \right|_{y=\eta} = -\frac{\omega^2 \eta}{k_a} \coth k_a, \quad (4)$$

and the dimensionless specific impedance is defined and calculated as

$$\begin{aligned} Z_+ &= \frac{p_+}{\rho_0^* c_0^* d\eta^*/dt^*} = \frac{p_+}{i\omega \eta} \\ &= \begin{cases} +i \frac{k_0}{k_a} \coth k_a, & \text{for } c < 1, \\ -i \frac{k_0}{|k_a|} \cot(|k_a|), & \text{for } c > 1. \end{cases} \end{aligned}$$

For long waves,  $|k_a| \ll 1$ , both  $\coth k_a$  and  $\cot(|k_a|)$  are positive. The effect of fluid loading on the flexible wall dynamics is analogous to added mass for subsonic waves, but for supersonic waves it is the spring stiffness.

Ignoring the bending stiffness, the dynamics of the flexible wall is

$$T \frac{\partial^2 \eta}{\partial x^2} - D \frac{\partial \eta}{\partial t} - p_+ = M \frac{\partial^2 \eta}{\partial t^2}.$$

Inserting the fluid pressure  $p_+$  of Eq. (4) we obtain

$$M\omega^2 - iD\omega + \frac{\omega^2}{k_a} \coth k_a = M c_T^2 k^2. \quad (5)$$

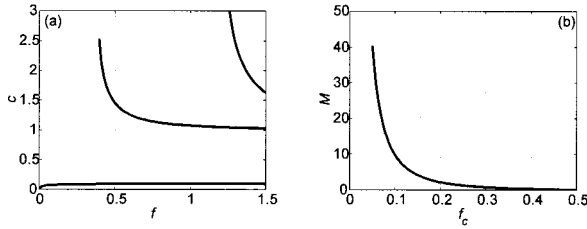


FIG. 2. Eigenwaves on a flexible wall. (a) is the wave speed for  $M=0.41$ ,  $c_T=0.1$ , and (b) gives the critical frequency (in abscissa) above which supersonic waves may prevail.

The characteristic wave speed  $c$  of a lossless wall ( $D=0$ ) is found numerically for each frequency by plotting the difference of the left-hand and right-hand sides of Eq. (5) as a function of wave speed  $c$ . The eigenwave speed for the specific case of  $M=0.41$  and  $c_T=0.1$ , which are derived from a realistic set of dimensional parameters to be used later, is shown in Fig. 2(a). The lower curve represents subsonic wave,  $c < 1$ . As  $f \rightarrow \infty$ ,

$$k_a \rightarrow k_0 \rightarrow \infty, \quad \frac{\coth k_a}{k_a} \rightarrow 0,$$

Eq. (5) becomes  $M = Mc_T^2/c^2$  or  $c = c_T$ ; the wave speed approaches  $c_T$  at high frequencies. Also, there exists a transonic class of eigenwaves. As frequency increases, each branch of the eigenwave speed descends from infinity (supersonic) to unity (sonic) before settling down to a value slightly less than unity (subsonic, outside the frequency range shown in Fig. 2). The frequency of the first (left-most) singularity ( $c \rightarrow \infty$ ) may be called the critical frequency above which supersonic waves may occur. Denote the critical condition by subscript  $c$ . We have  $|k_a| \rightarrow k_{0c} = \pi f_c$ , so that, from Eq. (5),

$$M = \frac{\cot k_{0c}}{k_{0c}},$$

which is shown in Fig. 2(b). The critical frequency  $f_c$  reduces as  $M$  increases.

When the flexible wall has damping, the characteristic wave number is complex ( $k = k_r + ik_i$ ) and the procedure of finding eigenvalues would be somewhat more complicated. Suppose we have negative imaginary part,  $k_i < 0$ , for decaying waves. The work done on the wall by the waves in fluid is  $P = -\frac{1}{2} \text{Re}(p_+ \dot{\eta}_*^*)$  per unit length per unit time, where the asterisk in subscript represents the conjugate of a complex number. Substituting the perturbation pressure of Eq. (4) gives

$$P = \frac{\omega^3}{2} \text{Im} \left( \frac{\coth k_a}{k_a} \right) |\eta|^2.$$

The sign of  $P$  determines the direction of energy flow. Since  $k_a^{-1} \coth k_a$  is an even function of  $k_a$ , we may consider only the positive root of  $k_a = +\sqrt{(k_r^2 - k_i^2 - k_0^2) + 2ik_r k_i}$ . For subsonic waves with slight damping  $k_a$  should be in the fourth quadrant of the complex plane and may be split into real and imaginary parts as  $k_a = k_{ar} - ik_{ai}$ , where both  $k_{ar}$  and  $k_{ai}$  are real and positive. First we look at the phase angle of  $\coth k_a$ , which may be expanded as

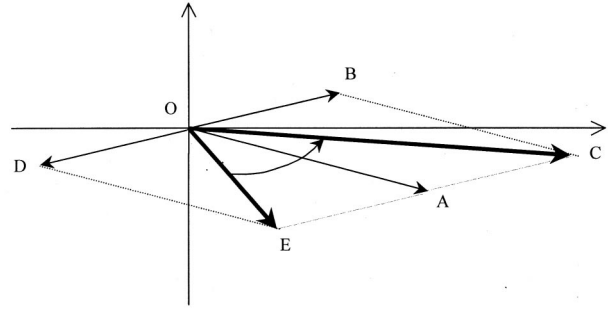


FIG. 3. Illustration of the range of phase angle of  $\coth k_a$  on the complex plane when  $k_a$  has a small negative imaginary part. The phase angle of  $\coth k_a$  is found to be the angle between OE and OC and within the range of  $(0, \pi/2)$ .

$$\coth k_a = \frac{e^{k_a} + e^{-k_a}}{e^{k_a} - e^{-k_a}} = \frac{e^{-ik_{ai}} + e^{-2k_{ar} + ik_{ai}}}{e^{-ik_{ai}} - e^{-2k_{ar} + ik_{ai}}}.$$

The range of phase angle of  $\coth k_a$  is shown in the geometric illustration on the complex plane in Fig. 3. OA represents  $e^{-ik_{ai}}$ , which has a length of unity and is in the fourth quadrant, and OB represents  $e^{-2k_{ar} + ik_{ai}}$ , which has a length less than unity. The numerator ( $e^{-ik_{ai}} + e^{-2k_{ar} + ik_{ai}}$ ) is OC while the denominator ( $e^{-ik_{ai}} - e^{-2k_{ar} + ik_{ai}}$ ) is OE. As quotient of OC/OE, the phase angle of  $\coth k_a$  is the angle from OE to OC, which is positive and less than  $\pi/2$ . Since  $k_a$  has a small negative phase angle, the phase angle of  $k_a^{-1} \coth k_a$  is surely in the range of  $(0, \pi)$ . Since the imaginary part of  $k_a^{-1} \coth k_a$  is positive, hence  $P > 0$ . The fact that  $P > 0$  means that sound energy is continuously pumped into flexural waves and gets dissipated by the wall. Such energy pumping is expected considering the absence of acoustic sink inside the fluid.

We now investigate, for a frictionless system, the proportion of energy flux in the fluid compared with that propagating in the flexural waves. The local energy flux in the fluid,  $I_f(y)$ , is found as follows,

$$p = -\frac{\partial \phi}{\partial t} = -i\omega \phi, \quad u = \frac{\partial \phi}{\partial x} = -ik \phi,$$

$$I_f(y) = \frac{1}{2} \text{Re}(p u_*) = \frac{1}{2} \omega k |\phi|^2,$$

in which  $\phi$  was given in Eq. (3). The total sound energy flux in fluid is found by integration across the channel height,

$$I_f = \int_0^1 I_f(y) dy = \frac{k\omega^3}{2k_a^2} \eta_0^2 \Lambda,$$

where

$$\Lambda = \int_0^1 \left[ \frac{e^{k_a(y-1)} + e^{-k_a(y-1)}}{e^{k_a} - e^{-k_a}} \right]^2 dy = \frac{e^{2k_a} - e^{-2k_a} + 4k_a}{(e^{2k_a} + e^{-2k_a} - 2)2k_a}.$$

The time-mean vibrational energy density and the wave energy flux in the structure are, respectively,

$$\frac{dE_s}{dx} = \frac{1}{4} \left[ M \left| \frac{\partial \eta}{\partial t} \right|^2 - T \text{Re} \left( \frac{\partial^2 \eta}{\partial x^2} \eta_* \right) \right],$$

$$I_s = c \left( \frac{dE_s}{dx} \right) = \frac{1}{2} M c \omega^2 \eta_0^2.$$

The structure-fluid energy flux ratio is found to be

$$\mathcal{R} = \frac{I_s}{I_f} = \frac{M}{\Lambda} (1 - c^2),$$

which is a function of mass  $M$  and two other parameters:  $k_0 = \pi f$  and  $c$ . We consider moderate frequency  $f$  and analyze two extremes of wave speed: one very close to the normal speed of sound,  $c \rightarrow 1$ , and the other very slow,  $c \rightarrow 0$ :

- (i) For  $c \rightarrow 1$ ,  $k_a \rightarrow 0$ ,  $\Lambda \rightarrow k_a^{-2} \rightarrow \infty$ ,  $\mathcal{R} \rightarrow M \omega^2 (1 - c^2)^2 \rightarrow 0$ .
- (ii) For  $c \rightarrow 0$ ,  $k_a \rightarrow k \rightarrow \infty$ ,  $\Lambda \rightarrow (2k_a)^{-1} \rightarrow 0$ ,  $\mathcal{R} \rightarrow 2M \omega / c \rightarrow \infty$ .

The conclusion is that when waves are slowed down, most energy is imparted to the structural vibration.

## II. FORMULATION FOR THE SOUND-PANEL INTERACTION

We now investigate the interaction between an incident traveling wave of unit amplitude,

$$p_i = e^{i(\omega t - k_0 x)}, \quad (6)$$

and a finite compliant panel hinged at  $x = \pm L/2$  where  $L = L^*/h^*$  is the dimensionless panel length. Reaction of external fluid is ignored as in the previous section. The panel response is found in three steps. First, the incident wave is allowed to pass through the channel assuming no panel response. The panel is then subject to pressure excitation  $p_i$ . Second, the pressure induced on the panel surface by its own vibration of unit vertical velocity is found; this pressure may be called the local radiation impedance  $Z$ , bearing in mind that  $Z$  may not be a function of the material property alone. Third, the panel vibration is determined by the coupled dynamics:

$$M \frac{\partial^2 \eta}{\partial t^2} + D \frac{\partial \eta}{\partial t} + p_w - T \frac{\partial^2 \eta}{\partial x^2} = 0, \quad (7)$$

in which the total wall pressure  $p_w$  consists of contributions from the incident wave and the radiation of its own:

$$p_w = p_i + Z \partial \eta / \partial t.$$

We now focus on the panel vibration velocity,

$$V = \frac{\partial \eta}{\partial t},$$

instead of displacement  $\eta$ . Equation (7) may be rewritten as

$$\left( \mathcal{L} + Z - \frac{M c_T^2}{i \omega} \frac{\partial^2}{\partial x^2} \right) V + p_i = 0, \quad (8)$$

where

$$\mathcal{L} = i M \omega + D \quad (9)$$

is a structural mechanics operator. Given the noise frequency,  $\mathcal{L}$  can be seen as a local panel property.

## A. Expansion of the axial modes

In order to satisfy the structural boundary conditions of the panel,  $V|_{x=\pm L/2} = 0$ , we expand  $V$  in terms of sine transform using a local dimensionless variable  $\xi = (x/L + \frac{1}{2})$  which spans from 0 to 1 for the panel:

$$V = \sum_{j=1}^{\infty} V_j \sin j \pi \xi, \quad V_j = 2 \int_0^1 V(\xi) \sin j \pi \xi d\xi. \quad (10)$$

In reality a finite number of axial panel modes, say,  $N=20$ , will be sufficient. The sine transform of Eq. (8) becomes a set of equations

$$\mathcal{L}_j V_j + \sum_{n=1}^N V_n Z_{jn} = -I_j, \quad j = 1, 2, 3, \dots, N, \quad (11)$$

in which  $\mathcal{L}_j$  is defined as

$$\mathcal{L}_j = \mathcal{L} + \frac{M c_T^2}{i \omega} \left( \frac{j \pi}{L} \right)^2 \quad (12)$$

and may be seen as the  $j$ th mode property constant of the panel,  $Z_{jn}$  is the radiation impedance on the  $j$ th mode induced by the vibration of the  $n$ th mode, which will be given in the next subsection, and  $I_j$  is the sine transform of the incident wave:

$$\begin{aligned} I_j &= 2 \int_0^1 e^{-i k_0 L (\xi - 1/2)} \sin(j \pi \xi) d\xi \\ &= 2 j \pi e^{i k_0 L / 2} \left[ \frac{e^{i(-k_0 L + j \pi)} - 1}{(k_0 L)^2 - (j \pi)^2} \right]. \end{aligned}$$

When  $k_0 L \rightarrow j \pi$ , the square bracket becomes  $\frac{0}{0}$  and its value is found to be  $(2i k_0 L)^{-1}$ . The solution to Eqs. (11) is

$$[V] = -\mathcal{Z}^{-1}[I], \quad (13)$$

where

$$\begin{aligned} [V] &= \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \vdots \\ V_N \end{bmatrix}, \quad [I] = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \vdots \\ I_N \end{bmatrix}, \\ \mathcal{Z} &= \begin{bmatrix} Z_{11} + \mathcal{L}_1 & Z_{12} & \cdots & Z_{1N} \\ Z_{21} & Z_{22} + \mathcal{L}_2 & \cdots & Z_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{N1} & Z_{N2} & \cdots & Z_{NN} + \mathcal{L}_N \end{bmatrix}. \end{aligned}$$

The impedance matrix  $[Z_{jn}]$  will be found below.

## B. Modal radiation impedance

The pressure perturbation induced by the wall oscillation was given by [Doak (1973), Eq. (123)] and is rewritten below in our notations and simplified for the two-dimensional configuration:

$$p(x, y, t) = \frac{1}{2} \sum_{m=0}^{\infty} c_m \psi_m(y) \int_{-L/2}^{+L/2} \psi_m(y') V(x', t) \times [\mathbf{H}(x-x') e^{-i\omega(x-x')/c_m} + \mathbf{H}(x'-x) e^{+i\omega(x-x')/c_m}] dx', \quad (14)$$

where  $\mathbf{H}$  is the Heaviside function,  $\delta$  is the Kronecker delta,  $y'$  is the coordinate of the panel which is 0 in this case,  $\psi_m$  is the channel modal function defined as

$$\psi_m(y) = \sqrt{2 - \delta_{0m}} \cos(m\pi y),$$

and  $c_m$  is the complex modal wave speed defined as

$$c_m = \frac{1}{\sqrt{1 - (m\pi/\omega)^2}}$$

for the cut-on modes but

$$c_m = \frac{i}{\sqrt{(m\pi/\omega)^2 - 1}}$$

for the cut-off modes.

For panel oscillation mode  $n$  of unit amplitude,

$$\hat{V}_n(\xi', t) = e^{i\omega t} (\sin n\pi\xi'),$$

where  $\xi' = (x'/L + \frac{1}{2})$ , the fluid loading on the panel surface can be calculated from Eq. (14) by substituting  $V$  with  $\hat{V}_n$  and setting  $y=0$ . Its complex amplitude, denoted by  $p_n$ , is sine transformed to give  $Z_{jn}$ ,

$$Z_{jn} = 2 \int_0^1 p_n \sin(j\pi\xi) d\xi = L \sum_{m=0}^{\infty} c_m (2 - \delta_{0m}) \mathcal{I}_2(m, n, j), \quad (15)$$

where  $\mathcal{I}_2$  is the double integration

$$\mathcal{I}_2(m, n, j) = \int_0^1 \left[ \int_0^1 \mathcal{I}(\xi, \xi') d\xi' \right] \sin(j\pi\xi) d\xi.$$

Here  $\mathcal{I}$  is defined as

$$\mathcal{I}(\xi, \xi') = \sin(n\pi\xi') [\mathbf{H}(\xi - \xi') e^{-ik_m L(\xi - \xi')} + \mathbf{H}(\xi' - \xi) e^{+ik_m L(\xi - \xi')}],$$

and  $k_m = \omega/c_m$ . The result of the double integration is

$$\mathcal{I}_2(m, n, j) = \frac{n\pi j \pi (e^{in\pi} - e^{-ik_m L})(e^{in\pi} + e^{ij\pi})}{[(k_m L)^2 - (n\pi)^2][(k_m L)^2 - (j\pi)^2]} - \frac{ik_m L \delta_{jn}}{(k_m L)^2 - (n\pi)^2}. \quad (16)$$

Note that the impedance matrix  $[Z_{jn}]$  is symmetrical. Note also that for frequency below the first cut-on of the channel, only the plane wave mode,  $m=0$ , has real wave number  $k_0 = \omega/c_0$ , and the chances are that  $k_0 L/\pi$  is an integer. In such case the expression for  $\mathcal{I}_2$  becomes  $\frac{0}{0}$ -type, giving

$$\mathcal{I}_2|_{k_0 L/\pi = n \neq j} = \frac{i[1 + (-1)^{n+j}]j\pi}{2[(k_0 L)^2 - (j\pi)^2]}.$$

Replacing  $j$  by  $n$  gives  $\mathcal{I}_2$  for  $k_0 L/\pi = j \neq n$ . When  $j=n$  and  $\delta_{jn}=1$ ,  $\mathcal{I}_2$  given in Eq. (16) becomes the type of  $\frac{\infty}{\infty}$  which is also shown to be finite:

$$\mathcal{I}_2|_{k_0 L/\pi = j = n} = \frac{1}{4} - \frac{3i}{4k_0 L}.$$

As  $m \rightarrow \infty$ ,  $c_m \rightarrow i\omega/m\pi$ ,  $k_m \rightarrow -im\pi$ ,  $\mathcal{I}_2 \rightarrow m^{-3}$ , so that the series for  $Z_{jn}$  in Eq. (15) converges as  $m^{-4}$ . Calculation shows that, for  $L \leq 10$ , less than 20 channel modes are needed for an accuracy better than 0.3%.

### C. Transmission loss

The total sound waves transmitted to the downstream is the sum of incident wave and the wave radiated by the vibrating panel. For an excitation frequency below that of the first cut-on of the channel, the radiated wave may be calculated from Eq. (14),

$$p_{\text{rad}} = \frac{L}{2} \sum_{j=1}^{\infty} V_j e^{i\omega t} \int_0^1 \sin(j\pi\xi') e^{-ik_0 L(\xi - \xi')} d\xi',$$

$$\frac{p_{\text{rad}}}{p_i} = \frac{L}{2} \sum_{j=1}^{\infty} V_j \int_0^1 \sin(j\pi\xi') e^{ik_0 L(\xi' - 1/2)} d\xi' = \frac{L}{2} \sum_{j=1}^{\infty} V_j j \pi e^{-ik_0 L/2} \left[ \frac{e^{i(k_0 L - j\pi)} - 1}{(k_0 L)^2 - (j\pi)^2} \right],$$

in which the square brackets takes the value of  $i/2k_0 L$  when  $k_0 L/\pi = j$ . The transmission loss is found by

$$\text{TL} = -20 \log_{10} \left| \frac{p_i + p_{\text{rad}}}{p_i} \right|.$$

The reflection wave is found similarly,

$$p_{\text{ref}} = \frac{L}{2} \sum_{j=1}^{\infty} V_j e^{i\omega t} \int_0^1 \sin(j\pi\xi') e^{+ik_0 L(\xi - \xi')} d\xi', \frac{p_{\text{ref}}}{p_i} = \frac{L}{2} \sum_{j=1}^{\infty} V_j \int_0^1 \sin(j\pi\xi') e^{-ik_0 L(\xi' - 1/2)} d\xi' = \frac{L}{2} \sum_{j=1}^{\infty} V_j j \pi e^{+ik_0 L/2} \left[ \frac{e^{i(-k_0 L - j\pi)} - 1}{(k_0 L)^2 - (j\pi)^2} \right],$$

which may be obtained by substituting  $k_0$  with  $-k_0$  in  $p_{\text{rad}}/p_i$ .

### III. ANALYSES OF FLEXURAL WAVES

For easy comparison, all examples are based on the same set of data regarded as the default. Deviations of individual parameters will be highlighted in figure captions and text. In order to produce a typical value of mass ratio  $M$ , we consider a channel of height 10 cm, a panel of density 1000 kg/m<sup>3</sup> (close to rubber), and thickness 0.05 mm. Taking air density as 1.225 kg/m<sup>3</sup>, we have

$$M = M^*/\rho_0^* h^* = (1000 \times 0.05 \times 10^{-3}) / (1.225 \times 0.1) = 0.41.$$



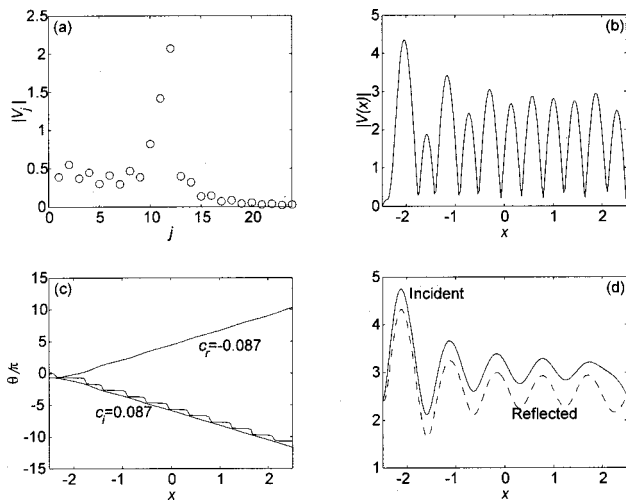


FIG. 4. Basic case of  $M=0.41$ ,  $D=0$ ,  $c_T=0.1$ ,  $L=5$ , and  $f=0.2$ .  $TL=15.0$  dB. (a) Amplitude of modal response  $|V_j|$  which peaks at  $j=12$ . (b) Axial distribution of panel vibration amplitude showing standing wave pattern of short wavelength and strong response near the leading edge. (c) Phase ( $\theta$ ) distribution of panel vibration decomposed into incident and reflected waves both travelling at a speed of 0.087. The line with ripples is for the total vibration velocity ( $V$ ). (d) Amplitudes of the decomposed traveling waves.

The critical frequency below which there is only subsonic waves is  $f_c=0.36$ , or 612 Hz, according to the theory of infinite flexible walls or Fig. 2(b). The following data are chosen to be the default values which will not be stated again in examples:

$$M=0.41, \quad c_T=0.1, \quad D=0, \quad f=0.2. \quad (17)$$

For instance, when considering supersonic waves we only specify  $f=0.4 > f_c$  together with the panel length  $L=10$  while  $M$ ,  $D$ , and  $c_T$  take the above default values, respectively.

### A. The basic case

We first test the “default” subcritical frequency  $f=0.2$  with  $L=5$  and the results are shown in Fig. 4. As shown in Fig. 4(a), the main response of the panel is centered on the 11th axial mode, but Fig. 4(b) shows that the response at the leading edge is apparently stronger than elsewhere. The transmission loss is 15.0 dB. The standing wave may be decomposed into two traveling waves:

$$V(x) = \sum_{j=1}^N V_j \sin(j\pi\xi) = \sum_{j=1}^N \frac{1}{2i} [V_j e^{+ij\pi\xi} - V_j e^{-ij\pi\xi}].$$

The first term in the square brackets represents the upstream-traveling reflection wave, while the second term represents the downstream-traveling incident wave. The phase angles of the decomposed vibration velocities are shown in Fig. 4(c) while the amplitudes are in Fig. 4(d). The phase distributions are rather linear and the phase speed of 0.087 is slightly less than the value given by the theory of infinite panel, 0.0934, cf. Fig. 2(a).

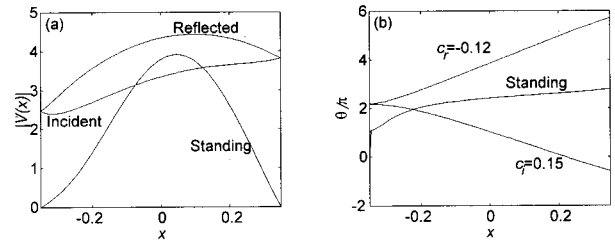


FIG. 5. Response of a short panel ( $L=0.7$ ). (a) Amplitudes of incident, reflected and standing waves. (b) Phase distribution showing slightly different mean phase speed for the incident and reflected waves. The standing wave is seen to be “in-phase” for most of the length except the region near the leading edge.

### B. Short panel

Figure 5 is an example of a short panel ( $L=0.7$ ). The phase distributions of Fig. 5(b) indicate that the vibration is almost in-phase over most of the panel length like a piston. The wave speed is higher than the *in vacuo* wave speed of the tensioned panel, implying that air exerts either positive virtual stiffness or negative virtual mass on the panel.

### C. Supercritical case

We now raise the frequency (0.4) above the predicted critical value of 0.36 for a longer panel ( $L=10$ ). The modal response  $|V_j|$  is plotted in Fig. 6(a) showing two local maxima, one at  $j=44$  and another at  $j=2$ . The wave speed embedded in each panel mode,  $c_j^*$ , is calculated from its wavelength  $\lambda_j^*$ :

$$c_j^* = \frac{\omega^* \lambda_j^*}{2\pi}, \quad \lambda_j^* = \frac{2L^*}{j} \rightarrow c_j^* = \frac{c_j^*}{c_0^*} = \frac{fL}{j}.$$

The dimensionless wave speed  $c_j$  is 0.092 and 2.025 for  $j=44$  and 2, respectively, confirming our earlier prediction of coexisting subsonic and supersonic flexural waves. Figure 6(b) plots the real part of the complex modal vibration amplitude,  $\text{Re}[V(x)]$ , which represents a snapshot of the real vibration velocity when  $e^{i\omega t}=1$ , showing short and subsonic waves modulated by long supersonic waves. In reality it is possible that the subsonic “ripples” may be easily damped by structural viscosity, leaving the long supersonic wave as the dominant feature of the panel response. The transmission loss in this case is only 1.1 dB, although the magnitude of panel response is comparable to the subcritical case of Fig. 4.

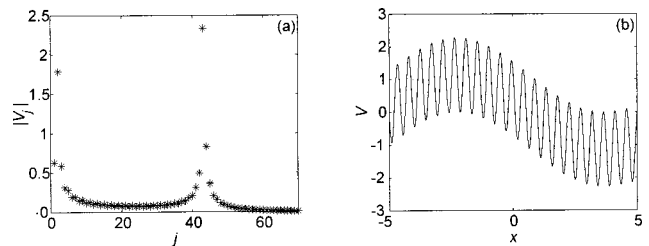


FIG. 6. Supersonic flexural waves on a panel of  $L=10$  at frequency  $f=0.40$ . (a) shows two local maxima of  $|V_j|$ , one at  $j=2$  and the other at  $j=44$ . (b) is  $\text{Re}[V(x)]$  representing the real vibration velocity at one particular moment when  $e^{i\omega t}=1$ .

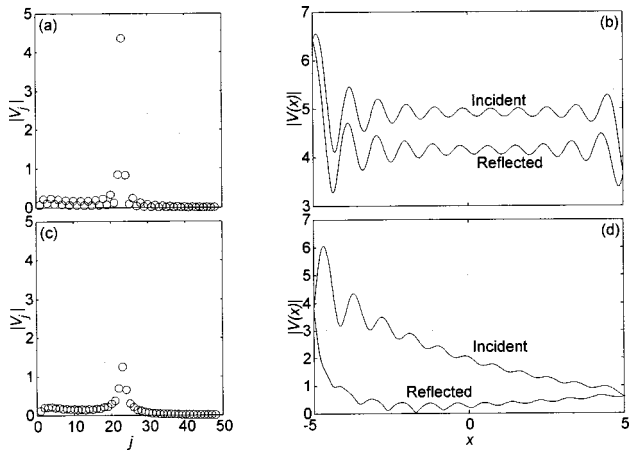


FIG. 7. Comparison of flexural waves on an undamped and damped ( $\bar{D}=0.2$ ) panel of  $L=10$ . (a) and (b) are for the undamped panel producing a transmission loss of  $TL=10.27$  dB, while (c) and (d) are for the damped with  $TL=25.97$  dB. (a) and (c) are plots of modal response while (b) and (d) are the decomposition of flexural waves as incident and reflection waves. (d) shows the effectiveness of wave damping.

#### IV. PERFORMANCE OF PANEL AS A MUFFLER

The performance of the flexible panel as a means of passive noise control is now analyzed. The main conclusion is that the combination of noise reflection and damping of slow flexural waves could be very powerful for a broad frequency band.

##### A. Damping of flexural waves

The level of structural damping  $D$  is estimated by its *in vacuo* dynamics by using Eq. (11) with  $p_i=0$ ,  $Z=0$ :

$$i\omega M + D + \frac{Mc_T^2}{i\omega} \left( \frac{j\pi}{L} \right)^2 = 0.$$

For small  $D$ , the eigen angular frequency is

$$\omega = \pm c_T \left( \frac{j\pi}{L} \right) + i \frac{D}{2M},$$

which may be decomposed into real and imaginary parts:  $\omega = \omega_r + i\omega_i$ . The fractional rate of oscillation reduction per cycle is

$$\bar{D} = 1 - e^{-\omega_i(2\pi/\omega_r)} \approx \omega_i \frac{2\pi}{\omega_r} = \frac{DL}{jc_TM} \rightarrow D = \frac{jc_TM\bar{D}}{L}.$$

Here  $\bar{D}$  is a function of material property, tension force applied, and frequency. However, for the purpose of our estimation it suffices to note that the normal range of material property  $\bar{D}$  is 0.001–0.01 for metals and 0.05 for nonmetallic materials such as natural rubber. For materials specially developed for vibration isolation,  $\bar{D}$  can be much higher. We take  $\bar{D}=0.2$  and reexamine the subcritical case of  $f=0.2$ . The results of the undamped and damped panels of  $L=10$  are given in the upper and lower parts of Fig. 7, respectively. The amplitudes given in Fig. 7(a) and (c) show that the damped panel has much lower response. Figure 7(b) com-

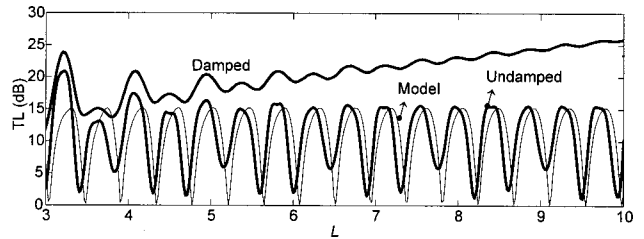


FIG. 8. Variation of transmission loss with panel length for the undamped, damped ( $\bar{D}=0.2$ ) and one-dimensional model showing steady increase of TL for the damped panel and close agreement between the one-dimensional model and the performance of the undamped panel.

pares the incident and reflected waves on the undamped panel. The amplitudes of the two waves do not change appreciably although the response at the leading edge is larger. The transmission loss is  $TL=10.27$  dB. Figure 7(d) is for the damped panel showing obvious decay of both incident and reflected flexural waves. Notice that the reflected wave reaches zero amplitude midway, but the requirement of boundary condition at the leading edge causes it to grow again in its direction of travel to the left. This is, of course, part of scattering. The transmission loss for the damped panel is 25.97 dB, 15.7 dB higher than the undamped counterpart. The portions of acoustic energy reflected, damped, and transmitted are, respectively, 69.53%, 30.22%, and 0.25% for the damped panel, while those for the undamped panel are 90.61%, 0%, and 9.39%. The comparison shows that acoustic energy is effectively transferred from air to the panel for dissipation.

##### B. Effect of panel length

The variation of TL with  $L$  is shown in Fig. 8 for panels without damping and with damping of  $\bar{D}=0.2$ . Regular patterns are established for panels longer than about 5 while irregularities in shorter panels reflect the strong effect of edge scattering. The pattern of TL for the undamped panel is similar to that of an expansion chamber where passbands result from double reflections or standing waves. We validate this point by a simple one-dimensional model in which the panel section is replaced by a rigid duct of the same dimension but is filled with a gas of specific impedance  $\rho^*c^* = c_{in}(\rho_0^*c_0^*)$  in which  $c_{in}$  is the incident panel wave speed found from the calculation as shown in Fig. 4(c). The transmission loss is then

$$TL_{1D} = 10 \log_{10} \left[ \cos^2(k_{in}L) + \frac{1}{4}(c_{in} + c_{in}^{-1})^2 \sin^2(k_{in}L) \right],$$

where  $k_{in} = \pi f/c_{in}$ , which is plotted as the thin line in Fig. 8. The agreement between the one-dimensional model and the two-dimensional analytical solution appears to be satisfactory for relatively long panels. The slight left-shift of the curve for the undamped panel relative to that of the 1-D model signifies a reduction of effective panel length due to the edge scattering, similar to the length correction for sound radiation at the open end of a duct.

For the damped panel, TL increases almost linearly with  $L$ , indicating that the mechanism of further noise reduction becomes purely dissipative. The noise reduction per unit

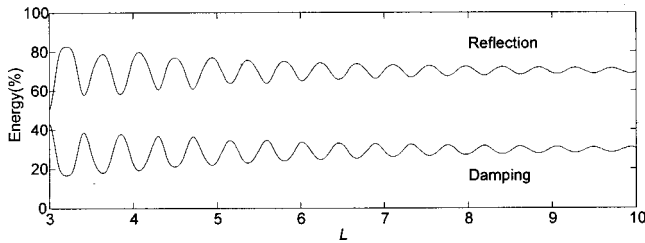


FIG. 9. Proportions of sound energy in the reflected wave and the loss by damping ( $\bar{D}=0.2$ ) with respect to panel length.

length, i.e., one channel height, is 2.0 dB in this example; 4.0 dB may be obtained when two such panels are used on the opposite sides of the channel, a performance which competes well with the best fibrous duct lining which also yields some 4 dB with linings on all four walls (cf. Ingard, 1994). It should be pointed out that the good performance of wave damping per distance is partly attributable to the fact that the wave speed on the flexible panel in our example is only 10% of the isentropic speed of sound.

The reduction of sound energy traveling to the right of the channel is achieved by two mechanisms: wave reflection and damping. Figure 9 shows that the variations in the damped energy and reflected sound energy are complementing each other as panel length varies. The percentage approaches 70% for reflection and 30% for damping for very long panels. The sum of the two approaches unity as transmission loss increases. For a panel as long as  $L=10$ , there is very little structure-borne wave at the downstream part of the panel.

### C. Spectrum of transmission loss

The variation of TL with frequency is shown in Fig. 10 for  $L=10$ . The results for the undamped panel are shown as the vertical lines, and its main features are similar with the variation of TL with length in that it has peaks and dips depending on the match between the local wavelength and the panel length. These variations are eliminated by a moderate level of panel damping ( $\bar{D}=0.1$ ) shown as the thick curves. In Fig. 10(a) we notice that the panel only reduces noise effectively for the subcritical frequency range of  $f < f_c = 0.36$ . For supercritical frequencies supersonic waves become the dominating feature and TL per distance is poor. Figure 10(b) shows the performance of two identical panels placed on the two opposing sides of the channel. The effect

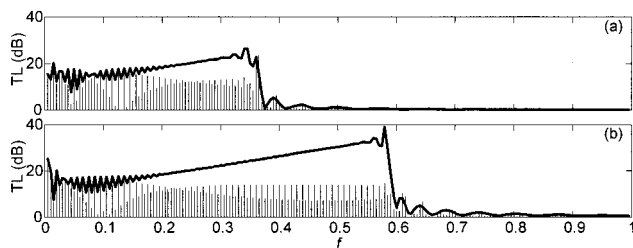


FIG. 10. The TL spectra of (a) single and (b) double panels of  $L=10$ . The thin vertical lines represent the performance without damping while the thick curves are for the damped ( $\bar{D}=0.1$ ) panels.

of having the extra panel is equivalent to placing a rigid wall in the middle of the duct. This is implemented in the calculation by first doubling the dimensionless length, such as  $L = L^*/h^*$ , then scaling the results back to the original definitions of dimensionless variables based on the full duct height  $h^*$ . Compared with the single panel, it has a wider stopband and a higher level of transmission loss.

## V. CONCLUSIONS

Theoretical analyses have been conducted to see whether a combination of wave reflection and damping by a segment of flexible wall in an otherwise rigid channel is effective. The answer is yes and the main mechanisms of the panel muffler are summarized as follows:

- (i) The axial wave speed in a channel with flexible walls can be controlled by the cross-section distensibility and the wave speed derived for the infinite flexible wall is approximated by that of a finite panel. It is also shown that the energy partition between the wall flexural waves and sound in fluid favors the former when the panel wave is much slower than sound in free space.
- (ii) The cross sections of the leading and trailing edges of the panel represent discontinuity in acoustic properties across which wave reflections occur. For a frictionless panel the spectrum of transmission loss is similar to that of an expansion chamber. Unsatisfactory passbands exist and can only be eliminated by structural damping.
- (iii) For a panel with considerable structural damping the coefficient of wave reflection at the leading junction is reduced. Much of the transmitted energy flux propagates as wall flexural waves which can be dissipated at a spatial rate much higher than is normally possible in air. The final transmitted wave is much weaker than that of a lossless panel. For panels of moderate length the spectrum of transmission loss is smooth without significant passbands.
- (iv) A crude estimate shows that the fractional damping rate per distance achieved over such a panel is comparable to or even higher than that of fibrous duct lining.

However, the initial study excludes many issues of practical importance.

- (i) The possibility of instability and sound generation by flow-induced vibration on the compliant panel. Initial reasoning suggests that such a problem may not occur as far as the mean flow is everywhere slower than the flexural wave speed.
- (ii) The effect of the mean flow on the acoustic performance of panel muffler is not considered yet.
- (iii) The effect of the external fluid has not been considered. For a long panel with subsonic wave speed, the effect is likely to be one of added mass. For supersonic waves, there will be considerable Mach wave radiation causing decay of flexural waves. For panels

of short to moderate length, which is the focus of this article, a local-reacting model for the external fluid may not be sufficient and a comprehensive analysis would be rather involved. The same applies to situations where the panel is supported by an air cavity. These are left to future studies.

- (iv) The performance of the panel muffler for incident waves of frequencies higher than the first cut-on frequency of the channel based on the normal speed of sound is also left to future studies.
- (v) Direct experimental validation of the findings is likely to be difficult since in reality a two-dimensional channel can only be approximated by a three-dimensional duct. Structural restrictions in the third dimension may have considerable effect on the modes of waves propagating in the duct.

## ACKNOWLEDGMENT

This work is supported by a grant from The Hong Kong Polytechnic University (Grant No. G-S739).

- Cummings, A. (1994). "Attenuation of sound in unlined ducts with flexible walls," *J. Sound Vib.* **174**, 433–450.
- Doak, P. E. (1973). "Excitation, transmission and radiation of sound from source distributions in hard-walled ducts of finite length (I): the effects of duct cross-section geometry and source distribution space-time pattern," *J. Sound Vib.* **31**, 1–72.
- Ffowcs Williams, J. E., and Hill, D. C. (1987). "On the scattering of evanescent waves into sound," *J. Fluid Mech.* **184**, 101–121.
- Huang, L. (1998). "Reversal of the Bernoulli effect and channel flutter," *J. Fluids Struct.* **12**, 131–151.
- Ingard, K. U. (1994). *Notes on the Sound Absorption Technology* (Noise Control Foundation, Poughkeepsie, NY).
- Lighthill, M. J. (1978). *Waves in Fluids* (Cambridge U.P., Cambridge).
- Munjal, M. L. (1987). *Acoustics of Ducts and Mufflers* (Wiley, New York).



# Experimental investigation of the acoustical characteristics of university classrooms

Murray Hodgson

*Occupational Hygiene Program and Department of Mechanical Engineering,  
University of British Columbia, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada*

(Received 26 June 1998; revised 18 May 1999; accepted 10 June 1999)

Acoustical measurements were performed in 30 randomly chosen, unoccupied classrooms at the University of British Columbia (UBC). Tests had previously been done in 46 unoccupied UBC classrooms, as well as in 10 of these when occupied by students. The results for the 10 classrooms were used to correct the "unoccupied" results to the half-occupied and fully occupied conditions. The objective of the work was to characterize the 30 classrooms, which were used in subsequent studies, to determine the acoustical quality of the UBC classroom stock and how this depends on the classroom design and the presence of students, and to elucidate characteristics of classroom acoustics relevant to optimal design. The results showed that the UBC classroom stock is of far from optimum acoustical quality when unoccupied, but is much better in the occupied condition. Generally, many classrooms have excessive reverberation and result in low speech levels, especially at the back of the rooms; in addition, they have excessively noisy ventilation systems. © 1999 Acoustical Society of America. [S0001-4966(99)00510-X]

PACS numbers: 43.55.Gx, 43.55.Hy, 43.71.Gv [JDQ]

## INTRODUCTION

An extensive, multi-phase study of classroom acoustics is being undertaken. Its ultimate aim is to determine how to optimize the acoustical design of university classrooms with respect to verbal communication. In the first phase of the work,<sup>1</sup> preliminary measurements were made in 46 classrooms at the University of British Columbia (UBC) when unoccupied and in 10 of them when the classrooms were occupied with students. Subsequent phases of the work involved 30 classrooms. This paper reports the results of acoustical tests done in these 30 classrooms when unoccupied. It also draws on results from Ref. 1 to consider occupied classrooms. The objectives of the current work were as follows:

- (1) to characterize the 30 classrooms used in subsequent work, with respect to their main physical and acoustical characteristics;
- (2) to elucidate the main acoustical characteristics of university classrooms which are relevant to classroom design; and
- (3) to determine the acoustical quality of the UBC classroom stock and how this depends on classroom design and the presence of students.

## I. SPEECH INTELLIGIBILITY

In university classrooms, the major acoustical concern is verbal communication. Nonoptimal acoustical conditions, resulting in reduced verbal communication, can cause two main problems. First, they can lead to reduced learning efficiency.<sup>2</sup> Second, they can lead to fatigue, stress, and health problems (headaches, sore throats) among lectures, who are forced to compensate for poor acoustical conditions by raising their voices, for example.<sup>3</sup>

The quality of verbal communication can be quantified by the "speech intelligibility." This quantity is the percentage of speech material that is correctly identified by the average listener. In the present study, speech intelligibility was assumed only to be due to the following two factors.

- (1) Signal-to-noise ratio—this is equal to the level of speech minus the level of background noise, both at the listener position. The speech level depends mainly on the speaker's voice level, the distance between the speaker and the listener, and on the acoustical characteristics of the classroom. The background noise results from noise from the ventilation system, classroom equipment, in-class student activity, and sources outside the classroom. The levels of these depend, among other factors, on the acoustical characteristics of the classroom. In this paper, long-term, average speech level, background-noise level, and signal-to-noise ratio are considered, as expressed in terms of the total, A-weighted values.
- (2) Reverberation—the reverberation in a room generally increases with room size, decreases with the amount of sound absorption in the room, and is affected by room shape and contents. It is known that reverberation affects speech intelligibility by affecting the early- and late-arriving energies and, more importantly, their ratio, the early-to-late energy fraction.<sup>4</sup> The early and late energies are usually defined as the total speech energy radiated by the speech source which arrives at a receiver position at times less than and greater than, respectively, about 50 ms after the arrival of the direct sound. Speech intelligibility is directly related to the early-to-late energy fraction. Excessive reverberation results in a reduced ratio of early-to-late energy and, therefore, in lower speech intelligibility. Reverberation can be quantified by a number of measures, besides the early-to-late energy fraction, including reverberation time (RT) and early decay time

(EDT). While, in principle, the early-to-late energy fraction (and speech intelligibility) can vary without affecting RT or EDT, in the case of real classrooms these measures tend to be highly correlated.<sup>5,6</sup> The quantities are usually measured in octave bands.

While the aim of the present investigation was to study the separate effects of signal-to-noise ratio and reverberation, and how they are influenced by room design, the ultimate acoustical conditions experienced by classroom users depend, of course, on the combined effect of these two factors on verbal communication. A number of measures have been proposed to quantify this combined effect. One of these is the speech transmission index (STI), details of which are presented in Ref. 7. In any case, if STI is a good measure of the combined effect of signal-to-noise ratio and of reverberation on verbal communication, then, clearly, the speech transmission index for infinite signal-to-noise ratio ( $STI_{\infty}$ ) is a good, alternative measure of the effect of reverberation alone. Thus, that measure was also used in the present work. Speech transmission index, as for RT and EDT, is measured in octave bands. However, in the case of speech transmission index, it is conventional to sum the octave-band values, after multiplying them by weighting factors which indicate their relative contributions to speech intelligibility, to obtain a single “global” value (GSTI,  $GSTI_{\infty}$ ). The most important octave-band frequencies for speech intelligibility are 2000 Hz, followed by 1000 and 4000 Hz.<sup>7,8</sup>

It is worth noting here that an implicit assumption of the STI approach is that its values define completely the conditions for speech, independent of the individual values of the two component factors, reverberation and signal-to-noise ratio (of which there are an infinite number of combinations), which result in a given STI.

The higher the speech level is, and the lower the background-noise level is, the higher the signal-to-noise ratio is and, thus, the speech intelligibility. In itself, reverberation is detrimental to verbal communication, because it results in greater late energy, as discussed above. That is, the optimum is zero reverberation. However, in rooms, reverberation also tends to increase speech levels, since it is associated with the reverberant sound field, which adds to the direct field. For this reason, optimal conditions for speech occur for some nonzero reverberation. Research has suggested that, to obtain high speech intelligibility for normal-hearing adults working in their first language, the 1000-Hz octave-band reverberation time must not exceed 0.5 to 0.7 s; with this reverberation time, the signal-to-noise ratio must exceed 15 dB.<sup>9</sup> Given typical speech levels,<sup>10</sup> this implies that the background-noise level must not exceed about 35 dB.<sup>9</sup> As mentioned, the requirements are even more stringent in the case of more acoustically challenged persons (children, hard-of-hearing, second-language); an optimum mid-frequency reverberation time of 0.4 to 0.5 s, a minimum signal-to-noise ratio of 20 dB, and a maximum background-noise level of 25 to 30 dB have been suggested.<sup>9</sup> An important issue, seldom discussed, is whether these criteria should apply to the occupied or the unoccupied classroom. Listeners experience the occupied classroom, but it is likely more effective to specify design

TABLE I. Range, average, and standard deviation of the main physical characteristics of the 30 classrooms.

Characteristic	Minimum	Maximum	Average	Standard deviation
Length (m)	4.4	24.9	10.6	4.9
Width (m)	4.4	21.0	10.4	4.1
Average height (m)	2.6	7.7	3.4	0.9
Volume (m <sup>3</sup> )	109.9	3890.1	518.2	724.0
Surface area (m <sup>2</sup> )	149.8	1631.5	404.5	331.3
Volume/surface area (m)	0.7	2.4	1.0	0.4
Seats	20	515	108	112

criteria for unoccupied classrooms to architects. Furthermore, if the criteria apply to the occupied rooms, should they apply when the room is fully occupied, half-occupied, or when?

In the absence of noise transmission from outside the room, the acoustical conditions in a classroom depend on three main factors, besides the sound sources: room geometry (size and shape); the total (surface and air) classroom absorption; and the number of people in the classroom. All three factors affect speech and background-noise levels, as well as reverberation.

## II. CLASSROOMS TESTED

Measurements were performed in two categories of classroom.

- (1) Randomly selected, unoccupied classrooms—tests were done in 30 unoccupied classrooms, chosen randomly from the UBC classroom list. This represents about 7% of the UBC classrooms. Of course, the results of tests in unoccupied classrooms are not typical of the acoustical conditions in a classroom when in use for lectures, since they do not account for the presence of students, who absorb and scatter sound. However, tests in unoccupied classrooms are much easier to do than in occupied classrooms. Table I presents data summarizing the main physical characteristics of the 30 classrooms. The classrooms varied from small lecture rooms with volumes around 100 m<sup>3</sup> and as few as 20 seats, to large auditoria with volumes over 3000 m<sup>3</sup> and over 500 seats. The volume-to-surface-area ratios varied from 0.7 to 2.4 m. The largest proportion of rooms had 40 to 60 seats, volumes from 250 to 500 m<sup>3</sup>, and volume-to-surface-area ratios of about 1.0 m. The classroom shapes also varied considerably. Regarding the plan shape, many were rectangular, with length less than, similar to, or greater than width. Some were somewhat fan shaped; two had other plan shapes. As for elevation, all had rectangular cross-section shapes of various aspect ratios. Many had rectangular longitudinal-section shapes of varying aspect ratio, and a flat floor. Some classrooms, particularly the larger ones, had sloped or tiered floors and seating. One or two had nonflat ceilings or reflectors, apparently designed to direct speech sound to the back of the classroom.
- (2) Unoccupied/occupied classrooms—in order to determine the effect of the presence of students on verbal commu-

nication and, thus, to correct the results from the unoccupied classrooms for the presence of students, tests were done in ten classrooms when both unoccupied and occupied by a number of students. These are described in Ref. 1.

During all of the tests, the ventilation systems were in normal operation. However, overhead or slide projectors, common sources of background noise in a classroom, were not in operation. Noise from outside the classroom was not a significant factor. Speech-reinforcement systems, installed in some classrooms, were not in operation. In the case of the occupied-classroom tests,<sup>1</sup> the students were asked to remain quiet; thus, the effect of background noise due to in-class student activity was not measured.

### III. EXPERIMENTATION AND ANALYSIS

In each classroom, measurements were made of the impulse responses between a source and each of three to eight microphone positions, distributed throughout the room as described below, using the maximum length sequence system analyzer (MLSSA).<sup>7</sup> Tests were performed at typical student positions, to determine the acoustical characteristics at such positions.

Improving on the tests reported in Ref. 1, which used an omni-directional source, a speech source, radiating with directivity approximating that of a typical human speaker, and meeting the criteria described in Ref. 8, was located at the typical lecturing position. This was usually at half-width and 1–3 m from the front-end wall, at 1.4 m height; the source “faced” towards the student-seating area. The source output spectrum was shaped, using a digital equalizer, to maximize the signal-to-noise ratio which, as indicated by the MLSSA results, exceeded 15 dB during all but a few measurements. The speech source was precalibrated in an anechoic chamber. Output octave-band sound-power levels were measured. In addition, values of octave-band  $STI_{\infty}$  generated in this “perfect” acoustical environment were measured and used to normalize subsequent  $STI_{\infty}$  values measured in “imperfect” environments, as described in Ref. 8.

Receiver positions were in seats, at typical ear height, 1.1 m above the floor and 0.6 m above the seat. Since most rooms were laterally symmetrical, receiver positions were usually located in only one-half of the classroom and were chosen to represent typical seating positions towards the front, middle, and back of the classroom, at the center and the side, as illustrated in Fig. 1. The range of source/receiver distances involved in the tests was 0.6–20.3 m. Typically, the student position closest to the instructor is at a distance of about 1, 2, and 4 m in small, medium, and large classrooms, respectively. The farthest receiver is at about 4, 8, and 16 m, respectively. That is, the farthest receiver is at a position corresponding to approximately two doublings of the source/receiver distance (dd), irrespective of classroom size.

At each position, and in octave bands from 250 to 8000 Hz, the following quantities were determined from the measured impulse responses.

(1) Sound propagation (SP)—this is the received steady-

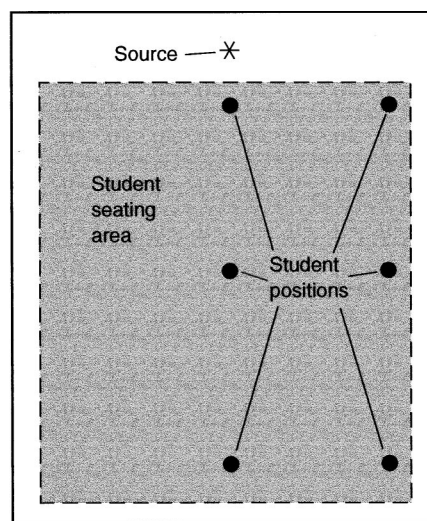


FIG. 1. Plan view of a typical classroom, showing seating area and instructor–source and student–receiver positions.

state sound-pressure level  $L_p$  minus the source sound-power level  $L_w$ , generally resulting in a negative value in decibels. It is a useful measure describing how the room and its contents, independent of the source output-level characteristics, affect the variation of sound-pressure level and, thus, speech level, with distance from a source. In addition, the measured octave-band sound-pressure levels, the known sound-power levels of the speech source, and sound-power levels for a typical person speaking at midway between normal and raised vocal effort<sup>8</sup> were used to calculate total A-weighted speech levels (SLA). Thus, results apply to the average speaker or, in other words, 50% of all speakers. For results to apply to the majority of speakers, lower levels would have to be used; typical standard deviations of voice levels are 4 to 5 dB.<sup>10</sup>

- (2) Early decay time (EDT)—this was the measure of reverberation based on the initial (0 to –10 dB) part of the sound decay.
- (3) Reverberation time (RT)—this was the measure of reverberation based on the –5- to –25-dB part of the sound decay.
- (4) Speech transmission index for infinite signal-to-noise ratio ( $STI_{\infty}$ ).

It is clearly convenient to work with a summary parameter value that combines the numerous octave-band values into a single number. Thus, global  $STI_{\infty}$  values were calculated from the octave-band  $STI_{\infty}$  values, using the weighting factors recommended in Ref. 8. Similarly, global EDT (GEDT) and RT (GRT) values were calculated. These are proposed, in the absence of a better method, and without precise scientific justification, as convenient summary measures that can be used to illustrate trends in the effect of classroom design on the resulting acoustical conditions. They weight the frequencies in approximately the correct proportion and vary with classroom design as do the individual octave-band results. If the weights indicate the rela-



TABLE II. Details of six classrooms for which detailed results are presented.

Room	Seating type	No. of seats	Length (m)	Width (m)	Height (m)	Global average absorption coefficient	No. students (occupied)
U1	Flat	24	5.8	8.2	2.7	0.20	...
U2	Flat	23	9.1	4.4	3.7	0.10	...
U3	Sloped	263	17.0	13.9	4.2	0.12	...
U4	Sloped	515	24.9	20.2	7.7	0.25	...
O1	Flat	120	10.6	11.6	3.1	0.27	14
O2	Sloped	451	25.2	21.5	6.1	0.16	350

tive importance of octave-band STIs to the total STI, and if speech intelligibility relates directly to reverberation, then the weights should also indicate, at least qualitatively, the relative importance of the octave-band reverberation contributions to overall speech intelligibility. Bradley<sup>11</sup> has recently worked with similar concepts after testing their validity more rigorously. The following weighting factors, in octave bands from 250 to 8000 Hz, were used: 0.0617/0.1671/0.2373/0.2648/0.2142/0.0549.<sup>8</sup> Note that these are different from the weighting factors suggested in Ref. 7 and used in the first phase of this work.<sup>1</sup> Ranges of  $GSTI_{\infty}$  values can also be associated with subjective acoustical-quality descriptors as follows:<sup>7</sup> 0.0 to 0.3 (bad); 0.3 to 0.45 (poor); 0.45 to 0.6 (fair); 0.6 to 0.75 (good); and 0.75 to 1.0 (excellent). These descriptors apply to global STI values, independent of the values of the two component factors, reverberation and signal-to-noise ratio, that determine a given value. Thus, for example, they apply in the present case of infinite signal-to-noise ratio. The  $GSTI_{\infty}$ , GRT, and GEDT results were further averaged over all of the measurement positions in each classroom to obtain room-average values.

In addition, measurements were made, at each position, of ventilation-noise levels, in the 31.5- to 8000-Hz octave bands and in total A-weighted decibels, using a sound-level meter.

It is important, from a design point of view, to determine to what extent typical classroom surfaces absorb sound. Therefore, diffuse-field reverberation theory and the octave-band reverberation times measured in the 30 unoccupied classrooms from this study, as well as the 46 unoccupied rooms from the previous study,<sup>1</sup> were both used to determine the average octave-band absorption coefficients of the classroom surfaces. From these, global absorption coefficients ( $G\alpha$ ) were calculated as a convenient, approximate summary measure, based on the same rationale as discussed above.

It was also of interest to determine to what extent people in a classroom (students) absorb sound. Therefore, diffuse-field reverberation theory and the reverberation times previously measured in the ten classrooms when both unoccupied and occupied<sup>1</sup> were used to determine the average absorption per student, in  $m^2$ .

#### IV. RESULTS

Test results for the unoccupied classrooms will be illustrated for four classrooms U1–U4, whose main characteristics are shown in Table II. Classroom U1 was small, paral-

lelepipedic with width greater than length, and had moderate absorption due to a suspended acoustical ceiling. It had the highest  $STI_{\infty}$ s and the lowest EDTs and RTs of the 30 classrooms measured. Classroom U2 was small, parallelepipedic with length greater than width, and had low absorption. It had the highest SLAs measured. Classroom U3 was a large classroom with rectangular floor-plan and length greater than width. It had inclined, wooden seating and low surface absorption. It had the lowest  $STI_{\infty}$ s and the highest EDTs and RTs of the classrooms measured. Classroom U4 was a very large, irregularly shaped auditorium with fairly high absorption due to a carpet and upholstered seating. It had the lowest SLAs measured. That is, these four classrooms represent all four combinations of high and low size and sound absorption and, thus, the full range of  $STI_{\infty}$ s, EDTs, RTs, and SLAs measured.

Results taken from Ref. 1, illustrating the effect of students on the classroom acoustical environment, will be presented for two classrooms, O1 and O2, whose main characteristics are also shown in Table II. Classroom O1 was approximately parallelepipedic, with width slightly greater than length and a slightly sloped floor, was moderately sized, and had quite-high absorption due to a carpet, wall absorption, and upholstered seating. Classroom O2 was a very large, irregularly shaped auditorium, with length greater than width, and with moderate absorption due to a carpet.

#### A. Unoccupied classrooms

Figure 2 shows the variation of octave-band and global  $STI_{\infty}$  with source/receiver distance in classrooms U1–U4. With the exception of more significant fluctuations at low frequency, presumably due to modal effects, generally  $STI_{\infty}$ s vary little with distance at all frequencies. In order to investigate this in more detail, logarithmic regression was used to fit the  $GSTI_{\infty}$  versus distance data. In individual classrooms,  $GSTI_{\infty}$  varied with source/receiver distance at rates varying between  $-0.07$  and  $+0.02$  per distance doubling. The average (standard deviation) of the variations was  $-0.02$  (0.02), confirming that variations are typically small. According to the discussion above of typical source/receiver distances,  $GSTI_{\infty}$  would be expected to decrease by only about 0.04 from the front to the back of a classroom. The rate of decrease of  $GSTI_{\infty}$  with distance generally increased with classroom size and total classroom absorption.

Generally,  $STI_{\infty}$ s increased with frequency, especially in the classrooms containing significant sound absorption. This is as expected, since  $STI_{\infty}$  is inversely related to reverbera-



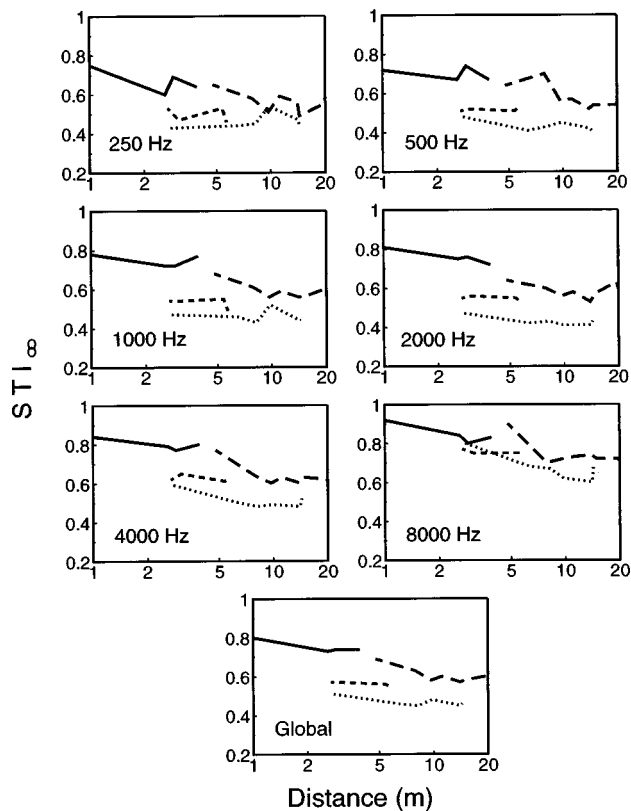


FIG. 2. Variation of measured octave-band and global  $STI_{\infty}$  with source/receiver distance in four unoccupied classrooms: solid line, U1; short-dashed line, U2; dotted line, U3; and long-dashed line, U4.

tion, which is inversely related to the total room absorption; thus,  $STI_{\infty}$  is directly related to the total room absorption. Room absorption, due to air absorption and porous surfaces, results in a total room absorption which tends to increase with frequency. As for the variation of  $STI_{\infty}$  from room to room, Fig. 2 illustrates the key effects. Classroom U1, with small size and moderate absorption had the highest values. Classroom U3, with large size and low absorption, had the lowest values. Classrooms U2, with low size and absorption, and U4, with large size and moderate absorption, had intermediate values of  $STI_{\infty}$ . This will be discussed further in Sec. V D.

Figure 3 shows the frequency distribution of the room-average  $GSTI_{\infty}$ s measured in the 30 unoccupied classrooms. Only two (7%) of the classrooms had room-average  $GSTI_{\infty}$ s above 0.75 (excellent); none was below 0.45 (poor). Twice as many (60%, compared to 33%) of the classrooms had good conditions ( $0.45 < GSTI_{\infty} < 0.6$ ) as had fair conditions ( $0.6 < GSTI_{\infty} < 0.75$ ). We will return in Sec. V A to the question of the quality when occupied.

Regarding EDT and RT, values were generally very similar. Linear regression on the correlation between the global room-average values found the following best-fit relationship:  $GRT = 1.012 \cdot GEDT$  ( $R^2 = 0.96$ ), confirming this fact. Figure 4 shows the room-average, octave-band, and global EDTs in classrooms U1–U4. The EDTs generally decreased with increasing frequency, particularly in treated rooms. As discussed above for  $GSTI_{\infty}$ , this is expected, since EDT is directly related to total room absorption, which

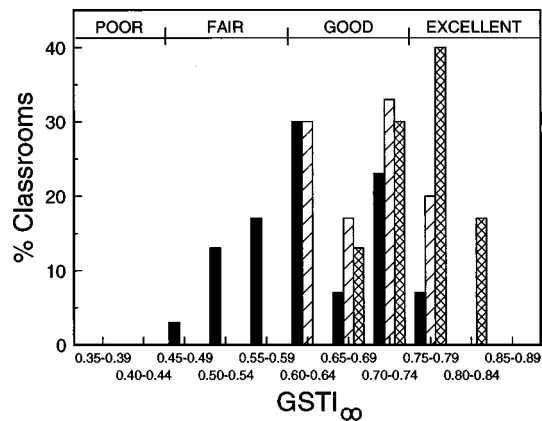


FIG. 3. Frequency distribution of room-average global  $STI_{\infty}$  in 30 classrooms: solid block, unoccupied (measured); slashed block, half-occupied (predicted); and hatched block, fully occupied (predicted).

generally increases with frequency. Classroom U1 (small, absorbent) had the lowest EDT and classroom U3 (large, nonabsorbent) generally had the highest. Classrooms U2 (small, nonabsorbent) and U4 (large, absorbent) had intermediate EDTs. Again, these results will be discussed in more detail in Sec. V D.

Figure 5 shows the frequency distribution of the measured GEDT values in the 30 unoccupied classrooms. Measured GEDTs exceeded 0.5 s in all but two (93%) classrooms, 0.7 s in 16 (53%), and exceeded 1.5 s in one classroom (GEDTs as high as 2.2 s were reported in Ref. 1). The largest number (13, or 43%) of the classrooms had GEDTs in the range 0.40–0.59 s. Figure 6 shows the correlation between room-average  $GSTI_{\infty}$  and GEDT. There is clearly a strong correlation between these two measures, as expected. The relationship appears to be substantially linear over the range of values measured. Linear regression found the following best-fit relationship:  $GSTI_{\infty} = 0.79 - 0.20 \cdot GEDT$  ( $R^2 = 0.72$ )—see Fig. 6. A slightly better fit, describing the expected rise in  $GSTI_{\infty}$  for lower GEDT, and the expected asymptotic behavior for higher GEDT, was obtained with a power-law regression:  $GSTI_{\infty} = 0.58 \cdot GEDT^{0.28}$  ( $R^2 = 0.75$ )—see Fig. 6.

Figure 7 shows the distribution of total A-weighted background-noise levels (BNA) measured in the 30 unoccu-

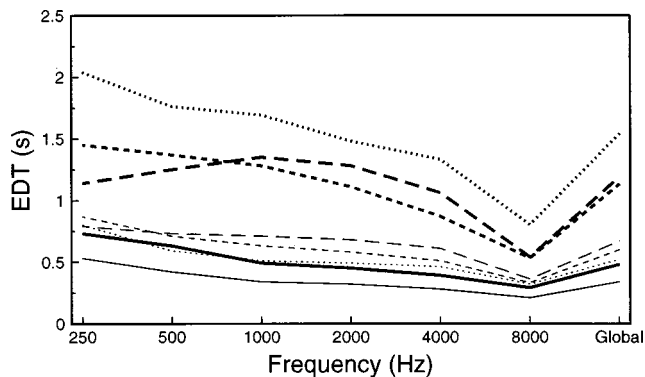


FIG. 4. Room-average, octave-band, and global EDT in four classrooms when unoccupied, as measured (thick lines), and fully occupied, as predicted (thin lines): solid line, U1; short-dashed line, U2; dotted line, U3; and long-dashed line, U4.

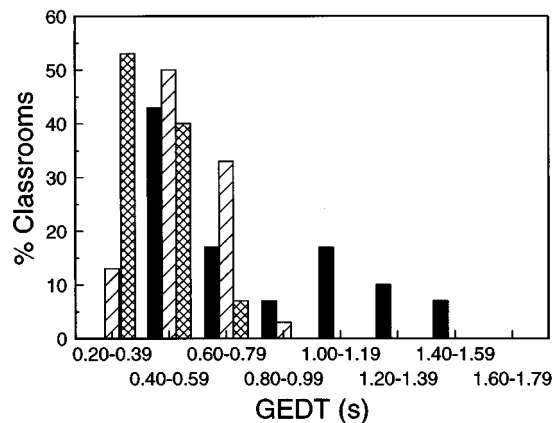


FIG. 5. Frequency distribution of room-average global EDT in 30 classrooms: solid block, unoccupied (measured); slashed block, half-occupied (predicted); and hatched block, fully-occupied (predicted).

occupied classrooms. The BNA exceeded 35 dB in all but one (i.e., in 97%) classroom, and 45 dB in 12 (40%). Of course, the ventilation noise varied with position in a classroom in a way which depended mainly on the position of the ventilation inlets and outlets.

Figure 8 shows the variation of octave-band SP with source/receiver distance in classrooms U1–U4; Fig. 9 shows the corresponding SLA results. Levels were highest, with SLAs in the range 63–65 dB, in classroom U2 (small, low absorption), and lowest, with SLAs in the range 47–53 dB, in classroom U4 (large, high absorption). Intermediate SLAs were measured in the other two classrooms (U1: small, moderate absorption, and U3: large, low absorption). See Sec. VD for further discussion.

In the small classrooms, particularly when nonabsorbent, the speech level varied little with position. In large rooms, particularly when absorbent, the speech level decreased significantly with distance from the source, leading to significantly lower speech levels at the back of the room. Generally levels tended to decrease with distance at all frequencies; with distance plotted on a logarithmic scale, the decrease was often fairly linear (see Figs. 8 and 9). In order to investigate this in more detail, logarithmic regression was used to fit the SLA versus distance data; resulting fits were

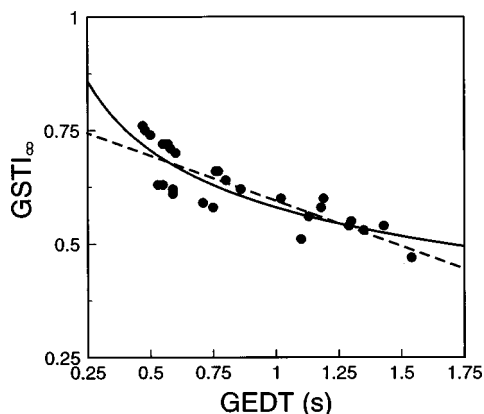


FIG. 6. Correlation between room-average, measured, global  $STI_{0}$ , and global EDT in 30 unoccupied classrooms, and the best-fit curves obtained by (solid line) linear and (dashed line) power-law regressions.

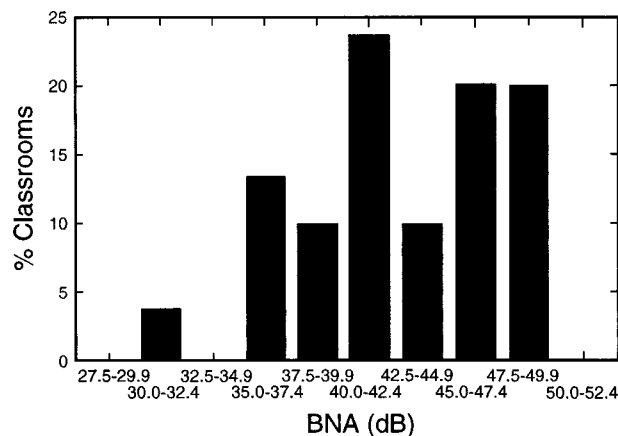


FIG. 7. Frequency distribution of room-average, measured A-weighted total background-noise levels in 30 unoccupied classrooms.

usually excellent (average  $R^2=0.85$ ). There was a tendency for levels at side positions to be 0.5–1 dB higher than levels at central positions, for the same source/receiver distance, presumably due to sound reflection from the side-wall. In individual classrooms, SLA decreased with source/receiver distance at rates varying between 0.3 and 4.0 dB per distance doubling, the rate of decrease being highest in larger rooms with more absorption. The average (standard deviation) of the variations was 1.9 (0.9) dB. These rates of decrease correspond to a range of decreases of speech level from the front to the back of a classroom of between about 0.5 and 8 dB, with an average decrease of 4 dB. Of course, the reduced SLA at the back of (especially large and absorptive) classrooms can lead to a reduced signal-to-noise ratio and significantly poorer conditions for verbal communication, as is well known. Regarding classrooms U1–U4, and the results shown in Fig. 9, SLA decreased with distance at the following rates:

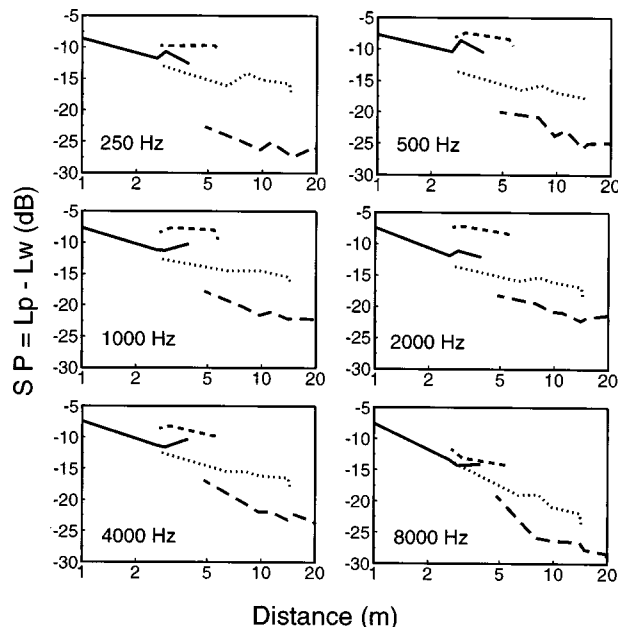


FIG. 8. Variation of measured octave-band SP, with source/receiver distance in four unoccupied classrooms: solid line U1; short-dashed line, U2; dotted line, U3; and long-dashed line, U4.

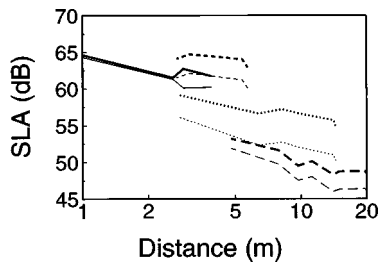


FIG. 9. Variation of SLA with source/receiver distance in classrooms U1–U4 when unoccupied, as measured (thick lines), and fully occupied, as predicted (thin lines): (solid line) U1; (short-dashed line) U2; (dotted line) U3; and (long-dashed line) U4.

1.1 dB/dd (U1); 1.0 dB/dd (U2); 1.6 dB/dd (U3); 2.4 dB/dd (U4).

Note also that the shapes of these curves, in particular, the significant slopes in larger, more absorbent classrooms, indicate that prediction by diffuse-field steady-state theory may be inaccurate. Levels generally decrease with distance; approximately constant reverberant levels predicted by this theory only occurred in small, low-absorption rooms.<sup>12</sup>

In the absence of special acoustical treatment, the main features found to significantly increase the amount of sound absorption above and beyond the ambient absorption in “basic” rooms, which did not have these features, were carpets, absorptive wall and/or ceiling treatments, and upholstered seating. In order to determine the absorptive properties of typical classroom surfaces, the classrooms were subdivided into groups according to whether or not they had these features, and group-average surface absorption coefficients were calculated. Figure 10 shows the average octave-band and global surface-absorption coefficients which can thus be attributed to each feature in these classrooms. For example, classrooms without carpets, absorptive wall and/or ceiling treatments, and upholstered seating had, on average, a global absorption coefficient of 0.09. The presence of a carpet, absorptive wall and/or ceiling treatments, or upholstered seating increased the global coefficient by 0.05, 0.08, or 0.05, respectively. Note that these values are likely higher than

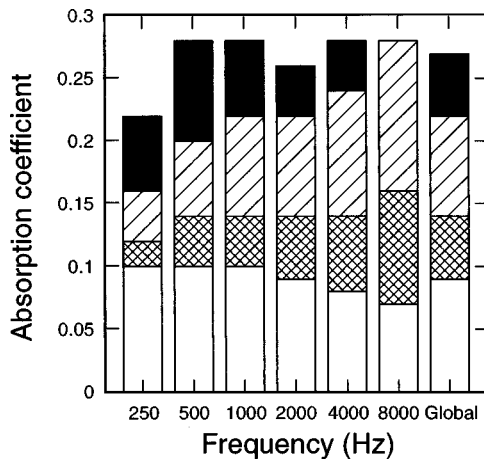


FIG. 10. Average octave-band and global surface absorption coefficients in unoccupied classrooms without and with acoustical features: white block, without (“basic”); cross-hatched block, wall/ceiling absorption; slashed block, carpet; and solid block, upholstered seating. Adapted from Ref. 1.

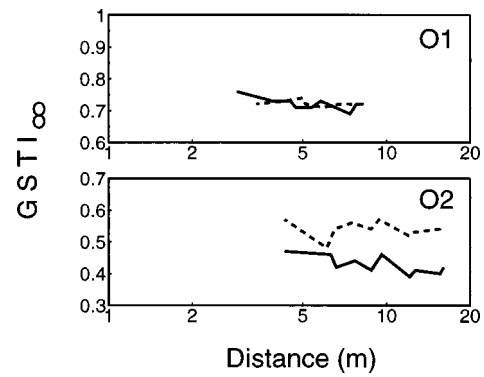


FIG. 11. Variation of measured global  $STI_{\infty}$  with source/receiver distance in classrooms O1 and O2: (solid line) unoccupied and (dashed line) occupied. Adapted from Ref. 1.

those associated with the true absorption of the surfaces because of scattering resulting from the presence of the classroom contents (desks and chairs), which act as “fittings” which reduce reverberation times, increasing apparent absorption coefficients.<sup>13</sup>

## B. Occupied classrooms

Figure 11 shows the variation with distance of the change, due to students, of the  $GSTI_{\infty}$  in classrooms O1 and O2. The presence of students generally had little effect on  $GSTI_{\infty}$  in smaller classrooms, especially when absorbent, such as classroom O1, and increased the  $GSTI_{\infty}$  in larger, less-absorbent classrooms, such as classroom O2. Similarly, as illustrated in Fig. 12, in small, absorbent classrooms students had little effect on reverberation, whereas in large classrooms, especially when nonabsorbent, they reduced the reverberation significantly (in classroom O2, GEDT decreased by 44%). As for speech levels, Fig. 13 shows the change in the 1000-Hz octave-band SP in classrooms O1 and O2, respectively, due to students. In classroom O1, levels decreased by from 0 to 1 dB. In classroom O2, they decreased by 2 to 3 dB.

Figure 14 shows the average and standard deviations of the octave-band and global absorption-per-student results. The absorption introduced by a student in a classroom increases with frequency from about 0.4 to 1.1 m<sup>2</sup> (global increase of 0.8 m<sup>2</sup>). These values agree reasonably well with

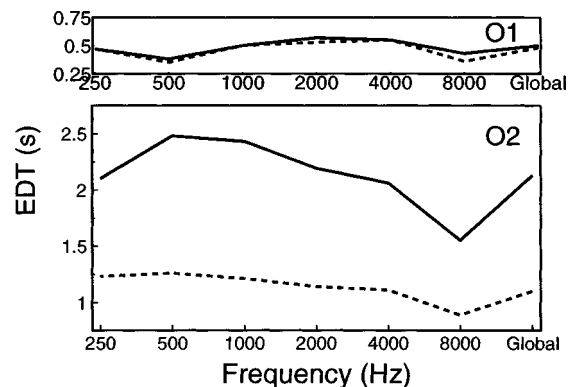


FIG. 12. Variation of measured octave-band and global EDT with source/receiver distance in classrooms O1 and O2: (solid line) unoccupied and (dashed line) occupied. Adapted from Ref. 1.

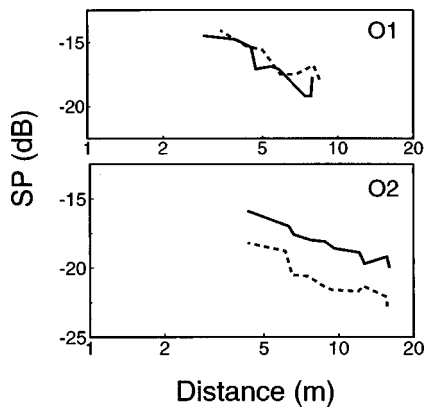


FIG. 13. Variation of measured 1000-Hz SP with source/receiver distance in classrooms O1 and O2: (solid line) unoccupied and (dashed line) occupied. Adapted from Ref. 1.

previously published ones;<sup>14</sup> that they are slightly higher is, again, likely due to the “fitting” effects discussed above.

## V. DISCUSSION

### A. EDT, $STI_{\infty}$ , and SLA in occupied classrooms

The results for EDT (or RT),  $STI_{\infty}$ , and SLA in the unoccupied classrooms, and for the changes of these quantities due to students, can be combined to estimate EDT,  $STI_{\infty}$ , and SLA in the occupied classrooms. This can be done with the help of classical diffuse-field theory, on the assumption—albeit, questioned above—that it is applicable to classrooms. Since the acoustical conditions in an occupied classroom depend on the number of students, corrections were done for the cases of the 30 classrooms when half-occupied and fully occupied. For each of the 30 classrooms, corrections were applied to the room-average, octave-band EDTs and global values calculated. Corresponding corrected global  $STI_{\infty}$ s were determined from the best-fit, power-law relationship discussed in Sec. IV A. Similarly, corrections were applied to the octave-band SPs and SLAs—this at each source/receiver distance. Logarithmic regression was applied

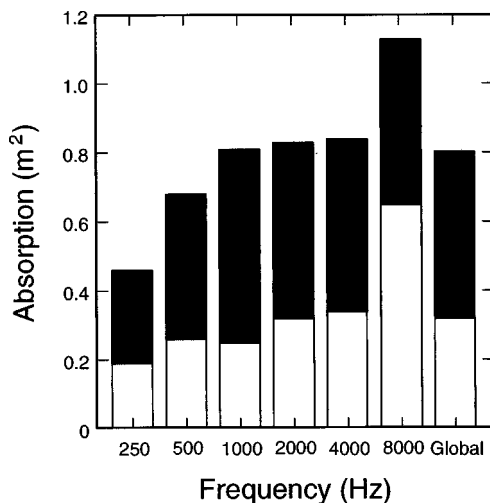


FIG. 14. Average (whole bar) and standard deviation (white bar) of the average absorption per person, measured in ten classrooms. Adapted from Ref. 1.

to the resulting “occupied” SLA versus distance data, as was done for the “unoccupied” data, in order to determine how the presence of students would affect the spatial rate of decrease of SLA.

Regarding GEDT, students reduced this quantity in all cases. The range of GEDTs was 0.36 to 0.86 s when half-occupied and 0.30 to 0.67 s when fully occupied (as compared to 0.47 to 1.54 s unoccupied). GEDTs in the half- and fully occupied classrooms decreased to from 0.49 to 0.83 and 0.33 to 0.72 of the unoccupied values, respectively. How these reductions varied from classroom to classroom is illustrated in Fig. 4, which shows the fully occupied, octave-band, and global GEDTs in classrooms U1–U4. The reduction in GEDT due to students was greatest in classroom U3 and least in classroom U1; the reduction was greatest in rooms with low unoccupied-room absorption and increased with room size. Figure 5 shows the frequency distribution of GEDT in the half- and fully occupied rooms. In the half-occupied case, GEDT exceeded 0.7 s in only 7 (23%) rooms and 0.5 s in only 14 (67%). In the fully occupied case, GEDT never exceeded 0.7 s and exceeded 0.5 s in only 7 (23%).

Regarding  $STI_{\infty}$ , students increased this quantity in direct relation to the reduction of reverberation. The range of  $GSTI_{\infty}$ s was 0.61 to 0.77 when half-occupied and 0.65 to 0.82 when fully occupied (as compared to 0.47 to 0.76 unoccupied). The  $GSTI_{\infty}$ s in the half- and fully occupied classrooms increased by a factor of from 0.04 to 0.12 and 0.07 to 0.19, respectively, relative to the unoccupied values. The increase was greatest in classrooms with low unoccupied-room absorption, and increased with room size. Figure 3 shows the frequency distribution of  $GSTI_{\infty}$  in the half- and fully occupied rooms. In the half-occupied case,  $GSTI_{\infty}$  was above 0.75 (excellent) in 6 (20%) rooms; the remainder (10%) were “good.” In the fully occupied case,  $GSTI_{\infty}$  was excellent in 17 (57%) rooms; the rest (43%) were “good.”

As for SLA, students decreased speech levels in all cases. The decrease in level increased with source/receiver distance. Thus, the spatial rate of decrease, and the decrease in speech level from the front to the back of the classroom, increased with students. In the half- (fully) occupied rooms, speech levels ranged from 47.0 to 64.4 (46.0 to 64.3) dB, and were 0.2 to 3.2 (0.3 to 4.9) dB lower than in the unoccupied rooms. In the extreme case of the fully occupied classrooms, the spatial rate of decrease varied from 0.6 to 4.5 (average = 2.3, standard deviation = 1.0) dB/dd, corresponding to expected front/back SLA differences varying from about 1 to 9 (average = 4.5) dB. Figure 9 shows the variations with distance of the SLAs in classrooms U1–U4 when fully occupied. The reduction of SLA, the increase of spatial rate of decrease, and the increase of front/back speech-level difference all were greatest in rooms, such as U2 and U3, with low unoccupied-room absorption, and increased with room size, being smallest in small, higher-absorption classroom U1.

In summary, the presence of students in a classroom has a significant effect on the acoustical conditions and, as a result, on verbal communication in a classroom. This effect must be taken into account in classroom design, and when setting classroom acoustical criteria. For example, at UBC, if unoccupied classrooms are considered, only small, absorbent



rooms meet the reverberation criteria. However, if the fully occupied condition is considered, only large classrooms do not meet the criteria.

### B. Combined effect of signal-to-noise ratio and reverberation

The above discussion and analysis only consider the effect of changes in reverberation and speech level independently, as per the objectives of the study. Of course, in practice, as mentioned above, different classroom designs and acoustical treatments affect both reverberation and speech level at the same time. With respect to the resulting combined effect of, for example, the presence of students on verbal communication, unfortunately the positive effect of decreased reverberation is offset by the negative effect of a reduction in speech level and an increase in noise due to student activity.<sup>15</sup> As discussed above, this can be quantified using the speech transmission index in its general form<sup>7</sup> and prediction models. Whether the result is better or worse conditions for verbal communication is not immediately obvious and likely depends on the details of the room. This will be discussed in detail elsewhere.

### C. Signal-to-noise ratio in occupied classrooms

In the analysis presented above, the range of SLAs was calculated for each classroom when half- and fully occupied, on the assumption that the instructor was speaking at half-way between an average normal and raised voice. In the best case of classroom U1 when half-occupied, SLAs varied from about 61 to 64 dB; in the worst case of classroom U4 when fully occupied, they varied from about 47 to 52 dB. If a signal-to-noise ratio of 15 dB is required for optimum verbal communication, then the background-noise level in these two rooms must not exceed 46 and 32 dB, respectively. Viewed differently, background-noise levels as high as about 49 dB were measured in the 30 unoccupied classrooms. We will ignore the expected reduction in noise level due to sound absorption by students, and its possible increase due to student activity which has been shown to be significant.<sup>15</sup> Then, this level of background noise would result in signal-to-noise ratios varying between 12 and 14 dB, and between -2 and +3 dB, in the best and worst classrooms U1 and U4, respectively. Of course, this discussion also ignores other factors, such as instructors changing their vocal efforts to compensate for noise, differences in speech levels generated by male and female instructors, and instructors whose speech levels are significantly higher or lower than average.

### D. Optimal design

Let us briefly consider what the results presented above tell us about optimal classroom design. Clearly, background noise levels must be sufficiently low. Consider, in particular, the GSTL<sub>∞</sub> and GSLA results for classrooms U1–U4. Clearly the effect of different designs (i.e., room geometries and acoustical treatments) on these two important quantities is complex and in opposite directions with respect to conditions for verbal communication. To achieve optimum (i.e., sufficiently low) reverberation, a classroom should be small and

sound absorptive. To achieve optimum (i.e., sufficiently high relative to the background noise) speech levels, the classroom should be small and nonabsorptive. Is it more important to meet the reverberation criteria or the signal-to-noise level criteria? The results suggest that it is easier to meet the reverberation criterion as long as classrooms are used fully occupied. Apparently small classrooms are best, but how should sound absorption be optimized? One approach is to determine the classroom size, determine the amount of absorbent treatment required to meet the reverberation criteria, then choose the classroom shape and the location of the sound absorption to promote high speech levels, particularly at the back of the classroom where they are lowest. A more detailed discussion of this topic and, for example, of the effect of classroom shape and absorption distribution on the acoustical conditions for speech, will be presented elsewhere.

## VI. CONCLUSION

The results of the UBC<sup>1</sup> classroom acoustical survey show that the unoccupied classrooms have far from optimum acoustical quality. However, the picture is much brighter when occupied classrooms are considered—at least with respect to reverberation. This demonstrates the significant effect of the presence of students on the acoustical conditions in classrooms, emphasizing the need to account for them in classroom design, and when setting design criteria. Many classrooms have excessive reverberation. Many have high background-noise levels due to the ventilation system and other sources. Speech levels may be low, and may decrease significantly from the front to the back of the room.

## ACKNOWLEDGMENTS

The author would like to thank students Callum Campbell, H el ene Illaire, John Kim, Elizabeth Nethery, Waqar-Un-Nissa Ressler, and Tony Skrjanec for their hard work in completing the two surveys and the associated analyses.

<sup>1</sup>M. R. Hodgson, "UBC-classroom acoustical survey," *Can. Acoust.* **22**(4), 3–10 (1994).

<sup>2</sup>R. H etu, C. Truchon-Gagnon, and S. Bilodeau, "Problems of noise in school settings: a review of the literature and the results of an exploratory study," *J. Speech Lang. Path. Audiol.* **14**(3), 31–39 (1990).

<sup>3</sup>C. Truchon-Gagnon and R. H etu, "Noise in day-care centres for children," *Noise Control Eng. J.* **30**(2), 57–64 (1988).

<sup>4</sup>J. S. Bradley, "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.* **80**, 837–845 (1986).

<sup>5</sup>J. S. Bradley, "Speech intelligibility studies in classrooms," *J. Acoust. Soc. Am.* **80**, 846–854 (1986).

<sup>6</sup>J. S. Bradley, "Relationships among measures of speech intelligibility in rooms," *J. Audio Eng. Soc.* **46**(5), 396–404 (1998).

<sup>7</sup>H. J. M. Houtgast and T. Steeneken, "A physical method for measuring speech transmission quality," *J. Acoust. Soc. Am.* **67**, 318–326 (1980).

<sup>8</sup>*American National Standard Method for the Calculation of the Speech Intelligibility Index*, ANSI S3.79-1998.

<sup>9</sup>J. S. Bradley, "Uniform derivation of optimum conditions for speech in rooms," Report BRN 239, National Research Council Canada (1985).

<sup>10</sup>K. S. Pearsons, R. L. Bennett, and S. Fidell, "Speech levels in various noise environments," EPA Report EPA-600/1-77-025 (1977).

<sup>11</sup>J. S. Bradley, "Speech intelligibility in simulated classroom sound fields," *J. Acoust. Soc. Am.* **105**, 1259 (1999).

<sup>12</sup>M. R. Hodgson, "When is diffuse-field theory applicable?" *Appl. Acoust.* **49**(3), 197–207 (1996).

<sup>13</sup>M. R. Hodgson, "Measurements of the influence of fittings and roof pitch on the sound field in panel-roof factories," *Appl. Acoust.* **16**(5), 369–391 (1983).

<sup>14</sup>L. Cremer and H. A. Müller, *Principle and Applications of Room Acous-*

*tics, Vol. 1* (Applied Science, New York, 1982), Sec. II.6.6.

<sup>15</sup>M. Hodgson, R. Rempel, and S. Kennedy, "Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures," *J. Acoust. Soc. Am.* **105**, 226–233 (1999).

# On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility

J. S. Bradley, R. D. Reich, and S. G. Norcross

*Institute for Research in Construction, National Research Council, Montreal Road,  
Ottawa K1A 0R6, Canada*

(Received 24 November 1998; revised 2 June 1999; accepted 15 June 1999)

Speech intelligibility in rooms is influenced by room acoustics effects and by the signal-to-noise ratio (S/N) of the speech and ambient noise. Several measures such as useful-to-detrimental sound ratios and the speech transmission index predict the combined effects of both types of factors. These measures were evaluated relative to speech intelligibility test results obtained in simulated sound fields. The use of simulated sound fields made it possible to create the full range of combinations of room acoustics and S/N effects likely to be found in rooms for speech. The S/N aspect is shown to be much more important than room acoustics effects and new broadband useful-to-detrimental ratios were validated. Useful-to-detrimental ratios, speech transmission index measures, and values of the articulation loss for consonants were all reasonably accurate predictors of speech intelligibility. Further improvements to these combined measures are suggested.

[S0001-4966(99)01210-2]

PACS numbers: 43.55.Hy, 43.71.Gv [JDQ]

## INTRODUCTION

The intelligibility of speech in rooms is influenced by both the signal-to-noise ratio and the room acoustics characteristics of the space. The two types of effects are not independent and room acoustics characteristics can influence signal-to-noise ratios by modifying the received sound levels of the speech and noise. Few studies have systematically investigated the combined effects of signal-to-noise ratio (S/N) and room acoustics on speech intelligibility and the relation of speech intelligibility scores with measures that combine both types of factors because it is difficult to systematically vary room acoustics conditions in real rooms. That is, one can readily vary speech or noise levels to create a range of S/N values, but one cannot easily create a similar wide range of varying room acoustics conditions at a single location. It is usually difficult to create more than a small number of room acoustics conditions and changing between them usually takes too long to include varied room acoustics in a single test of speech intelligibility.

The current work investigates the combined effects of a wide range of S/N and room acoustics conditions using synthesized sound fields. Listeners sat in realistic three-dimensional sound fields in which either the S/N or the room acoustics characteristics (or both) could be varied instantaneously under computer control. The range of both types of variations was representative of all likely conditions in rooms for speech. The use of synthesized sound fields provides a unique opportunity to better understand the combined effects of S/N and room acoustics on speech intelligibility in rooms.

The effects of S/N on the intelligibility of speech have been investigated thoroughly many years ago and measures such as the Articulation Index<sup>1</sup> accurately indicate the expected effects of speech and noise levels on intelligibility. One can also use simpler measures such as the difference of the A-weighted speech and noise levels [S/N(A)] as approxi-

mate indicators of the effects of S/N on speech intelligibility.

Although rooms add a complex sequence of reflected sounds, room acoustics has traditionally been characterized simply by reverberation time. However, it is now known that the early decay time (the reverberation time measured from the slope of the first 10 dB of the decay) is usually a better indicator of perceived reverberance. It is also known that while some reflections degrade intelligibility, others can improve it. Since the pioneering work of Haas,<sup>2</sup> we know that we do not hear individual early-arriving sound reflections as separate events but they are subjectively integrated with the direct sound. The combined effect of these early-arriving reflections is to increase the intelligibility of speech because they make the direct sound seem louder. Later-arriving reflections are not integrated with the direct sound and degrade speech intelligibility by causing one speech sound to blur into the next. Typically, the boundary between early- and late-arriving reflections has been taken to be 50 ms for speech sounds.<sup>3,4</sup>

Because early-arriving reflections are beneficial to speech intelligibility and later-arriving reflections are not, the ratio of early-arriving to later-arriving sound has been used as an indicator of the effects of room acoustics on the clarity and intelligibility of speech.<sup>4</sup> For example,  $C_{50}$  is defined,

$$C_{50} = 10 \log \left\{ \frac{\int_0^{0.05} p^2(t) dt}{\int_{0.05}^{\infty} p^2(t) dt} \right\}, \text{ dB}, \quad (1)$$

where  $p(t)$  is the instantaneous pressure in the room impulse response. A linear early-arriving to total sound ratio (referred to as Deutlichkeit or Definition, ( $D$ )<sup>3</sup>) also has been proposed. Although the two measures can be exactly related mathematically, recent studies indicate that  $C_{50}$  is better because it is linearly related to subjective assessments of the clarity of speech sounds.<sup>5</sup>

Several measures have been proposed that combine the effect of room acoustics and S/N into a single quantity. Lochner and Burger<sup>6</sup> created the useful-to-detrimental ratio measure by adding a S/N component to the early-to-late-arriving sound ratio concept. Useful-to-detrimental ratios were defined as the logarithmic ratio of the useful to the detrimental sound. The useful sound was just the early-arriving speech energy and the detrimental sound was the sum of the later-arriving speech energy and the ambient noise. In Lochner and Burger's work the early energy sum involved a complicated scheme of weighting factors varying with the amplitude and time of arrival of the early reflections. The useful-to-detrimental ratio concept was validated by Latham<sup>7</sup> and further studies by Bradley<sup>8,9</sup> indicated that in real rooms simple (unweighted) sums worked as well as Lochner and Burger's original proposals. For a 50-ms early time limit, the useful-to-detrimental ratio is defined,

$$U_{50} = 10 \log \left\{ \frac{E}{L+N} \right\}, \text{ dB.} \quad (2)$$

Here,  $E$  is the early-arriving speech energy,  $L$  the later-arriving speech energy, and  $N$  the ambient noise energy.

Where steady state speech and noise levels as well as  $C_{50}$  values are separately available,  $U_{50}$  values can be calculated as follows:

$$U_{50} = 10 \log \left\{ \frac{E/L_{50}}{1 + (E/L_{50} + 1)N/S} \right\}, \text{ dB,} \quad (3)$$

where  $N$  is the ambient noise energy and  $S$  is the speech energy.  $E/L_{50}$  is the linear early-to-late arriving sound ratio. That is,

$$C_{50} = 10 \log(E/L_{50}). \quad (4)$$

Peutz<sup>10</sup> proposed the articulation loss concept. This is essentially the amount by which speech intelligibility scores are reduced below 100%. The calculated percent articulation loss for consonants ( $AL_{\text{cons}}$ ) is another measure that combines both room acoustics and S/N parameters to predict the expected effect on speech intelligibility. Several variations have been developed and this measure has mostly been used for evaluating sound reinforcement systems. Calculated  $AL_{\text{cons}}$  values were obtained following procedures defined in Ref. 11 and include the effects of direct and reverberant speech sound levels as well as noise levels and reverberation times.

The Speech Transmission Index (STI), developed by Houtgast and Steeneken,<sup>12,13</sup> is the most recent combined measure. It is based on the assumption that the degradation of speech intelligibility in rooms is related to reductions of the amplitude modulations of speech signals by both room acoustics and ambient noise. The STI is therefore based on modulation transfer functions. The reduction in amplitude modulations is assessed for the 98 combinations of 7 acoustical octave bands and 14 modulation frequency bands. These are reduced to a single number STI value in a manner somewhat similar to the Articulation Index procedure.<sup>13</sup> The STI measure has been shown to be successful for a wide range of applications. A simplification of the STI referred to as RASTI is based on only nine measured modulation reduc-

tions obtained in only the 500- and 2-kHz octave bands. The three types of combined measures ( $U_{50}$ ,  $AL_{\text{cons}}$ , and STI) are mathematically quite different and it is not clear whether they represent similar forms of combinations of room acoustics and S/N effects on speech intelligibility. A recent study showed that measured values of all three types of combined measures are quite strongly correlated for a large number of different sound fields.<sup>11</sup> However, it was not possible to evaluate the importance of differences among the measures because no speech intelligibility test results were available for the same sound fields. In fact, no previous study could be found that has extensively compared the merits of these three types of combined measures as predictors of speech intelligibility in rooms. Further, no previous studies have included extensive systematic examinations of the combined effects of room acoustics and S/N aspects on speech intelligibility and combined measures of speech intelligibility.

The present work was intended to first determine the relative importance of S/N and room acoustics effects on speech intelligibility over a broad range of both factors that were representative of conditions found in rooms for speech. It was further intended to evaluate whether existing combined measures satisfactorily predict the combined effects of S/N and room acoustics on speech intelligibility and whether there are substantial differences among the three types of combined measures. Many possible combined measures may give approximately correct predictions for more average conditions but when systematically tested may be found less satisfactory for various more extreme cases. For example, they may be satisfactory for average room acoustics and S/N, but not for a very good room acoustics and low S/N combination. It was also intended to compare the more sophisticated combined measures with simpler approaches such as combinations of S/N and decay times (reverberation time and early decay time). It was thought that the improved understanding that was expected to result from this new work would provide a more reliable basis for future attempts to determine acceptable criteria for speech in rooms. For example, some previous suggestions that very short reverberation times are preferred seem to ignore the benefits of reflected sound on intelligibility. When S/N and room acoustics are both considered, optimum conditions for speech seem to be less sensitive to reverberation time than sometimes suggested.<sup>14</sup>

## I. EXPERIMENTAL PROCEDURES

### A. Subjects and speech test

Speech intelligibility scores were obtained using a Fairbanks rhyme test as modified by Latham<sup>7</sup> and used in previous tests in rooms.<sup>8,9</sup> The original digital recording of the source material from the previous studies was transferred to compact disc for more convenient presentation in the current tests. Four lists of 50 words were used and subjects had to identify only the initial consonant of each test word. Test words were presented by a male talker in the carrier phrase, "Word number \_ is \_, write that down." The material was presented at an average rate of three syllables per second. A total of 10 subjects completed the tests. They were all adults



TABLE I. Orientation of loudspeakers relative to the listener. Horizontally straight ahead of the listener at ear level is 0 deg in both planes.

Loudspeaker	Horizontal angle, deg	Vertical angle, deg
Center low	0	0
Center high	0	20
Left low	-32	0
Right low	+32	0
Left high	-37	23
Right high	+37	23
Far left	-90	0
Far right	+90	0

and varied in age from 20 to 62 years old. They had no known hearing problems and were intended to be representative of a broad range of typical listeners. It was only possible to test one subject at a time in the simulated sound fields and no attempt was made to visually disguise the loudspeakers and anechoic room setup. Each subject listened to a complete list of 50 test words for each of 16 different sound field configurations after one complete practice test. The tests were presented in randomized order and subjects performed a sequence of four tests on four different days to avoid fatigue. Each test was preceded by three practice words heard via exactly the same sound field as the subsequent complete test. The final speech intelligibility scores were simply the percentage of correct responses. The average scores of all 10 subjects were used in the following analyses.

Of course other speech tests or even other talkers could lead to slightly different speech intelligibility scores. However, the procedures described here were thought to provide a valid and efficient basis for the comparisons of the combined effects of S/N ratios and room acoustics characteristics.

## B. The simulated sound fields

Sound fields were simulated in an anechoic room using an eight-channel electro-acoustic system. The sounds were reproduced by eight loudspeakers, one for each of the eight channels. The signal to each loudspeaker was created by various computer controlled time delays, equalizers, and digital reverberators. Each channel contributed a discrete reflection and a reverberant tail to the overall sound field. Yamaha DEQ7 digital equalizers were used to control the amplitude and delay time of the early reflections. Separate channels of Lexicon Reflex digital reverberators were used to create the reverberant decays from each of the eight loudspeakers. Bookshelf type loudspeakers were used (Paradigm Compact Monitor) with an on-axis frequency response that was equalized to be flat  $\pm 3$  dB from 80 Hz to 12 kHz. (Of course, in an anechoic room only the on-axis responses is relevant.) The loudspeakers were located approximately 1.8 m from the listener. The orientation of the eight loudspeakers relative to the listener are given in Table I.

The discrete reflections were distributed over the first 50 ms after the arrival of the direct sound and with amplitudes generally decreasing with time after the direct sound. Figure 1 illustrates the measured impulse response of one of the

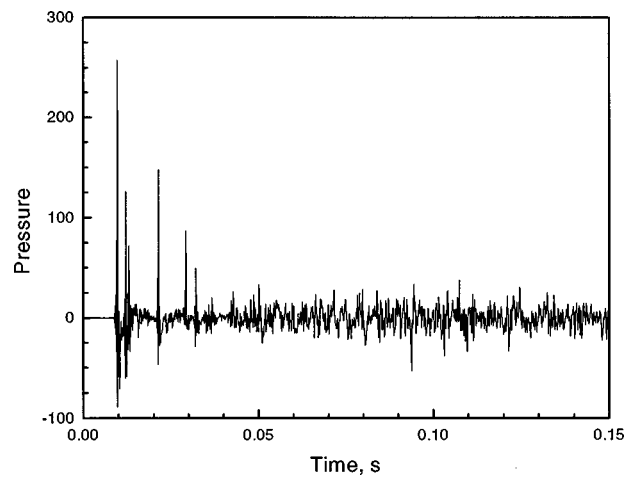


FIG. 1. Example of initial part of measured impulse response of simulated sound field.

simulated sound fields. To simulate a realistic room situation, the loudspeaker directly in front of the listener simulated the direct sound while the others produced early arriving reflections lower in amplitude than the direct sound. The speech source was clearly localized on the loudspeaker directly in front of the listener. (Subjects were free to turn their heads but would normally focus on the loudspeaker directly in front of them.) As illustrated in Fig. 1, the reverberant tails started at approximately 50 ms after the direct sound with small variations in the start time and reverberator setting to avoid highly correlated reverberant sounds from the eight loudspeakers. By adjusting the levels of early reflections as well as the level and reverberation time of later-arriving sound, sound fields were created with a range of  $C_{50}$  values.

In most real rooms the early sound field would include more than the direct sound and seven early reflections. Thus the simulations are only an approximation to the complexity of typical sound fields in rooms. Although other factors such as reflection density may be perceptually noticeable, it is thought that they are second order effects and that the total early sound energy and in particular the  $C_{50}$  value will indicate the predominant effects of the room acoustics on speech intelligibility.

In all cases the overall level of the speech at the listener was 55 dBA. This was measured at the location of the listener's head without the listener present. Ambient noise was simulated using an NC 40 shaped spectrum. The level of this noise was raised or lowered to create A-weighted speech to A-weighted noise levels of -5, 0, +5, and +10 dB. The noise was reproduced equally from all eight loudspeakers and the noise signals were passed through additional reverberators with very short reverberation times to avoid the unwanted effects of correlated noise signals at the listening position.

A total of 16 different sound fields were created consisting of the combinations of 4 different S/N values and 4 different room acoustics situations. The four S/N values as given above were -5, 0, +5, and +10 dB. The four different room acoustics cases are described in Table II in terms of the mid-frequency values of  $C_{50}$ , reverberation time (RT), and early decay time (EDT). In these sound fields the variation of

TABLE II. Measured mid-frequency (1 kHz) values of  $C_{50}$ , EDT, and RT for the four different room acoustics cases.

$C_{50}$	EDT	RT
-3.06	1.88	1.95
1.10	1.42	1.45
5.17	1.05	1.01
9.15	0.43	0.56

$C_{50}$ , EDT, and RT are clearly strongly related and for these particular sound fields any one of these quantities could be used to describe the room acoustics conditions. However, for simplicity sound fields will be identified by their  $C_{50}$  value. The four room acoustics cases were intended to represent the full range of  $C_{50}$  values likely to occur in rooms for speech. Previous results<sup>9</sup> suggested that the  $C_{50}=+9$  dB case would correspond to near perfect room acoustics conditions for speech intelligibility. A  $C_{50}$  of less than  $-3$  dB (corresponding to an RT of greater than 2 s) would not represent conditions likely to occur in rooms for speech. The fact that the EDT and RT values for each case are similar indicates that the sound decays are quite linear.

The four different S/N cases would correspond to ambient noise levels of 45 to 60 dBA. Previous studies have shown that a S/N of +15 dB corresponds to a speech intelligibility score of 100%<sup>8,9,15</sup> and that a S/N of +10 dB leads to only slightly reduced intelligibility. Ambient noise levels of greater than 60 dBA, corresponding to a S/N of less than  $-5$  dB, would relate to nearly impossible speech communication. Thus the range of ambient noise and S/N values represents the full range of conditions of practical interest.

Sound fields were measured using our Room Acoustics Measurement Software (RAMSoft). This measures the impulse response of the sound field and determines values of  $C_{50}$ ,  $C_{80}$ ,  $D$ , EDT, and RT in octave bands from 125 to 4 kHz. Impulse responses are obtained using a maximum length sequence signal and a Fast Hadamard transform procedure.<sup>16</sup> From the measured impulse responses and the speech and noise levels, further speech-related measures were calculated in a second speech measure calculation program that used the measured impulse responses and ambient noise levels as input. This produced values of the Articulation Index (AI), and signal-to-noise ratio of the A-weighted speech and noise levels [S/N(A)] as well as values of the three types of combined measures. Useful-to-detrimental ratios for 50- and 80-ms early time intervals (i.e.,  $U_{50}$  and  $U_{80}$ ) were calculated in octave bands from 125 to 4 kHz. ALcons values were calculated for the same octave bands. STI and RASTI values were calculated both with and without the effects of speech and noise levels. The complete measures including the effects of speech and noise levels are referred to as STI+N and RASTI+N. Excluding the effects of speech and noise levels they are referred to as STI-N and RASTI-N. These latter variations assess only the room acoustics effects. The procedures for calculating the various speech measures were successfully validated in previous studies by comparison with other measurement systems.<sup>17</sup>

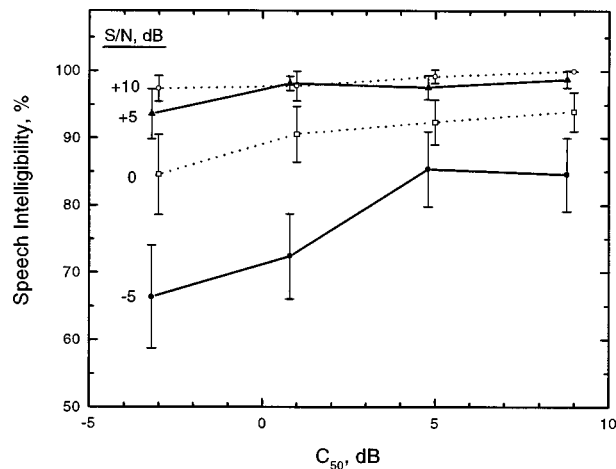


FIG. 2. Mean speech intelligibility scores for each of the 16 conditions and the corresponding 95% confidence limits. (The S/N = +5 dB and  $-5$  dB cases have been offset to avoid overlapping error bars.)

## II. SPEECH INTELLIGIBILITY RESULTS

The speech intelligibility test scores of the 10 subjects for each of the 16 sound fields were first subject to an analysis of variance. This showed highly significant main effects of both room acoustics ( $C_{50}$ ) and S/N ( $p < 0.001$ ). In addition there was also a significant interaction effect ( $p < 0.001$ ). However, the main effect of S/N ( $F$  ratio 119.06) on speech intelligibility scores was much stronger than that for  $C_{50}$  ( $F$  ratio 20.85) and the interaction effect ( $F$  ratio 4.65) was much weaker. Thus the S/N is the much more important determinant of speech intelligibility than  $C_{50}$  for the 16 test conditions in this experiment that represent the full range of acoustical conditions likely to occur in rooms for speech.

These analysis of variance results are illustrated in Fig. 2, which shows the mean speech intelligibility score for each of the 16 conditions as well as the 95% confidence limits. As the analysis of variance results indicate, there is an overall significant trend for speech intelligibility to increase with increasing  $C_{50}$  (improving room acoustics) and also with increasing S/N. Further, the effects of room acoustics ( $C_{50}$ ) are larger for lower S/N values (i.e., there is a significant interaction effect). For the highest S/N case (S/N = +10 dB), there is only a very small effect of varied  $C_{50}$  ( $\sim 2.5\%$ ) over the full range of likely room acoustics conditions. That is, if there are minimal S/N problems, the effect of room acoustics, from very bad to near optimum, changes speech intelligibility by no more than 2.5%. Even when S/N is only 0 dB, the maximum effect of  $C_{50}$  on mean speech intelligibility scores is no more than 10%. On the other hand, for near optimum room acoustics conditions (i.e.,  $C_{50}=+9$  dB), varying S/N varies speech intelligibility scores by as much as 15% and for less optimum room acoustics of  $C_{50}=+1$  dB, varying S/N leads to a 25% change in speech intelligibility scores. Clearly, the S/N is the much more important determinant of speech intelligibility than is  $C_{50}$ .

Further confirmation of the relative importance of the room acoustics and S/N factors was obtained by relating the test scores to separate measures of the room acoustics and S/N factors. As will be seen later (Fig. 3) the relationships

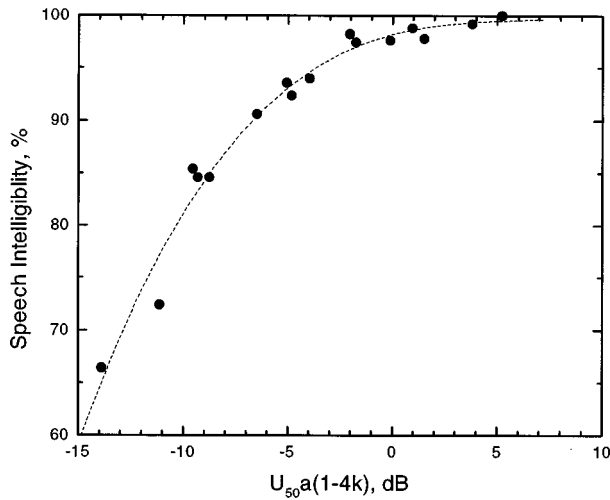


FIG. 3. Example of third order polynomial fit of broadband useful-to-detrimental ratio [ $U_{50a}(1-4k)$ ] to speech intelligibility scores.

between many acoustical measures and the speech intelligibility test scores were well described by third order polynomial fits to the data. Thus for convenient comparisons all analyses used only third order polynomial fits to the speech intelligibility scores. Articulation index values (AI) and A-weighted speech to noise ratios [S/N(A)] were used as measures of S/N effects. The polynomial fits of speech intelligibility scores to these measures resulted in  $R^2$  values of 0.859 for AI values and 0.779 for S/N(A) values. ( $R^2$  is the coefficient of determination and in all cases the number of data points is 16.)

Similar third order polynomial fits were calculated for a range of room acoustics measures. These included various octave band combinations of  $C_{50}$ ,  $C_{80}$ , and EDT as well as RASTI- $N$  and STI- $N$  values. The  $R^2$  values were all very similar and never exceeded 0.136.

Again the S/N effects are clearly the much stronger determinant of speech intelligibility and the S/N measures explained 78%–86% of the variance in responses. None of the room acoustics measures explained more than 14% of the variance in speech intelligibility scores.

### III. DEVELOPING A BROADBAND USEFUL-TO-DETRIMENTAL RATIO MEASURE

Before comparing the merits of the three types of combined measures, it was necessary to first construct a broadband useful-to-detrimental measure because this had not been done in previous studies. Three different approaches were tried. The first was to sum useful-to-detrimental ratios over various octave bands. The second approach was to use frequency-weighted sums, and the third approach was to base the broadband measure on the approach described in Eq. (3) and to calculate useful-to-detrimental ratios from broadband values of early-to-late ratios, speech, and noise levels. Although it was assumed that an early time interval of 50 ms was more appropriate for speech, broadband measures were also determined with an 80-ms early time interval.

For each sound field, useful-to-detrimental ratios were calculated from measured impulse responses combined with

TABLE III. Results of third order polynomial fits of various broadband useful-to-detrimental ratio measures to speech intelligibility scores. In all cases there were 16 data points.

Measure	$R^2$	Measure	$R^2$
$U_{50a}(1-4k)$	0.973	$U_{80a}(1-4k)$	0.943
$U_{50a}(500-4k)$	0.969	$U_{80a}(500-4k)$	0.958
$U_{50a}(A)$	0.972	$U_{80a}(A)$	0.969
$U_{50a}(AI)$	0.971	$U_{80a}(AI)$	0.969
$U_{50}[C_{50}(500-4k)]$	0.961	$U_{80}[C_{80}(500-4k)]$	0.958

speech and noise levels as indicated by Eq. (2). As in previous studies,<sup>8,9</sup> two variations of useful-to-detrimental ratios were calculated. In the first type, the early-and late-arriving speech energy as well as the ambient noise energy were all octave band measures. In the second type, the early and late speech energies were octave band measures but the ambient noise was a broadband A-weighted level. [This is indicated by the lower case ‘‘a,’’ e.g.,  $U_{50a}(1-4k)$ .] The latter type of measure had previously been found to be more successful as a predictor of speech intelligibility scores. In the present work this was again found to be true and only useful-to-detrimental ratios based on broadband A-weighted ambient noise levels are presented here.

The relative merits of various broadband measures were assessed by calculating third order polynomial fits with speech intelligibility scores. Figure 3 gives an example of such a polynomial fit for values of the measure  $U_{50a}(1-4k)$ , described below. Since the form of the fit and the number of data points (16) was the same in every case, the success of each measure can be compared in terms of the corresponding  $R^2$  values. These are given for the more successful combinations in Table III.

Two different octave band sums were quite successful. Summing over the three octave bands from 1 to 4 kHz was most successful and led to an  $R^2$  value of 0.973 for  $U_{50a}(1-4k)$  shown in Fig. 3. [ $U_{50a}(1-4k)$  indicates a useful-to-detrimental sound ratio with a 50-ms early time interval, summed over the octave bands from 1 to 4 kHz and where the noise component is the broadband A-weighted noise level.] Including the four octave bands from 500 to 4 kHz was essentially equally successful. Because the rhyme test evaluates the intelligibility of consonants, it is to be expected that higher frequencies will be better predictors of speech intelligibility scores in these tests. However, including the 500-Hz octave band as in  $U_{50a}(500-4k)$  may be a more generally applicable broadband measure.

The second type of broadband useful-to-detrimental ratio measure was obtained by frequency-weighted sums of the octave band useful-to-detrimental ratio values. Both A-weighted and AI (articulation index) weighted sums were calculated. As indicated by the  $R^2$  values in Table III, these were essentially equally well correlated with speech intelligibility scores as the best octave band sum type measure and the A-weighted sums were as successful as the AI weighted sums. These frequency-weighted sums may be more generally reliable as broadband measures because they do include effects in all frequency bands of interest.

The final type of broadband measure was based on the

approach of Eq. (3). It was assumed that this approach would be practically useful because several commercial measurement systems can provide values of  $C_{50}$ , as well as the speech and noise levels required as input to this equation. The most successful measure in this format was  $U_{50}[C_{50}(500-4k)]$ . This was based on the arithmetic average of  $C_{50}$  values over the octaves from 500 to 4 kHz. Combining these average  $C_{50}$  values with the A-weighted speech and noise levels produced  $U_{50}[C_{50}(500-4k)]$  values and the associated  $R^2$  value is given in Table III. This measure is essentially equally strongly correlated with speech intelligibility scores as are the A- and AI-weighted summed measures. When the differences among the various correlations in Table III for various  $U_{50}$  measures were tested, none of the differences were statistically significant. Thus all of the  $U_{50}$  measures are equally acceptable and the choice of measure would depend on other details such as whether octave band useful-to-detrimental ratio values are available.

Table III also gives  $R^2$  values for the corresponding useful-to-detrimental ratios based on an 80-ms early time interval. Although a 50-ms early time interval is usually recommended for speech, the results in Table III suggest that 80 ms could also be acceptable and  $U_{80}$  values are only slightly less well-correlated with speech intelligibility. When the differences among the various correlations with  $U_{80}$  measures in Table III were compared with the correlation for  $U_{50}a(A)$  values, only the correlation for the  $U_{80}a(1-4k)$  was significantly different ( $p < 0.05$ ).

On balance, the A-weighted sum measure [ $U_{50}a(A)$ ] represents the most satisfactory broadband measure. If it is necessary to calculate useful-to-detrimental ratios from  $C_{50}$  values, speech and noise levels, then the  $U_{50}[C_{50}(500-4k)]$  measure should be equally accurate. Some other measures may be very slightly better predictors of the current rhyme test scores but may not be as generally reliable. Measures using an 80-ms early time interval seem to be equally satisfactory.

Because different types of broadband useful-to-detrimental ratios are possible, it is useful to determine how they are inter-related. Plots of the relationships between pairs of the more successful broadband measures included in Table III were all simple linear relationships with slopes of approximately 1. Thus there are essentially only small constant differences between the pairs of measures. The differences would depend somewhat on the range of sound fields that are included but the following can be used as approximate initial conversion factors. The A-weighted broadband measure  $U_{50}a(A)$  produced values that were approximately 0.8 dB smaller than the values of the measures that summed 500 to 4 kHz data [i.e.,  $U_{50}a(500-4k)$  and  $U_{50}[C_{50}(500-4k)]$ ]. The  $U_{50}a(1-4k)$  values were approximately 2.4 dB smaller than the  $U_{50}a(A)$  values. The broadband measures based on an 80-ms early time interval were approximately 0.6 dB larger than the corresponding values based on a 50-ms early time interval.

TABLE IV. Comparison of  $R^2$  values from third order polynomial fits of combined measures to speech intelligibility scores.

Combined measure	$R^2$
$U_{50}a(A)$	0.972
$U_{50}[C_{50}(500-4k)]$	0.961
$U_{80}a(A)$	0.969
$U_{80}[C_{80}(500-4k)]$	0.958
STI+N	0.936
RASTI+N	0.738
AL <sub>cons</sub> (2k)	0.930

#### IV. COMPARISONS OF COMBINED MEASURES AS PREDICTORS OF SPEECH INTELLIGIBILITY

The newly derived broadband useful-to-detrimental ratio measures were next compared with other measures that combine the effects of both room acoustics and S/N. The various combined measures were compared by testing how well they correlated with the speech intelligibility scores. In all cases third order polynomials provide very good fits to the measured data. Because the number of data points was always 16, the  $R^2$  values for each third order polynomial fit can be used as a simple means of comparing the prediction accuracy of the combined measures. Figure 3 gave an example of such a third order polynomial fit with a related  $R^2$  value of 0.973. Table IV compares the resulting  $R^2$  values for seven different combined measures. Some of these are repeated from Table III for convenience. The most accurate predictor,  $U_{50}a(A)$ , explains 97% of the variance in speech intelligibility scores. The other useful-to-detrimental ratio measures explain 96%–97% of the response variance. The complete speech transmission index measure, STI+N, explained 94% of the response variance and RASTI+N only 74%. The AL<sub>cons</sub> measure is typically only used as a 2-kHz octave band measure. This combined measure explained 93% of the variance in speech intelligibility scores. When the differences among the correlations in Table IV were tested, only the correlation with RASTI+N values was significantly different to the correlation with  $U_{50}a(A)$  values. Thus averaged over all 16 sound fields all of the measures except RASTI+N are equally satisfactory predictors of speech intelligibility scores.

Although the average performance of most of the measures was similar, when examined in more detail particular extreme cases are more or less well predicted by some of the measures. Plots such as Fig. 3 are not particularly helpful because they only exhibit more or less scatter with no indication as to which combination of room acoustics and S/N caused the most problems. To better illustrate detailed properties of each of these measures, plots similar in form to Fig. 2 were produced. That is, speech intelligibility scores were plotted versus  $C_{50}$  values separately for each S/N case resulting in plots with four separate sets of data. These included both the measured speech intelligibility scores (as data points and 95% confidence limits) and the predicted speech intelligibility according to one of the combined measures (as lines).

Figure 4 shows the first of these plots comparing measured and predicted speech intelligibility obtained from the



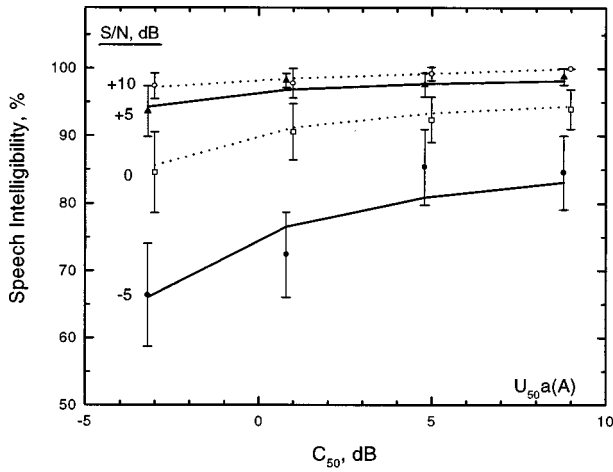


FIG. 4. Comparison of measured and predicted speech intelligibility scores based on the  $U_{50}a(A)$  measure. (The  $S/N=+5$  dB and  $-5$  dB cases have been offset to avoid overlapping error bars.)

$U_{50}a(A)$  measure. It is seen that the lines showing the predicted speech intelligibility scores fit the measured data quite well for all combinations of  $C_{50}$  and  $S/N$  values. Figure 5 provides very similar results for predictions based on the  $U_{50}[C_{50}(500-4k)]$  measure. In the high  $S/N$ —low  $C_{50}$  region ( $C_{50}=-3$  dB,  $S/N=+10$  dB) predictions agree a little less well with measured intelligibility scores than in Fig. 4 for  $U_{50}a(A)$  values. That is, for this extreme condition the predicted intelligibility is outside the confidence limits of the measured values. Although not shown, the plot for  $U_{80}a(A)$  values indicated similar agreement as would be expected from the high  $R^2$  value.

Figure 6 compares measured and predicted speech intelligibility based on the  $STI+N$  measure. Although on average the predictions match the measured values, there are areas where the fit is clearly less satisfactory. For the  $C_{50}=-3$  dB,  $S/N=+10$  dB case the predicted intelligibility is outside the confidence limits of the measured speech intelligibility scores. More generally the predicted intelligibilities seem to vary too much with variations in  $C_{50}$ . That is, the effect of room acoustics is too large relative to the effect of

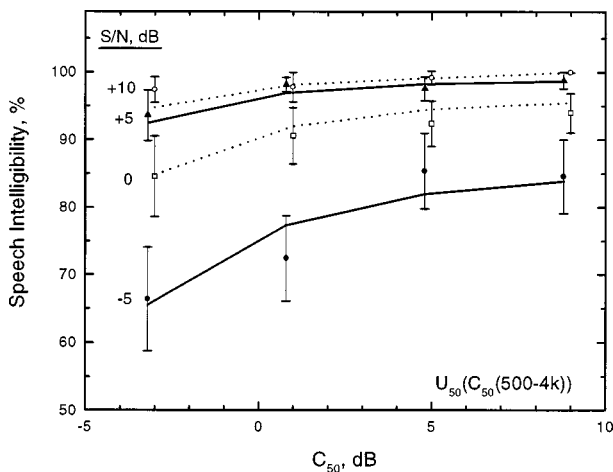


FIG. 5. Comparison of measured and predicted speech intelligibility scores based on the  $U_{50}[C_{50}(500-4k)]$  measure. (The  $S/N=+5$  dB and  $-5$  dB cases have been offset to avoid overlapping error bars.)

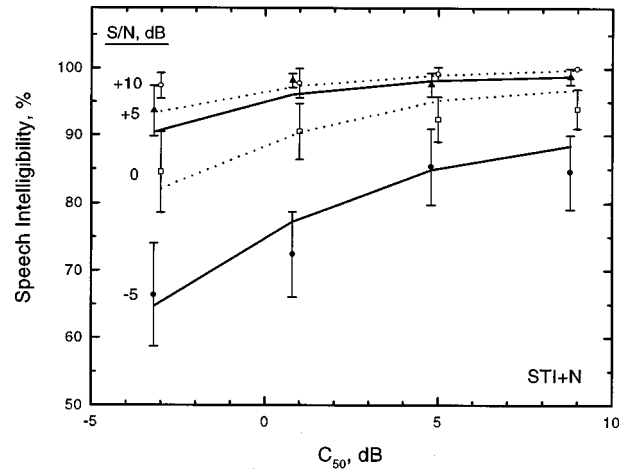


FIG. 6. Comparison of measured and predicted speech intelligibility scores based on the  $STI+N$  measure. (The  $S/N=+5$  dB and  $-5$  dB cases have been offset to avoid overlapping error bars.)

$S/N$ . Although these are small problems, it would seem possible that some adjustments to the  $STI+N$  measure would improve its prediction accuracy in some particular situations. Although not shown, the plot for  $RASTI+N$  values showed similar trends but with more scatter. This was expected as the  $RASTI+N$  measure is a simplification of the  $STI+N$  measure and is based on the same basic combination of room acoustics and  $S/N$  effects.

Figure 7 compares measured and predicted speech intelligibility based on the 2-kHz octave band  $AL_{cons}$  measure. Although the overall fit is quite good ( $R^2=0.930$ ), there are clearly areas where a better fit would improve the prediction accuracy of this measure.

In the past, acoustical conditions for speech have been assessed in terms of ambient noise levels and reverberation times. Linear combinations of these types of parameters can be devised to use as combined predictors of speech intelligibility. Several possibilities were tried and compared with the more sophisticated measures considered above. Linear combinations of  $EDT$ ,  $S/N(A)$ , and  $S/N(A)^2$  could explain ap-

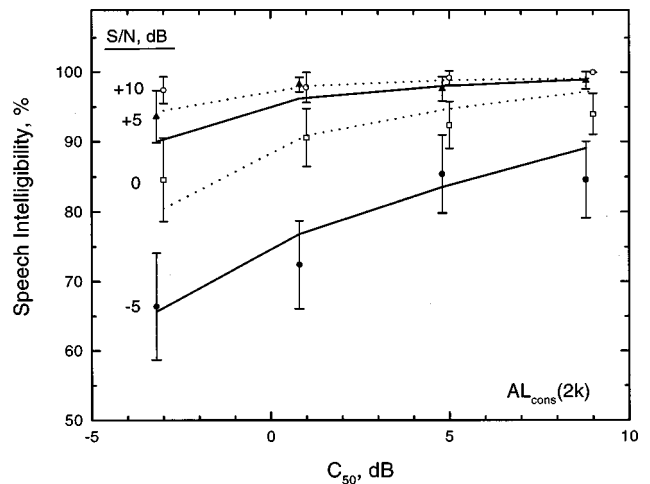


FIG. 7. Comparison of measured and predicted speech intelligibility scores based on the  $AL_{cons}(2k)$  measure. (The  $S/N=+5$  dB and  $-5$  dB cases have been offset to avoid overlapping error bars.)

proximately 90% of the variance in speech intelligibility scores. However, a linear combination of EDT, AI, and  $AI^2$  values explained 94% of the variance in speech intelligibility scores. When these possible combined predictors were examined in plots similar to those of Figs. 4–7, it was clear that these measures do not match the measured data well for all combinations of  $C_{50}$  and S/N. Of course, other more complex forms of combining these quantities might be more successful.

## V. IMPROVED COMBINATIONS OF ROOM ACOUSTICS AND S/N EFFECTS

While the three main types of combined measures are all quite highly correlated with speech intelligibility scores, there are some indications in Figs. 4–7 that some improvements are possible. That is, a slightly different combination of room acoustics measure and S/N measure might lead to better predictions of speech intelligibility in some particular cases. These improvements would probably not lead to large improvements in average prediction accuracy but could be significant for particular combinations of conditions. This was tested by varying the combination of room acoustics measure and S/N measure in both useful-to-detrimental ratios and in the speech transmission index values. The number and type of variations considered were quite small and only intended to indicate where more extensive studies could be useful.

To create new variations of useful-to-detrimental ratios a numerical constant ( $K$ ) was introduced into Eq. (3) as a multiplier of the linear N/S ratio factor. This factor was varied from 1 to 2 in steps of 0.1.  $U_{50}$  values were calculated using  $C_{50}(500-4k)$  and S/N(A) values as had been found successful in this work. For each new set of  $U_{50}$  values a third order polynomial fit was calculated with speech intelligibility scores. The corresponding  $R^2$  values went through a broad maximum with the highest  $R^2$  occurring for  $K=1.5$ . The  $R^2$  value increased to 0.964 from a value of 0.961 for  $K=1$  [i.e.,  $U_{50}[C_{50}(500-4k)]$  in Table III]. Although this is a very modest increase and not statistically significant, when the predictions were plotted similar to Fig. 5, there were small improvements in that the effect of  $C_{50}$  on speech intelligibility was slightly reduced relative to the effects of S/N.

The results in Fig. 6 suggest that improvements may also be possible for the speech transmission index measure. To test this idea STI values were calculated with a numerical factor multiplying the S/N component in each octave band. Increasing the importance of the S/N component again led to increased  $R^2$  values. When the S/N components were multiplied by 4, the  $R^2$  value associated with predictions of speech intelligibility scores from these modified STI+N values was increased to 0.969 from 0.936 for the unmodified STI+N. This is a significant improvement and Fig. 8 confirms that there is an improved relationship with speech intelligibility scores. These simple tests indicate that further changes to the useful-to-detrimental ratio and speech transmission index measures could lead to small improvements in predictions of expected speech intelligibility. These suggestions should be validated for a wide range of acoustical conditions in real rooms.

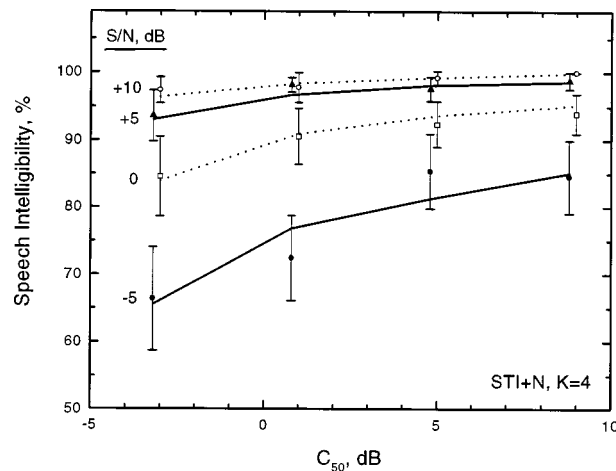


FIG. 8. Comparison of measured and predicted speech intelligibility scores based on a modified STI+N measure. (The S/N=+5 dB and -5 dB cases have been offset to avoid overlapping error bars.)

## VI. CONCLUSIONS

The following are the principal conclusions of this work: While room acoustics characteristics and signal-to-noise ratios both influence speech intelligibility in rooms, the effect of signal-to-noise ratio is much more important. This usually implies that it is most important to reduce ambient noise levels so that acceptable speech-to-noise ratios are achieved (i.e.,  $\geq +15$  dB). It also supports suggestions<sup>14</sup> that it will not normally be important to be overly concerned about achieving precisely the “correct” reverberation time and especially so if an adequate S/N is not first achieved. Broadband useful-to-detrimental ratios have been proposed and two practically useful forms were very accurate predictors of speech intelligibility scores. These are: (a) an A-weighted sum of octave band useful-to-detrimental ratios [ $U_{50a}(A)$ ] and (b) a measure calculated from measured  $C_{50}(500-4k)$  values (i.e., averaged over the 4 octaves from 500 to 4 kHz) and S/N(A) values using Eq. (3) and referred to as  $U_{50}[C_{50}(500-4k)]$ . The useful-to-detrimental ratios,  $AL_{cons}$  and STI+N values were all found to be reasonable predictors of the combined effects of room acoustics and S/N effects on speech intelligibility for the 16 simulated sound fields. However, combinations of EDT and AI were only a little inferior as predictors of speech intelligibility in the present study.

Although the combined measures were successful predictors of speech intelligibility, it was demonstrated that further improvements in the form of these measures are possible. In particular, it was possible to improve the prediction accuracy of the STI+N measure by increasing the importance of the S/N component of STI+N.

The simulated sound fields included the complete range of both room acoustics ( $C_{50}$ ) and S/N effects likely to occur in most rooms for speech. Thus the current results should be valid over a very wide and representative range of conditions. However, these results and the suggested improvements to the form of the combined measures should now be evaluated over a wide range of conditions in real rooms.

Such further investigations could include the investigation of various second order effects such as the overall density and spacing of early reflections on speech in rooms.

- <sup>1</sup>ANSI (1969). ANSI Standard S3.5-1969, "Methods for the Calculation of the Articulation Index" (American National Standards Institute, New York, 1969).
- <sup>2</sup>H. Haas, "The influence of a single reflection on the audibility of speech," *J. Aud. Eng. Soc.* **20**, 145–159 (1972).
- <sup>3</sup>R. Thiele, "Richtungsverteilung und Zeitfolge der Schallruckwürfe in Raumen," *Acustica* **3**, 291–302 (1953).
- <sup>4</sup>L. Cremer and H. A. Müller, *Principles and Applications of Room Acoustics*, Section III.2.2, Vol. 1 (Applied Science Publishers, London, 1982).
- <sup>5</sup>J. S. Bradley, R. Reich, and S. G. Norcross, "A just noticeable difference in  $C_{50}$  for speech," *Appl. Acoust.* **58**, 99–108 (1999).
- <sup>6</sup>J. P. A. Lochner and J. F. Burger, "The influence of reflections on auditorium acoustics," *J. Sound Vib.* **1**, 426–454 (1964).
- <sup>7</sup>H. G. Latham, "The signal-to-noise ratio for speech intelligibility—An auditorium acoustics design index," *Appl. Acoust.* **12**, 253–320 (1979).
- <sup>8</sup>J. S. Bradley, "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.* **80**, 837–845 (1986).

- <sup>9</sup>J. S. Bradley, "Speech intelligibility studies in classrooms," *J. Acoust. Soc. Am.* **80**, 846–854 (1986).
- <sup>10</sup>V. M. A. Peutz, "Articulation loss of consonants as a criterion for speech transmission in room," *J. Aud. Eng. Soc.* **19**, 915–919 (1971).
- <sup>11</sup>J. S. Bradley, "Relationships among measures of speech intelligibility in rooms," *J. Aud. Eng. Soc.* **46**, 396–405 (1998).
- <sup>12</sup>T. Houtgast and J. M. Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," *Acustica* **28**, 66–73 (1973).
- <sup>13</sup>T. Houtgast and 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).
- <sup>14</sup>R. Reich and J. S. Bradley, "Optimizing classroom acoustics using computer model studies," *Can. Acoust.* **26**, 15–21 (1998).
- <sup>15</sup>T. Houtgast, "The effect of ambient noise on speech intelligibility in classrooms," *Appl. Acoust.* **14**, 15–25 (1981).
- <sup>16</sup>J. S. Bradley, "Optimising the decay range in room acoustics measurements using maximum-length-sequence techniques," *J. Aud. Eng. Soc.* **44**, 266–273 (1996).
- <sup>17</sup>J. S. Bradley, "An international comparison of room acoustics measurements systems," IRC Internal Report No. 714 (January 1996).

# Time-frequency filters for target classification

P. Chevret,<sup>a)</sup> N. Gache, and V. Zimpfer

CPE Lyon, LISA(EP92 CNRS)/LASSSO, Laboratoire d'Acoustique, Systèmes, Signaux et Sonar,  
43 Bd du 11 nov. 1918, BP2077, 69616 Villeurbanne cedex, France

(Received 1 May 1998; revised 10 June 1999; accepted 7 July 1999)

This paper presents a new sonar target classification approach based on the use of time-frequency filters. Their design is carried out from the free field response of a reference target, and more precisely from the analysis of echo formation mechanisms in the time-frequency plane. The study of the relevance and the robustness of this approach in approximately real sonar conditions is conducted from experimental measurements in a tank. A data base is set up that contains a large set of target responses in the free field, near different interfaces and in waveguide situations. First, the efficiency of the method for the recognition of a nickel molybdenum spherical shell, corresponding to a class of man made targets whose size is much smaller than the sonar beam (finite size) is shown (100% of recognition). Second, a classification procedure between different targets of finite size is conducted: more than 85% of good classification is obtained (except for the marble solid target). Finally, in the presence of numerical noise, the method is found to be robust even for a low signal to noise ratio. © 1999 Acoustical Society of America. [S0001-4966(99)04510-5]

PACS numbers: 43.60.Gk, 43.60.Pt, 43.30.Gv [JCB]

## INTRODUCTION

The problem of discrimination of immersed targets was initiated in the late seventies with the works of Chestnut,<sup>1</sup> who tested multiple frequency based approaches, and J. F. Hoffman<sup>2</sup> who investigated time-domain techniques. Thanks to studies on marine mammals and especially on the sonar system of dolphins,<sup>3</sup> progress has been made on the understanding of their capability to discriminate objects and consequently new time-domain techniques, based on neural network inversions, have been developed.<sup>4,5</sup> Time-frequency approaches have also been used for target classification<sup>6,7,8</sup> and have shown high potentiality for the discrimination between solid and hollow targets<sup>9</sup> as well as for determining the target material.<sup>9,10</sup> Although these methods can be very good, they were only conducted in free field contexts and only a few relevant results were published concerning the influence of interfaces.

However, some measurements showed the deterioration of the sensitivity of the porpoises sonar when targets approach a reverberating bottom at short ranges and when they are moved away from the animal.<sup>11</sup> With respect to human algorithms, neural network based approaches show high potentialities in short range contexts (of the order of the target length) with a sandy ocean floor.<sup>12</sup> The main drawback of such learning procedures is that they necessitate a large amount of data to feed the network and the data need to be recorded, as far as possible, under similar conditions.

For long range propagation, where multiple reflexions occur, numerical simulations showed a large decrease of the performances in simple waveguide configurations. This has been shown in a previous paper<sup>13</sup> where the recognition algorithm was based on the combination of a colored bright spot model (extraction of pertinent parameters) and a  $k$  near-neighbor procedure for the classification.

The aim of this paper is to develop a new classification technique that should be robust to the bottom structure (variations of the mechanical properties and roughness) even for low signal-to-reverberation levels and should operate in a long range context. The technique is based on time-frequency filters that are derived from the free field response of the targets to be recognized (called reference targets). In fact, these filters behave like a time-frequency pass region that permits the extraction of the relevant information (surface wave patterns in this case) from the input signals.<sup>14</sup> Consequently, the method needs no *a priori* knowledge of the bottom structure nor the background noise. It only requires the free field response of the target to be recognized.

In Sec. I, the principle of the time-frequency filtering method, developed by Hlawatsch,<sup>15</sup> is briefly presented.

In Sec. II, we describe the recognition procedure based on the time-frequency filtering method. For this purpose, a tank experiment is conducted in order to model low frequency ( $\sim 500$  Hz) and long distant object situations ( $\sim 1$  km). A scale factor of about 1000 needs to be applied. The data base obtained from the measurements contains about one hundred acoustic echoes issued from targets of different shape, different material (solid or hollow) and located near various interfaces.

Section III gives two examples of target classification carried out from the experimental data base. In the first case, the procedure consists of the recognition of a nickel molybdenum spherical shell whose size is much smaller than the sonar beam (subsequently called finite size). The second case concerns the discrimination between different targets of finite size.

Finally, the last section is devoted to the study of the robustness of the method to noise. The previous experimental data base, as well as the reference signal, are adversely affected by simulated noise. The robustness is tested for various signal to noise ratios.

<sup>a)</sup>Electronic mail: chevret@cpe.fr



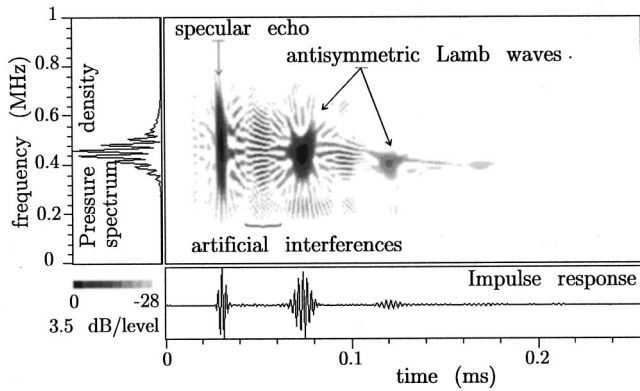


FIG. 1. Wigner-Ville distribution of a nickel molybdenum spherical shell response measured in free field ( $\phi=30$  mm;  $e=0.6$  mm).

## I. THE TIME-FREQUENCY FILTERING METHOD

The time-frequency filtering method has recently been proposed by Hlawatsch.<sup>15</sup> It consists of the construction of a time-frequency pass region that is used to extract the relevant information on signals. The method was successfully applied to noise filtering and component separation on transient theoretical signals.<sup>15,16</sup> The aim of this paper is to apply the time-frequency filtering approach to the problem of underwater target recognition and classification.

Previous studies<sup>6,7,8</sup> have shown that a time-frequency description, and typically the Wigner-Ville distribution, is relevant for the understanding of echo formation mechanisms and for the identification of surface waves that circumnavigate the targets. A filtering approach based on a time-frequency description should consequently be a relevant signal processing tool for the separation of echo components (from noise or reverberation) and target identification.

### A. Time-frequency distribution and echo formation mechanisms

The Wigner-Ville distribution WV of a signal  $s(t)$  is calculated from the usual expression:<sup>17,18</sup>

$$\text{WV}_s(t, f) = \int_{-\infty}^{+\infty} s\left(t + \frac{\tau}{2}\right) s^*\left(t - \frac{\tau}{2}\right) \exp(-2i\pi f\tau) d\tau. \quad (1)$$

As an example, the WV distribution associated with the experimental acoustic response of a nickel molybdenum spherical shell (diameter  $\phi=30$  mm; thickness  $e=0.6$  mm) in the free field is presented in Fig. 1. The experimental conditions are described in Sec. II A. The different patterns that appear on this image can be identified and separated into two categories.<sup>6,19,20</sup>

- (1) High energy patterns. The first one, non dispersive, is associated with the specular echo on the target and the two following patterns correspond to the arrival of surface waves (coinciding antisymmetric Lamb waves  $a_0^-$  and  $a_0^+$ ) that circumnavigate the target.
- (2) Artificial interferences due to the bilinear nature of the WV distribution.<sup>21</sup> They do not correspond to any physi-

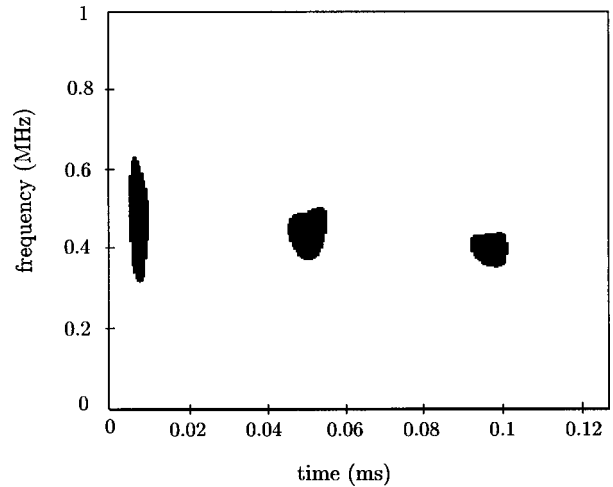


FIG. 2. Time-frequency support of the Wigner-Ville distribution from a nickel molybdenum spherical shell response measured in free field (black = 1; white = 0).

cal propagation phenomena but can be smoothed independently in time and frequency with the help of appropriate smoothing windows.

### B. Filter design

The function of a time-frequency filter is to extract from the signal to be analyzed the pertinent information associated with the circumferential waves previously described. The filter is designed from the WV distribution of a reference signal, and more particularly from its time-frequency support  $R$  containing the relevant information. This region can be derived either manually or with the help of image processing tools. Mathematically speaking, it can be described by an indicator function  $IR(t, f)$ :

$$IR(t, f) = \begin{cases} 1 & \text{if } (t, f) \in R \\ 0 & \text{if } (t, f) \notin R \end{cases}$$

In the shell scattering problem, the relevant information is composed of the specular reflection and the Lamb wave patterns. The indicator function associated with the nickel molybdenum spherical shell is presented in Fig. 2. In this figure, the specular and the surface wave patterns have been clearly isolated.

Although this function is not a valid Wigner distribution, it can still be written as a linear combination of the WV distribution of orthonormal signals  $u_k$ :<sup>15</sup>

$$IR(t, f) = \sum_{k=1}^{\infty} \lambda_k \text{WV}_{u_k}(t, f). \quad (2)$$

The  $u_k$  components define an orthonormal basis of a linear signal space, called  $S_R$ , of infinite dimension. Its WV distribution is localized in the time-frequency region  $R$  and defined as the sum of the WV of each  $u_k$ :

$$\text{WV}_{S_R} = \sum_{k=1}^{\infty} \text{WV}_{u_k}(t, f).$$

In Eq. (2),  $u_k$  and  $\lambda_k$  are, respectively, the eigensignals and eigenvalues of the following eigenequation:

$$\int_{-\infty}^{+\infty} H_R(t, t') u_k(t') dt' = \lambda_k u_k(t)$$

where the Hermitian kernel  $H_R$  is obtained from the region  $R$  ( $IR$  is the Weyl symbol of  $H_R$ ) using the expression<sup>15</sup>

$$H_R(t, t') = \int_{-\infty}^{+\infty} IR\left(\frac{t+t'}{2}, f\right) \exp(2i\pi(t-t')f) df.$$

$u_k$  and  $\lambda_k$  are completely determined by the region  $R$ . They are called eigensignals and eigenvalues of the time-frequency region  $R$ .

As the set of eigensignals  $u_k$  defines an orthonormal basis of infinite dimension, a reduced set of  $u_{k(k=1, N)}$  defines a finite eigensubspace  $\mathcal{U}_R^{(N)}$  (of dimension  $N$ ) of  $S_R$ . The orthogonal projection operator  $h$ , written as

$$h(t, t') = \sum_{k=1}^N u_k(t) u_k^*(t'),$$

represents the kernel of the linear application (endomorphism)  $g$  of  $\mathcal{U}_R^{(N)}$ :

$$g: x(t) \rightarrow \int_{-\infty}^{+\infty} h(t, t') x(t') dt'. \quad (3)$$

This orthogonal projection of a signal  $x(t)$  can be interpreted as the output of a linear, time-varying system, i.e., a time-varying filter. Now, the following question arises: what is the dimension of  $\mathcal{U}_R^{(N)}$  that minimizes the differences between the WV distributions of  $S_R$  and  $\mathcal{U}_R^{(N)}$ ? The estimation of  $N$  is carried out by maximizing a concentration function and minimizing a localization error of  $\mathcal{U}_R^{(N)}$  in  $R$ . The concentration  $\rho$  is defined as the part of the energy of  $\mathcal{U}_R^{(N)}$  contained in the region  $R$ . In other words, it is the ratio of the energy contained in  $R$  to the total energy of the eigensubspace.

$$\rho(\mathcal{U}_R^{(N)}, R) = \frac{\iint_{(t,f) \in R} \text{WV}_{\mathcal{U}_R^{(N)}}(t, f) df dt}{\iint_{(t,f)} \text{WV}_{\mathcal{U}_R^{(N)}}(t, f) df dt}.$$

It can be shown that maximizing the concentration is achieved by selecting the  $N$  highest eigenvalues for the construction of  $\mathcal{U}_R^{(N)}$ . The localization error defines the difference between the localization in the time-frequency plane of the region  $R$  and the WV distribution of  $\mathcal{U}_R^{(N)}$

$$\begin{aligned} \epsilon^2(\mathcal{U}_R^{(N)}, S_R) &= \int \int_{(t,f) \in R} [1 - \text{WV}_{\mathcal{U}_R^{(N)}}(t, f)]^2 df dt \\ &+ \int \int_{(t,f) \notin R} [0 - \text{WV}_{\mathcal{U}_R^{(N)}}(t, f)]^2 df dt. \end{aligned}$$

In order to minimize this error, the selected eigenvalues must be greater than 1/2. Consequently, the optimal set of orthonormal eigensignals (the optimal dimension of  $\mathcal{U}_R^{(N)}$ ) is provided by the  $N$  eigenvalues greater than 1/2.

Figure 3 shows the distribution of the eigensubspace  $\mathcal{U}_R^{(N)}$  in the case of a nickel molybdenum spherical shell. The optimal dimension of the subspace is 5. The localization error and the concentration are, respectively, 1% and 90%.

In summary, the filter design consists of four successive steps:

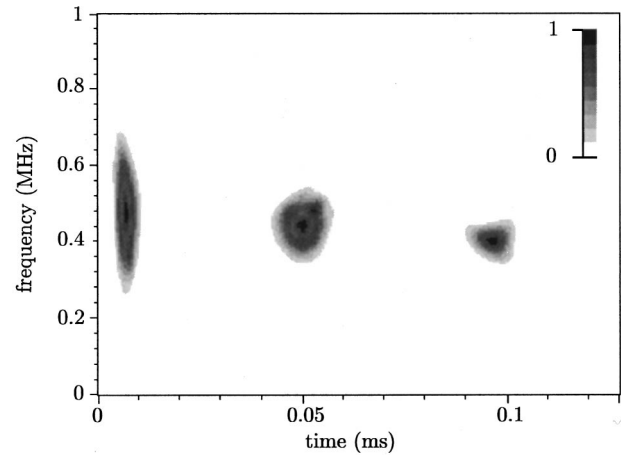


FIG. 3. Time-frequency representation of the finite dimension eigensubspace  $\mathcal{U}_R^{(N)}$  for a nickel molybdenum spherical shell response measured in free field.

- compute the WV distribution of the reference signal,
- build the indicator function  $IR(t, f)$ ,
- compute the kernel  $H_R(t, t')$  and perform its eigenanalysis,
- keep only the  $N$  eigenvalues greater than 1/2,
- construct the impulse response  $h(t, t')$  of the filter.

## II. THE RECOGNITION PROCEDURE

The recognition of targets in the free field is performed in the following way. First, the filter is designed from the response of a reference target (to be recognized). Second, this filter is applied to the signal to be analyzed and the output response is simply compared to the reference signal.

In the case of a complex multipath environment like the ocean, target impulse responses are much longer than the free field response and are adversely affected by many mechanisms such as volume (local inhomogeneities, fish shoals, etc.) and surface reverberation due to surface roughness. In this context, a recognition procedure should be able, in addition to the recognition operation itself, to separate targets and spurious components.









Such a recognition procedure is presented in this section. For this procedure, a large set of echo targets is recorded from a series of experimental measurements conducted in a tank.

From a selection of responses among the complete experimental data base, a study of echo formation mechanisms is first presented that points out the necessity of setting up a specific procedure to filter bottom reverberation echoes. The implementation of the filtering procedure is then described.

### A. Tank experiments

A scaled experiment in a tank has been set up that provides a large and eclectic set of acoustic responses of immersed targets located near very different kinds of bottoms. The tank dimensions are 2-m long, 1-m wide, and 1-m high. We considered four targets which are different in their geometry and their mechanical properties (Table I). The first one is a spherical hollow target (shell) made of nickel mo-

TABLE I. Target shape and mechanical properties ( $e$ : thickness,  $\phi$ : diameter,  $l$ : length).

	Nickel Molybdenum spherical shell ( $\phi = 30$ mm, $e = 0.6$ mm)
	Aluminium cylindrical shell ( $\phi = 40$ mm, $e = 2$ mm) Considered as infinite compared to the transducer beam. The inner fluid is air.
	Same as  , except the inner fluid is water.
	Aluminium cylindrical solid target with hemispherical endcaps ( $\phi = 30$ mm, $l = 75$ mm). The major axis is parallel to the transducer beam.
	Same as  , except the attitude in the transducer beam: the major axis is perpendicular to the transducer beam.
	Solid spherical target of marble ( $\phi = 30$ mm)

lybdenum. Its diameter and thickness are respectively 30 mm and 0.6 mm. The second target is a long tube of aluminum. It is considered as an object of infinite size in the sense that its length is much greater than the sonar beam. Its diameter and thickness are, respectively, 40 mm and 2 mm. In a first configuration, the tube is air-filled, in a second one it is full of water. The third target is a solid cylinder with hemispherical endcaps. It is made of aluminum and its diameter and length are, respectively, 30 mm and 75 mm. Firstly, the major axis of the cylinder is parallel to the transducer beam, secondly it is perpendicular to the beam. The last target is a solid sphere made of marble. Its diameter is 30 mm. Three series of experiments have been carried out (Fig. 4):

- (1) In the free field, where the targets are located far from any boundaries [Fig. 4(a)]. This gives 6 signals, one for each target.
- (2) Near a single interface (semi-infinite field) with a large variety of interfaces ranging from smooth to rough surfaces: the water-air interface, a layer of sand (thickness=20 cm), a gravel layer (thickness=2 cm; mean diameter of the grain:  $5\text{ mm} \pm 0.1\text{ mm}$ ) and a marble block (thickness=10 cm). The separation distance between the source/receiver and the target is about 60 cm. First, targets are laid down on the bottom then they are moved from the interface to the center of the water column by steps of 1 cm (three positions). This second series of measurements gives a set of 96 responses [Fig. 4(b)].
- (3) In a waveguide (shallow water). In this case, the separation distance between the transmitter and the target (the molybdenum spherical shell) is about 1.2 m. Three waveguide configurations are considered. In the first one, the height of the water column is 30 cm and the bottom is made up of sand. In the second and third trials, the height of the water column is 15 cm. In the last trial, a thin layer of gravel (2 cm thick and 50 cm long) is laid down equidistant from the source and the target [Fig. 4(c)]. One signal per configuration is recorded.

The whole data base consequently contains a total of 105 signals. For all the trials, the transmitter is a broadband

transducer. The beamwidth at 500 kHz is  $6^\circ(-3\text{ dB})$ . The transmitted signal is a pulse of about  $6\ \mu\text{s}$ . The frequency band is centered around 500 kHz (350 kHz–650 kHz) ( $-3\text{ dB}$ ). In order to eliminate electronic noise, each recorded signal is the average of 50 returns.

## B. Echo formation mechanisms

A qualitative analysis of echo formation mechanisms can be conducted in the very few cases where the echo com-

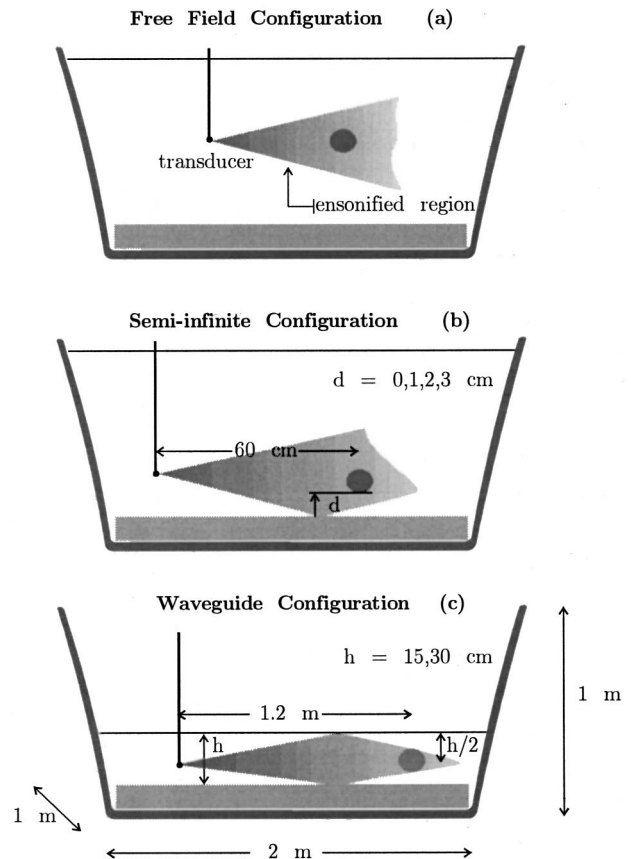


FIG. 4. Experiment configurations.

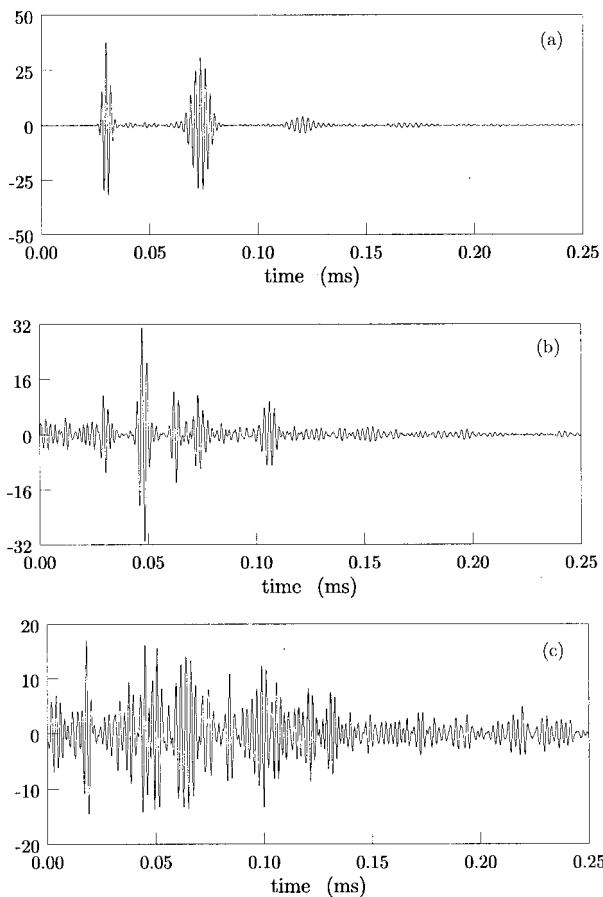


FIG. 5. Target acoustic responses showing the influence of the presence of interfaces (a) nickel molybdenum spherical shell in free field (b) same shell 2 cm above a sandy bottom (c) same shell 3 cm above a gravel layer.

ponents are clearly separated and can be easily identified. This is illustrated in Fig. 5 which depicts two amongst the more representative responses recorded.

The first one [Fig. 5(b)] corresponds to the response of the nickel molybdenum spherical shell near a sandy bottom and the second one to the same target in the vicinity of a gravel layer [Fig. 5(c)]. For a direct comparison, the free field response of the shell is also depicted [Fig. 5(a)]. These figures clearly show the influence of the bottom. In the first case, the attenuating properties of the sandy bottom are responsible for a separation of the echo components that makes the response easily interpretable. The detailed analysis of this particular echo structure is presented in Fig. 6. In this signal, each acoustic path can be easily identified by simple travel time calculations. We can distinguish:

- a high amplitude component labeled (2). This corresponds to the constructive superposition of 2 paths of same length but of opposite direction (ABCA and ACBA). Both are associated with a single reflection on the bottom;
- the specular echo corresponds to a direct monostatic reflection on the target (1). Its amplitude is about half as high as the previous component;
- the Lamb wave components that circumnavigate the shell are generated by the incident wave, impinging on the target. They are designated by the number (3);

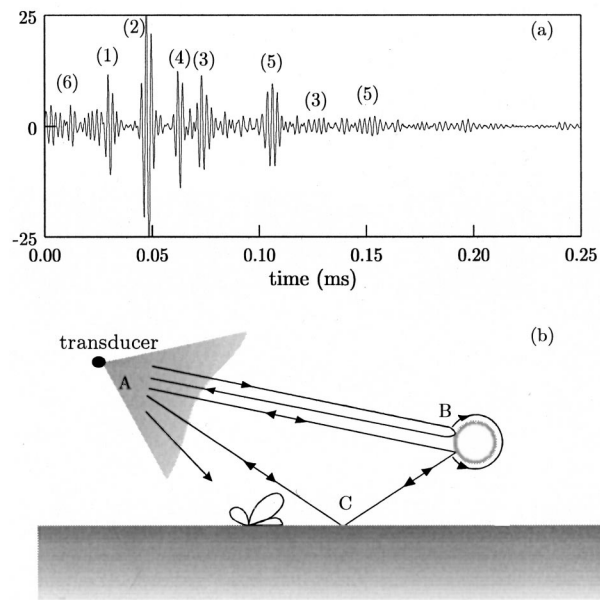


FIG. 6. Echo formation analysis from the experimental response of a spherical shell located 2 cm above a sandy bottom.

- the path ACBCA that corresponds to a second specular reflection on the target generated after a reflection on the bottom. The associated path is labeled (4);
- Lamb wave components (5) produced by the previous specular reflection;
- small amplitude components (6) preceding the direct specular reflection (1). These components correspond to bottom reverberation;

In such a basic case, for which a component identification is possible, the filtering process can be directly applied to the portion of signal that contains the direct specular reflection. The target recognition procedure then reduces to a comparison between the output of the filter and the reference signal, by means of a quadratic error for instance. On the other hand, for the majority of the recorded responses, the signal presents a very complex structure due to the superimposition of target echoes and high level reverberation components [Fig. 5(c)]. In this case, a more complex recognition procedure has to be set up in order to separate reverberation from target echoes.

### C. Description of the recognition procedure

In a first step, the time-frequency filter is synthesized, as explained in Sec. IB, from the impulse response of a reference target to be recognized. Then, the recognition procedure is developed according to the two following operations.

- (1) *The recognition operation itself that is applied as a sliding window along the input signal.* For each location of the filtering window, a quadratic error between the normalized envelopes of the reference response and the filter output is calculated. The minimum of the error along the entire input signal corresponds to the window location at which the successive echoes contained in the reference signal may have been identified. If this minimum is



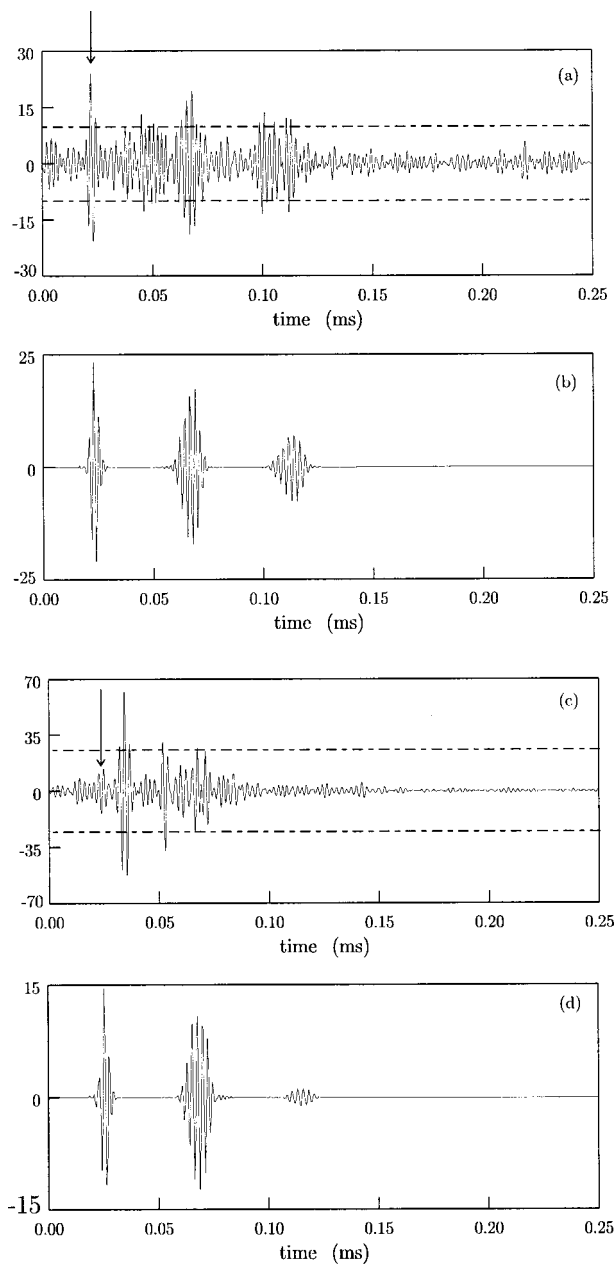


FIG. 7. Input and output of the reference filter for (a) the nickel molybdenum spherical shell located at 1 cm below the free surface (c) the cylindrical solid target with hemispherical endcaps located at 2 cm above the block of marble. (b) and (d) are their respective filter output. The filter is set up from the nickel molybdenum spherical shell response. The arrows show the component that was identified as the specular echo.

small enough (compared to an arbitrary threshold  $\mathcal{E}$ ) then the reference target is clearly identified in the input signal.

Figure 7 is the result of this filtering process applied to different impulse responses near a single boundary. The reference target is the nickel molybdenum spherical shell. The first signal [Fig. 7(a)] is the response of the shell above a gravel layer and the second one [Fig. 7(c)] is the response of a cylindrical solid target with hemispherical endcaps above a marble block (Table I). The arrows on these figures give the time the specular echo was identified with a very small error in both cases (respectively, 6.3% and 3% of the reference signal energy).

The signals in Fig. 7(b) and (d) are the corresponding filter output. This result is not convincing for the second target because it is identified as the nickel molybdenum shell. In fact, in this case, the identified specular echo is not a reflexion from the target but a reverberation pattern with low energy content (maximum amplitude of 15 against 70 for the input signal). This pattern must be rejected.

- (2) *The elimination of reverberation echoes.* To overcome this possible confusion, a signal to reverberation ratio SRR threshold is defined that separates the reverberation components and the target echoes. This SRR threshold is given by the ratio of the maximum amplitudes of the output and input signals. The choice of its value is arbitrary and is linked to the problem of target echo detection which is not the topic of this paper. In our case, in order to minimize confusions between target and reverberation echoes, an “optimal value” for this threshold was found to be 8 dB. In Fig. 7, the SRR threshold is shown by the horizontal dashed lines. The components lying between these lines are automatically considered as reverberation patterns.

A schematic description of the entire recognition procedure is given in Fig. 8.

### III. APPLICATIONS OF THE RECOGNITION PROCEDURE

In this section, we present two applications of the recognition procedure previously described. The first one concerns the recognition (among a large variety of different objects) of the nickel molybdenum spherical shell (corresponding to a man made target of finite size). The second application concerns the classification between different targets of finite size.

#### A. Recognition of a spherical shell

The recognition procedure presented in Sec. II C and summarized in Fig. 8 is now applied to each signal of the experimental data base. The targets are being classified in two categories: the nickel molybdenum shell and the set of other targets. The reference target used for the filter design is the nickel molybdenum spherical shell (representative of shells of finite size).

For each signal of the data base, the filter is applied as a sliding window. The error between the envelop of the reference response and that of the filter output is calculated for each position of the window and the minimum value is compared to the arbitrary threshold  $\mathcal{E}$ . The value of this threshold is chosen so that a maximum rate of good recognition of the spherical shell is obtained. In our case,  $\mathcal{E}$  will be kept to the value of 6.5%. For this optimization, the other targets may not be well classified. Finally, whether the error is smaller or greater than  $\mathcal{E}$ , the target is identified, respectively, as a nickel molybdenum shell or as another target.

Table II gives the results of classification that are very good in all cases: 100% recognition of the spherical shell and more than 88% of the other targets are correctly classified whatever the bottom structure. When averaging over the en-

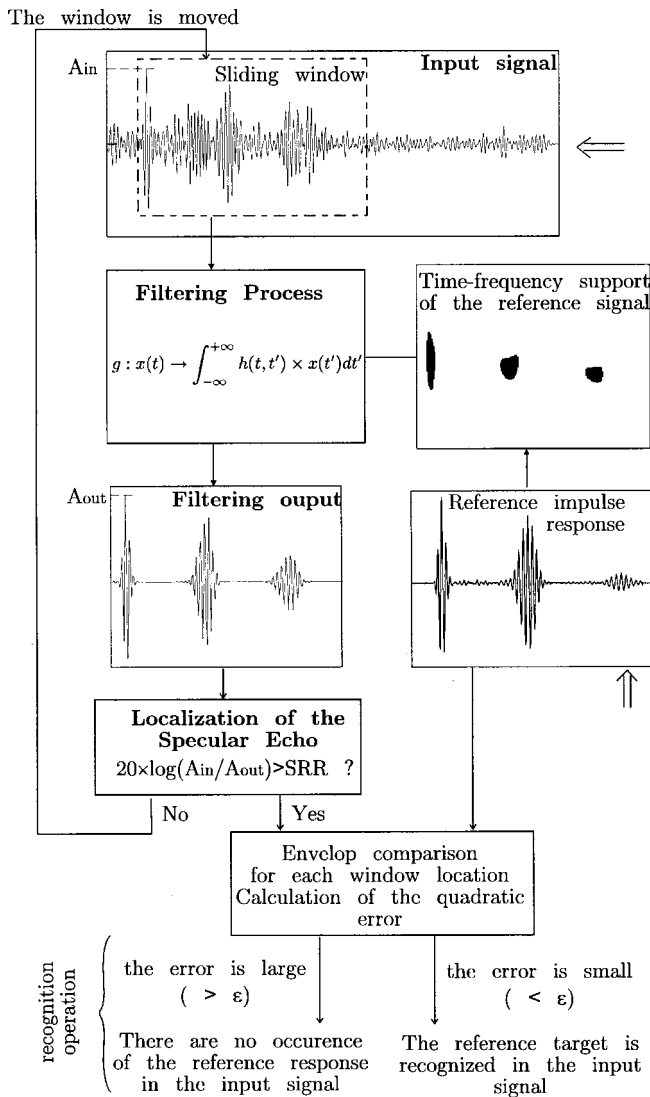


FIG. 8. Schematic representation for the recognition procedure.

tire set of other targets weighted by the number of signals for each target, the rate of good recognition reaches 96%.

Another interesting result, as shown in Table III, is the influence on the classification of the interface structure (roughness and mechanical properties). Very good classification rates are obtained for each configuration (90% or more) showing that the method is highly robust to variations of the roughness and mechanical properties of the bottom.

Finally, regarding the waveguide context, the method is very efficient (100% recognition of the spherical shell). However, the analysis has to be extended since the trans-

TABLE II. Good classification rates (SRR threshold=8 dB and  $\mathcal{E}=6.5\%$ ). Two classes are considered (1) shells of finite dimensions and (2) the other targets. The filter is processed from the free field response of the nickel molybdenum spherical shell.

Spherical Shell	Other targets				
100%	100%	95%	95%	100%	88%

TABLE III. Good classification rates according to the bottom structure (SRR threshold=8 dB and  $\mathcal{E}=6.5\%$ ). The filter is processed from the free field response of the nickel molybdenum spherical shell.

Free field	Sand	Marble	Gravel	Free surface	Waveguide
100%	100%	90%	96%	100%	100%

ducer beam is narrow, so may not produce a sufficient insonification of the boundaries.

## B. Classification of targets of finite size

In this section the recognition method is extended to the discrimination between different targets of finite size. The responses corresponding to the infinite cylinders are removed from the data base. Three classes are defined, each class corresponding to a particular target: the nickel molybdenum spherical shell, the cylinder with hemispherical endcaps (independently of its attitude in the sonar beam) and the marble spherical solid target. Consequently, the data base to be processed now contains 54 signals:

- 3 responses in the free field (one per target),
- 48 responses near interfaces (3 targets  $\times$  4 immersions  $\times$  4 interfaces),
- 3 responses in the waveguide.

With respect to the classification procedure, the following steps are conducted.

- (1) The filter associated with each target to be recognized is constructed from its free field impulse response. For the cylinder with hemispherical endcaps, two filters are processed, one for each attitude in the sonar beam.
- (2) The set of 54 signals is projected on each filter.
- (3) The class to which a signal belongs is determined by the minimum projection error on the set of filters.

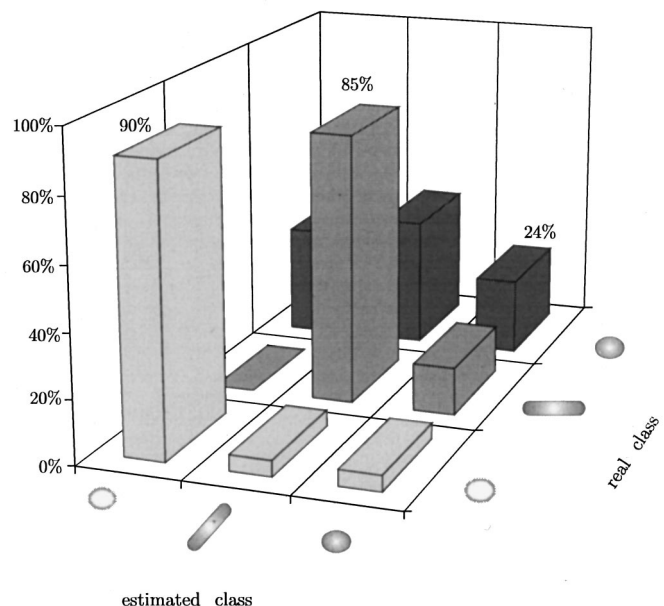


FIG. 9. Classification rates between targets of finite dimensions (SRR threshold=8 dB).

TABLE IV. Classification rates for the spherical shell and the other targets in presence of noise (SRR threshold=8 dB;  $\mathcal{E}$ =6.5%). Two configurations: (1) in free field and (2) near a single interface. No noise added on the reference signal. 50 noise realizations on the data bases.

SNR	Free field noisy data base		Semi-infinite noisy data base	
	Spherical shell	Other targets	Spherical shell	Other targets
20 dB	100%	100%	95%	93%
15 dB	100%	100%		
10 dB	100%	100%	77%	88%
5 dB	86%	96%		

The result of this procedure is presented in Fig. 9. The recognition rates for the spherical shell and the cylindrical solid target with hemispherical endcaps are very high (respectively, 90% and 85%). On the other hand, the marble target is badly recognized (24%). The reason for this low recognition rate cannot be clearly established. Nevertheless, it indicates in this case, that the surface waves do not present any visible specificity in the time-frequency plane (for example, a dispersive behavior) so as to achieve a clear identification of the target in all situations.

#### IV. ROBUSTNESS TO NOISE

In order to approach real ocean conditions, robustness to noise of the recognition procedure is studied in the same manner as in Sec. III A. For this, Gaussian white noise is added to each signal of the experimental data base which has been split into two, depending on whether the target is in the free field or near a single interface.

- *Free field target responses.* For each recorded signal, 50 realizations of noise are added, giving a total of  $6 \times 50 = 300$  signals. In an oceanic context, this set of data may be representative of target responses in a deep channel without boundary reflections.
- *Responses near a single interface.* 20 realizations of noise are added to each response, so that the data base contains  $96 \times 20 = 1920$  signals. These signals have more similarities with target responses in shallow water where the presence of interfaces (water-air interface and bottom) is of prime importance.

##### A. Noisy data base

For each data base, the recognition procedure is performed independently. The results are reported in Table IV. For the free field data base, the recognition rates are of 100% for a SNR up to 10 dB. A deterioration of the performances is only observed for very small SNR (5 dB). In this case, 86% of the shell responses are recognized. In the presence of interfaces, high recognition rates are still obtained (greater than 95%). Consequently, noise influence is very weak on the time-frequency classification procedure when the targets are far from any boundaries. Its effect is more important near interfaces because noise and reverberation components may add in a constructive manner to produce, on the resulting signal, echoes of high amplitude (higher than the specular echo).

TABLE V. Classification rates for the spherical shell and the other targets in presence of noise (SRR threshold=8 dB,  $\mathcal{E}$ =6.5%; SNR on the data base=10 dB). Two configurations: (1) in free field and (2) near a single interface. 10 noise realizations added on the reference signal and 50 on the data bases.

SNR On the reference	Free field noisy data base		Semi-infinite noisy data base	
	Spherical shell	Other targets	Spherical shell	Other targets
25 dB	100%	94%	88%	82%
20 dB	100%	94%	90%	79%
17 dB	100%	88%	93%	68%
15 dB	100%	78%	94%	44%

##### B. Noisy reference and data base

In an oceanic context, this situation is encountered when the reference target response is to be measured *in situ* where noise is systematically present on the signal.

In this study, the presence of noise is taken into account by using 10 realizations of Gaussian white noise added independently to the reference target response. The 10 associated time-frequency supports are then calculated and the final reference signal is obtained from their intersection in the time-frequency plane. The number of 10 realizations is chosen from the following considerations: too small a value may not be sufficient to eliminate noise patterns on the final time-frequency support and a too large one will be responsible for a disappearance of the relevant echo target patterns. The consequences are: in the first case, a decrease in the reference target recognition rates, and in the second one on the contrary, a misclassification of the other targets.

The classification results as a function of the SNR on the reference are presented in Table V. Noise (SNR=10 dB) is also added to the experimental data bases (free field and semi-infinite configuration). As previously, the classification rates are higher in the free field case (at least 78%) than near interfaces. Nevertheless, in that configuration, real deteriorations only appear for a SNR beyond 17 dB.

An increase of the recognition rates (from 88% to 94%) is to be noted when decreasing the signal to noise ratio. Although this result may look surprising, it can be explained by the way the reference filter was set up: when the noise level is increased, the time-frequency support contains, in addition to surface wave regions, patterns associated with the added noise. This leads to an increase in the number of output signals that are similar to the reference response and consequently an increase of good classification rates for the spherical shell. In addition, a large decrease in the classification rate of the other targets occurs (from 82% to 44% in the semi-infinite case).

In conclusion, while the measurement of the reference response should be made in a controlled environment (as little noise as possible), the time-frequency filtering method should also be successful in real ocean conditions.

#### V. CONCLUSION

The classification method we have presented in this paper is a new approach based on the use of time-frequency filters. The synthesized filters can be viewed as time-

frequency pass regions that are constructed from the response of a reference target measured in the free field. There are several advantages of using such a technique for target classification: first, it only requires the response of a known target (or class of targets) in the free field and can be applied whatever the bottom structure and variations. Second, as it is based on the appearance and behavior of surface waves around targets, it is more aspect-angle independent than a single time or frequency based approach. For the same reason, it is more robust to reverberation and background noise perturbations.

The relevance of time-frequency filters for sonar target recognition and classification is clearly shown in this paper, even for low signal to reverberation levels and distant objects (5 to 50 times the target length). The study is conducted from target responses measured in a tank for three typical configurations: the free field, near many different interfaces and in waveguides.

A high efficiency of the method is obtained with respect to the recognition of a nickel molybdenum spherical shell (representative of man made objects) among a large variety of other targets over bottoms with very different roughness: 100% of the shells and 96% of the other targets are well classified. Then a classification between different targets of finite size is successfully performed (more than 85% of good classification) except for the marble target (24%) whose time-frequency signature does not present as much specificity as metal targets. The study of the robustness to noise conducted in the last section of this paper is also really encouraging: the classification rates obtained from noisy data bases (in the free field and semi-infinite contexts) are systematically higher than 75%. In order to come closer to real conditions of measurements in oceans, the response of the reference target was also adversely affected by noise. Some expected degradations of the performances are observed, but the method is still robust (at least 68% of good classification for a SNR of 10 dB on the data base and 15 dB on the reference).

As the method shows a good robustness to high reverberation levels, variations in the properties of interfaces and long distances, it should be extended to situations closer to real ocean conditions, in particular to the shallow water context. This should be achieved first by improving the tank experimental conditions and second by testing the method with target responses measured under real ocean conditions.

## ACKNOWLEDGMENTS

This work was supported by the French Ministry of Defense DGA/DCE/GESMA.

- <sup>1</sup>P. C. Chesnut and R. W. Floyd, "An aspect-independent sonar target recognition method," *J. Acoust. Soc. Am.* **70**, 727–734 (1981).
- <sup>2</sup>J. F. Hoffman, "Classification of spherical targets using likelihood and quadrature components," *J. Acoust. Soc. Am.* **49**, 23–30 (1971).
- <sup>3</sup>D. A. Helweg, H. L. Roitblat, P. E. Nachtigall, and M. J. Hautus, "Recognition of aspect dependent three dimensional objects by an echolocating bottlenose dolphin," *J. Exp. Psychol.* **22**, 19–31 (1996).
- <sup>4</sup>W. W. L. Au, "Comparison of sonar discrimination: Dolphin and an artificial neural network," *J. Acoust. Soc. Am.* **95**, 2728–2735 (1994).
- <sup>5</sup>H. L. Roitblat, P. W. Moore, P. E. Nachtigall, R. H. Penner, and W. W. L. Au, "Natural echolocation with an artificial neural network," *Int. J. Neural Networks* **1**, 239–248 (1989).
- <sup>6</sup>F. Magand and P. Chevret, "Time-frequency analysis of energy distribution for circumferential waves on cylindrical elastic shells," *Acta Acust. (China)* **82**, 707–716 (1996).
- <sup>7</sup>D. M. Drumheller, D. H. Hughes, B. T. O'Connor, and C. F. Gaumont, "Identification and synthesis of acoustic scattering components via the wavelet transform," *J. Acoust. Soc. Am.* **97**, 3649–3656 (1995).
- <sup>8</sup>L. R. Dragonette, D. M. Drumheller, C. F. Gaumont, D. H. Hughes, B. T. O'Connor, N.-C. Yen, and T. J. Yoder, "The application of two-dimensional signal transform to the analysis and synthesis of structural excitations observed in acoustical scattering," in *Proceedings of the IEEE, Special Issue on Time-Frequency Analysis*, Vol. 84, pp. 1249–1263 (1996).
- <sup>9</sup>F. Magand, "Reconnaissance de cibles par sonar actif large bande. Application à des coques de forme simple et à la classification des espèces de poissons en mer," Ph.D. thesis, INSA Lyon, 1996 No. 96ISAL.
- <sup>10</sup>P. Chevret, F. Magand, and L. Besacier, "Time-frequency analysis of circumferential wave energy distribution for spherical shells. Application to sonar target recognition," *Appl. Signal Processing* **3**, 136–142 (1996).
- <sup>11</sup>A. E. Murchinson, "Maximum detection range and range resolution in echolocating bottlenose porpoises," Ph.D. thesis, University of California, Santa Cruz, 1980.
- <sup>12</sup>R. P. Gorman and T. J. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," *Neural Networks* **1**, 75–89 (1988).
- <sup>13</sup>P. Chevret, "Man made object classification in shallow water: a hybrid approach," in *3rd International Conference on Theoretical and Computational Acoustics*, New York, July 1997.
- <sup>14</sup>N. Gache, P. Chevret, and V. Zimpfer, "Target classification near complex interfaces using time-frequency filters," in *IEEE, ICASSP-98*, Seattle, USA, 4, 2433–2436 (1998).
- <sup>15</sup>F. Hlawatsch and W. Kozek, "Time-frequency projection filters and time-frequency signal expansions," *IEEE Trans. Signal Process.* **42**, 3321–3334 (1994).
- <sup>16</sup>F. Hlawatsch, W. Kozek, and W. Krattenthaler, "Time-frequency subspaces and their application to time varying filtering," in *Proc. IEEE ICASSP-90*, Albuquerque, New Mexico, USA, 1607–1610, 1990.
- <sup>17</sup>P. Flandrin, *Temps-fréquence* (Hermès, Paris, 1993), p. 394.
- <sup>18</sup>L. Cohen, *Time-frequency analysis*, Processing series (Prentice Hall, Signal, 1995), p. 299.
- <sup>19</sup>G. S. Sammelmann, D. H. Trivet, and R. H. Hackman, "The acoustic scattering by a submerged, spherical shell. I: The bifurcation of the dispersion curve for the spherical antisymmetric Lamb wave," *J. Acoust. Soc. Am.* **85**, 114–124 (1989).
- <sup>20</sup>G. S. Sammelmann and R. H. Hackman, "The acoustic scattering by a submerged, spherical shell. II: The high-frequency region and the thickness quiresonance," *J. Acoust. Soc. Am.* **89**, 2096–2103 (1991).
- <sup>21</sup>F. Hlawatsch and P. Flandrin, "The interference structure of the Wigner distribution and related time-frequency representations," in *The Wigner Distribution: Theory and Applications in Signal Processing*, edited by W. Mecklenbräuker and F. Hlawatsch (Elsevier, New York, 1997).



# Matched field inversion of broadband data using the freeze bath method

Lothar Jaschke and N. Ross Chapman

*School of Earth and Ocean Sciences, University of Victoria, Victoria, BC V8W 3P6, Canada*

(Received 7 January 1999; revised 25 June 1999; accepted 9 July 1999)

This paper presents a statistical freeze bath method for geoacoustic inversion that emphasizes the search for a distribution of models that fit the data well. Contrary to simulated annealing optimization, the freeze bath samples the multidimensional model parameter space at constant freeze probability, corresponding to fixed temperatures for each parameter. The sampling process uses a heat bath algorithm as a Boltzmann sampling tool to carry out a global search over the model parameter space. The conventional heat bath algorithm is modified to sample on a fuzzy grid in order to access the entire range of parameter values. The inversion provides a set of good models that indicates how well the model parameters are constrained by the data, and reveals the degree of correlation between parameters. The efficiency of the search process is improved by reparameterizing the original model parameters to a new set based on the eigenvectors of the model covariance matrix. The inversion performance of the freeze bath is demonstrated using simulated data for a simple geoacoustic model. The method is applied to shallow water broadband data obtained during the Haro Strait geoacoustic tomography experiment to estimate a geoacoustic profile at the site. © 1999 Acoustical Society of America. [S0001-4966(99)04910-3]

PACS numbers: 43.60.Pt, 43.30.Xm [DLB]

## INTRODUCTION

In ocean acoustics applications, the inverse problem of determining ocean bottom properties from acoustic field measurements is generally posed as an optimization problem. An example of this approach that has been applied successfully to shallow water data is matched field inversion (MFI). This method makes use of an efficient global search algorithm to explore a multidimensional model parameter space in order to determine the optimum model. Search processes such as simulated annealing (SA),<sup>1,2</sup> genetic algorithms (GA)<sup>3,4</sup> and hybrid methods have been applied to data from various shallow water environments to invert for geoacoustic properties. Although these methods can produce acceptable geoacoustic parameter estimates, it is much more difficult to determine meaningful objective measures of the uncertainty of the results. The inversion of geophysical data is always subject to errors that arise due to experimental errors, modeling errors, or a combination of both. In practice, many models may fit the data well, and the true environmental model is not necessarily the one that yields the optimum fit. Thus, the need to characterize the uncertainty of the estimate is as fundamentally important as the objective of obtaining the estimate itself.

An alternative approach to solving the inverse problem is provided by the statistical formulation of Tarantola<sup>5</sup> in which the solution is given in terms of the *a posteriori* distribution or (normalized) probability density in model space. However, a complete specification of the *a posteriori* probability density is generally not possible in the multidimensional model parameter space, and one must resort to practical methods of sampling the distribution to determine measures such as the moments or the marginal density functions. The calculation of these measures requires an efficient means of drawing samples from the *a posteriori* distribution,

and various techniques such as Monte Carlo sampling, SA and GA have been proposed for applications in geophysics and ocean acoustics.

In this paper we present a statistical freeze bath method for obtaining a representation of the *a posteriori* distribution of models. The goal in our formulation of the problem is to find geoacoustic models that explain measured acoustic field data. The approach is similar to simulated annealing optimization as described by Basu and Frazer,<sup>6</sup> in that we search for sets of models that optimize a cost function that characterizes the misfit between the observed and theoretically predicted data. However, unlike SA which cools an initially high temperature melt to a final lowest energy state, the freeze bath samples the model space at constant temperatures that correspond to a uniform freeze probability for all the parameters. An optimally performing freeze bath creates a set of models that consists of a high percentage of different low energy models that fit the data well. Moreover, the control parameters have intuitive interpretations, and the method itself is straightforward to apply. Contrary to SA optimization, the freeze bath method emphasizes the search for various different low energy models rather than the search for the best model itself. This behavior makes the freeze bath method especially suitable for inversion in real geophysical problems.

The freeze bath method implemented here was developed to invert broadband acoustic field data from the Haro Strait geoacoustic tomography experiment.<sup>7</sup> The method itself is described in Sec. I and is demonstrated in Sec. II with simulated data for the Haro Strait environment. The cost function for the search process is based on the correlation of measured and modeled broadband waveforms corresponding to the signals that were received at a vertical line array (VLA) from the light bulb sound sources that were used in

the experiment.<sup>7</sup> Ray theory is used to model the acoustic propagation to the VLA. Although the environmental model is very simple, it is adequate to reveal basic problems due to parameter sensitivities and correlations that are encountered in geoacoustic inversions in shallow water. We show in Sec. II that the efficiency of the sampling process is greatly improved by reparameterizing the initial geoacoustic model using a new independent set of parameters determined from the covariance of the sampled models. Application to the Haro Strait experimental data is presented in Sec. III and the results are summarized in the final section.

## I. FREEZE BATH INVERSION METHOD

### A. The inverse problem

For a particular geoacoustic parameterization of the ocean bottom, each model  $\mathbf{m}$  can be represented by a set of values  $m_i$  for the model parameters

$$\{m_i\} = \mathbf{m} \in \mathbf{M} \quad \text{for } i = 1, \dots, N_M, \quad (1)$$

where  $N_M$  denotes the dimension of the model space  $\mathbf{M}$ . Similarly the data  $\mathbf{d}$  may be written as

$$\{d_j\} = \mathbf{d} \in \mathbf{D} \quad \text{for } j = 1, \dots, N_D, \quad (2)$$

representing an element of the  $N_D$ -dimensional data space  $\mathbf{D}$ .

Geoacoustic inversion is fundamentally limited by the problem that many models may fit the data well. The reasons for the nonuniqueness of the model solutions are well known, and apparent causes can lie both in theory errors associated with the propagation model or the form of the geoacoustic model and the sensitivity of the model parameters, and in data errors. Thus, for inversion problems of the type considered here, the most important question is which set  $\{m_i\}$  or distribution  $\sigma(\mathbf{m})$  of models fits the data well, rather than what single model is the optimum model. In contrast to determining the optimum model, the distribution of models  $\sigma(\mathbf{m})$  is itself the complete solution to the inverse problem.

For any arbitrary multi-dimensional distribution  $\sigma(\mathbf{m})$ , the knowledge of its moments, such as the mean Eq. (3),

$$\langle \mathbf{m} \rangle = \int \sigma(\mathbf{m}) \, d\mathbf{m}, \quad (3)$$

and the model covariance matrix Eq. (4),

$$\text{cov}(\mathbf{m}) = \int [\mathbf{m} - \langle \mathbf{m} \rangle][\mathbf{m} - \langle \mathbf{m} \rangle]^T \sigma(\mathbf{m}) \, d\mathbf{m}, \quad (4)$$

and marginal density function Eq. (5),

$$\sigma_i(\mathbf{m}) = \int dm_1 \int dm_2 \cdots \int dm_{i-1} \int dm_{i+1} \cdots \int dm_{N_M} \sigma(\mathbf{m}), \quad (5)$$

are sufficient to describe the distribution. These integrals can be evaluated numerically by sampling the model space in some fashion. Thus, the central issue is how to draw samples from the distribution. Monte Carlo sampling methods have been used in various applications in geophysics.<sup>8</sup> However, more efficient algorithms make use of the concept of impor-

tance sampling in which the selection of sample points is mainly taken from regions where  $\sigma(\mathbf{m})$  is large. Methods such as simulated annealing and genetic algorithms have been applied in geophysics<sup>6,9</sup> and ocean acoustics<sup>10</sup> to evaluate the measures associated with the distribution. In this paper, an alternate approach using a statistical freeze bath method is described.

### B. Sampling on a fuzzy grid

The freeze bath method is based on the same search mechanism as used in the heat bath algorithm<sup>11</sup> in simulated annealing. In this approach, a control parameter analogous to the temperature of a thermodynamic system is reduced according to a designed schedule, to simulate the process of annealing a crystal. The typical heat bath algorithm is as follows:

- (1) Initialize  
Choose a random starting state  $\mathbf{m} = (m_1, m_2, \dots, m_{N_M})$  and starting temperature  $T = T_{\text{start}}$ .
- (2) Sweep  
Let  $\{\mu_1, \mu_2, \dots, \mu_k, \dots, \mu_K\}$  be values within the allowed search range of  $m_1$ . Let  $\{E_1, E_2, \dots, E_k, \dots, E_K\}$  be the corresponding values of a cost function,  $E(\mathbf{m})$ , with the other components of  $\mathbf{m}$  fixed at their current values. Select the new value of  $m_1$  by sampling from the distribution  $P_k \propto \exp\{-E_k/T\}$ . Visit the remaining components  $m_k$  of  $\mathbf{m}$  and update them by the same method.
- (3) Cool  
Reduce  $T$  slightly and repeat the sweep step.
- (4) Stop  
A subjective abort criterion may be defined. For instance, if the model  $\mathbf{m}$  has not changed in the last 50 sweeps, then the system is frozen; stop.

Sweeping through the parameter space during a heat bath iteration requires the evaluation of the energy function at a number of possible values for each model component. Traditionally, these values were taken from a regular grid<sup>6,11</sup> within predefined (or *a priori*) bounds for the maximum and minimum allowed value.

However, the drawback of the regular grid search is the inherent resolution limit that is given by the grid spacing. We present a new method, sweeping on a *fuzzy* grid, which allows model components to attain every possible value without a prohibitive amount of computing time. The allowed range of values for a model component  $\mu$  is divided into  $K$  subranges, as shown in Fig. 1. Each sweep includes two steps. In the first step, a random value  $\mu_k$  is generated in each subrange (open circle) and the corresponding energy  $E_k$  is calculated. The component that was sampled in the previous sweep,  $\mu_{\text{prev}}$  (filled circle), is also included in the set of possible values. In the second step, the new component is chosen by sampling from the distribution  $p_k \propto \exp(-E_k/T)$ . If the newly sampled value is identical with the previously sampled value, the iteration is said to be frozen. The probability for a frozen iteration to occur depends on the temperature  $T$ . In the figure, the sweep  $i$  is an example of a frozen iteration. Since the values  $\mu_1 \dots \mu_K$  are different for each

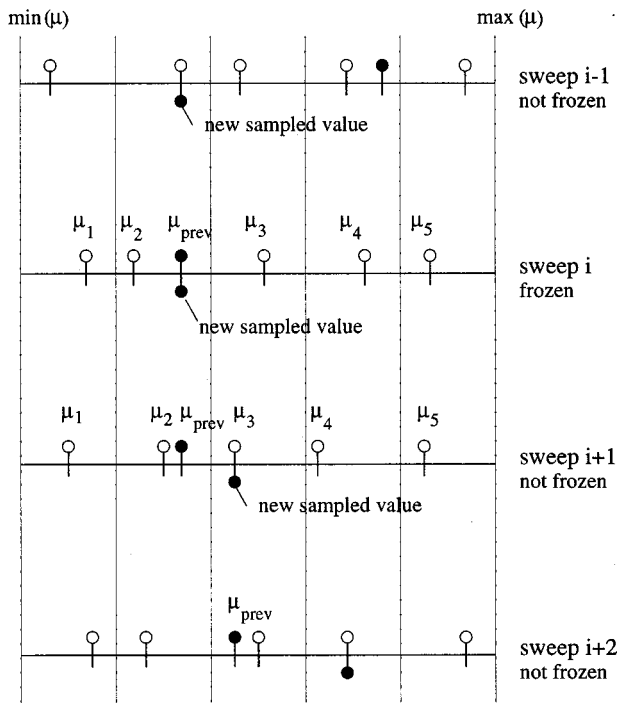


FIG. 1. Sweeping on a fuzzy grid.

iteration, the overall resolution is not limited to the step size of a regular grid. Each parameter requires only  $K$  additional forward model evaluations per sweep, corresponding to the number of subranges; a value of  $K=5$  provided effective sampling and was numerically efficient.

### C. The freeze bath concept

The fundamental difference between the freeze bath method and SA is the distinct handling of the temperature during repeated heat bath sweeps. Whereas simulated annealing is associated with a stepwise decrease of temperature, the freeze bath samples new models at constant temperature during the entire process. The set of sampled models is used to generate a distribution of models that fit the data well. Thus, the freeze bath method provides a measure of the uncertainty bounds for a parameter estimate, in addition to an indication of the best estimated value.

In the physical annealing process of a pure system, the temperature at which a physical system of particles changes its state of aggregation from liquid to solid is called the freezing temperature. When the system is brought to equilibrium at the freezing temperature, frequent transitions between states of the liquid phase and the solid phase occur. Simulating this behavior, the freeze bath method repeatedly samples models at a constant temperature that gives preference to low energy models, but also allows transitions into regions of higher energy. In analogy with the thermodynamic process, the temperature at which the freeze bath method operates is defined as the freezing temperature  $T_*$ . We can describe the freezing temperature intuitively as the temperature at which the sampling of a parameter is characterized by a general preference of low energy states but with an adequate probability to escape from energy minima.

In the freeze bath method the system is initially cooled down from a high temperature to  $T_*$ . Once the cooling process has reached  $T_*$ , the ambient temperature is no longer reduced after each sweep, but held constant. Repeated sweeping at  $T_*$  produces a set  $\{\mathbf{m}\}$  of  $S$  models,  $\{\mathbf{m}\} = \mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_S$ , which are distributed like samples drawn from the Boltzmann distribution<sup>11</sup>

$$P(\mathbf{m}) = \frac{\exp\left(-\frac{E(\mathbf{m})}{T}\right)}{\sum \exp\left(-\frac{E(\mathbf{m})}{T}\right)}, \quad (6)$$

with  $T=T_*$ . Clearly, the distribution density is highest for low energy configurations.

### D. Freeze probabilities

The freeze bath method requires the  $i$ th model component to be sampled at its appropriate freezing temperature  $T_{*i}$ , or equivalently, at a temperature that yields a suitable ratio of frozen to nonfrozen heat bath iterations. In order to quantify the tendency for the model component  $m_i$  to remain at a constant value, we define its freeze probability  $P_{*i}$  at the ambient temperature  $T$  as the probability that its value after the next heat bath sweep is equal to its present value. An estimate of this statistical parameter can be obtained by repeating the heat bath algorithm at constant temperature  $T$  and computing the average

$$P_{*i} \approx \frac{1}{S} \sum_{j=1}^S \# \text{ frozen iterations}, \quad T = \text{const.} \quad (7)$$

where the sum is taken over  $S$  iterations.

Alternatively, the change of  $P_{*i}$  over a range of temperatures can be estimated by evaluating Eq. (7) with subsets of parameter values that are sampled during an SA process. In this case, each subset consists of a number of sweeps and the temperature is approximated by the average temperature of that subset. The range of temperatures within each subset should be reasonably small.

Freeze probability curves provide the information that is needed to tune the method to a particular inversion problem. We assume that, for an optimal freeze bath performance, the freeze probability is the same for all model components. This means that a set of control parameters (individual  $T_{*i}$ 's) can effectively be reduced to a single control parameter  $P_*$ . In a manner similar to the way that the annealing schedule controls the behavior of the simulated melt in a SA application, the freeze probability influences the characteristics of the model set that is sampled by a freeze bath run.  $P_*$  represents a trade-off parameter between average energy and model variety of the model set that is sampled during a freeze. Generally, a low freeze probability will result in a large variety of sampled models with a relatively high energy average. In contrast to that, a high freeze probability will yield a lower energy average, but also a set that lacks variety.

Theoretically, an infinitely long freeze with a uniform sampling temperature results in the Boltzmann equilibrium distribution Eq. (6). By monitoring the energy distribution of the sampled model set  $\{\mathbf{m}\}$ , its degree of convergence to the

TABLE I. Parameter values for the *true* environment.

	Water	Sediment	Half-space
layer thickness [m]	200.0	40.0	-
$c_p$ at top [m/s]	1482.5	1550.0	1900.0
$c_p$ gradient [1/s]	0.0	2.0	0.0
$c_s$ [m/s]	-	80.0	400.0
$\rho$ [g/cm <sup>3</sup> ]	1.05	1.7	1.9
$k_p$ [dB/m/kHz]	0.0	0.03	0.03
$k_s$ [dB/m/kHz]	-	1.0	1.0

Boltzmann distribution can be used to define a stopping criterion. Contrary to that, freezing with uniform *freeze probability* results in an equilibrium distribution different from the Boltzmann distribution. We have not been able to construct the analogous multiple-temperature equilibrium distribution explicitly. However, the particular shape of the limiting energy distribution is only of minor interest. We expect the energy distribution to be a smooth function of energy, with its maximum at low energies. Clearly, one should continue sampling until the energy distribution has converged.

## II. SIMULATION

### A. Simulation environment

In order to demonstrate freeze bath inversion, synthetic data were generated to simulate the broadband signals from the Haro Strait geoacoustic tomography experiment.<sup>7</sup> The inversion is based on matching the broadband waveform with replica waveforms calculated for candidate environments. The synthetic signals and the replica waveforms were computed by convolving a light bulb waveform recorded in the experiment with the impulse responses calculated using the ray theory code GAMARAY.<sup>12,13</sup> This approach provided an efficient and accurate method for use in a global search process with broadband data.

The cost function,  $E(\mathbf{m})$ , was formulated as a function of the correlation between replica and measured waveforms:<sup>14</sup>

$$E(\mathbf{m}) = \frac{1}{2} \left( 1 - \frac{1}{N_{\text{rec}}} \times \sum_{i=1}^{N_{\text{rec}}} \frac{\int q_i(t) \cdot p_i(t) dt}{\sqrt{\sum_{i=1}^{N_{\text{rec}}} \int q_i^2(t) dt} \cdot \sqrt{\sum_{i=1}^{N_{\text{rec}}} \int p_i^2(t) dt}} \right). \quad (8)$$

In this form,  $E$  is related to the average crosscorrelation of measured waveforms and their modeled counterparts. The normalization is chosen to limit the range of  $E$  to the interval  $[0, +1]$ , with  $E=0$  being a perfect waveform match for all  $N_{\text{rec}}$  time series.

Since the dominant features of the Haro Strait data are the reflections from the sea bottom and a subbottom layer, the environmental model was a simple three-layer system consisting of an isospeed water layer and two solid sediment layers. The compressional speed in the upper sediment layer was inhomogeneous with depth; all other geoacoustic parameters were constant. The model parameter values were taken from Hamilton<sup>15</sup> and are listed in Table I. The environment is assumed to be range independent, and the synthetic data calculated to simulate a light bulb signal for a range of 570 m and source depth of 70 m are shown in Fig. 2. The data are calculated for receivers spaced at 6.25 m, with the top receiver at 30 m. The bandwidth for the 70-m light bulb is 300 Hz, centered at 600 Hz. The first signal represents the initial bottom reflection and the second signal is the reflection from the lower sediment boundary. The hydrophone numbers refer to the number scheme of the Haro Strait VLA.

Previous studies of the sensitivity of the geoacoustic model parameters indicated that the most sensitive parameters were the layer depth,  $d$ , and the gradient,  $c'_L$ , and com-

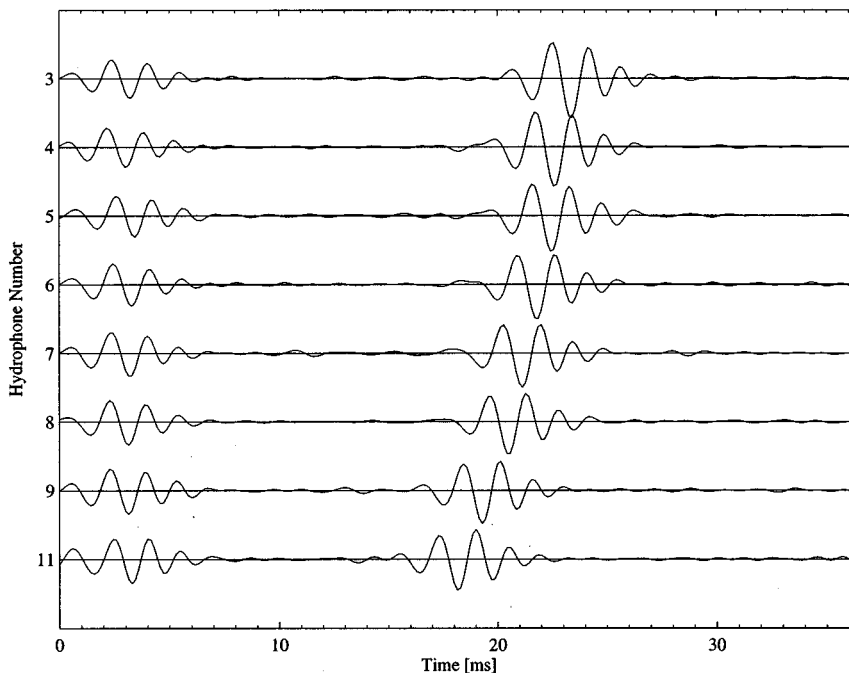


FIG. 2. Synthetic pressure series generated by GAMARAY for the environment in Table I.



TABLE II. Parameters included in the freeze bath search and their allowed range of values. All other parameters were held constant at values given in Table I.

Parameter	min	max	true
$d$ [m]	20	60	40
$c_L$ [m/s]	1520	1600	1550
$c'_L$ [1/s]	0.0	5.0	2.0
$c_H$ [m/s]	1700	2100	1900

pressional speed,  $c_L$ , at the top of the upper sediment layer, and the compressional speed of the lower sediment half-space,  $c_H$ .<sup>7,16</sup> We have restricted the inversion to estimate values of only these parameters, holding all others at their true values. The parameters and their lower and upper bounds are listed in Table II.

### B. Freeze in initial model parameters

Freeze probability curves estimated following the procedure described in Sec. I D are shown in Fig. 3 for the four model parameters. As seen in the figure, the three parameters with roughly the same sensitivity have similar freeze probabilities at all temperatures. Parameter sampling temperatures were estimated for a common freeze probability of  $P_* \approx 0.65$ . This value of  $P_*$  yields a sampling temperature  $T_*$  for the most sensitive parameter (the layer thickness  $d$ ) that is approximately the temperature at which the energy drops to relatively low values in the SA run.

The energy of models that were sampled during  $S = 1000$  freeze bath sweeps is presented in Fig. 4. It can be seen that the energy frequently attained low values throughout the entire freeze bath. Phases of low energy were interrupted by occasional jumps into higher energy regions. In total, 60% of the models have an energy  $E < 0.05$ . The lowest encountered energy was  $E = 0.001$ .

In order to estimate the marginal model densities  $\sigma_i(\mathbf{m})$ , we divide each parameter range in 20 bins and approximate  $\sigma_i(\mathbf{m})$  by the average

$$\sigma_i(\mathbf{m}) \approx \frac{\#(S_{ij})}{S}, \quad \text{for } \min(\text{bin}_j) \leq m_i < \max(\text{bin}_j), \quad (9)$$

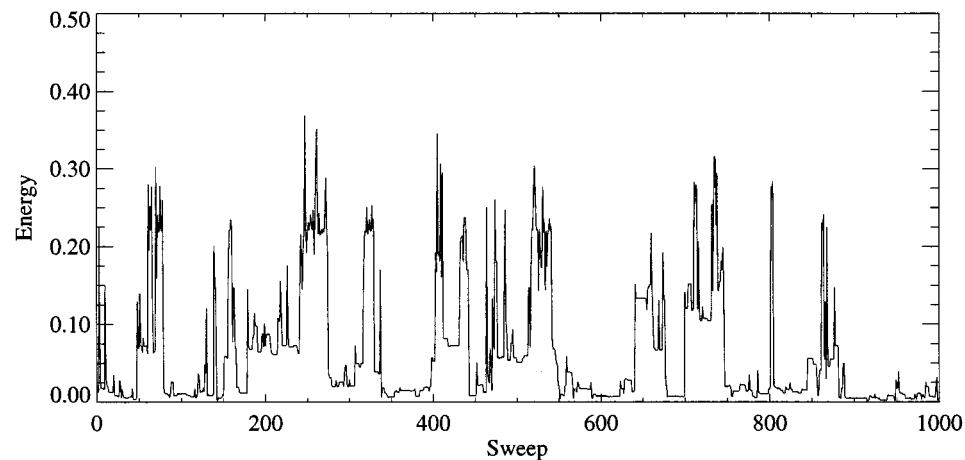


FIG. 4. Model energy during a freeze bath with 1000 sweeps.

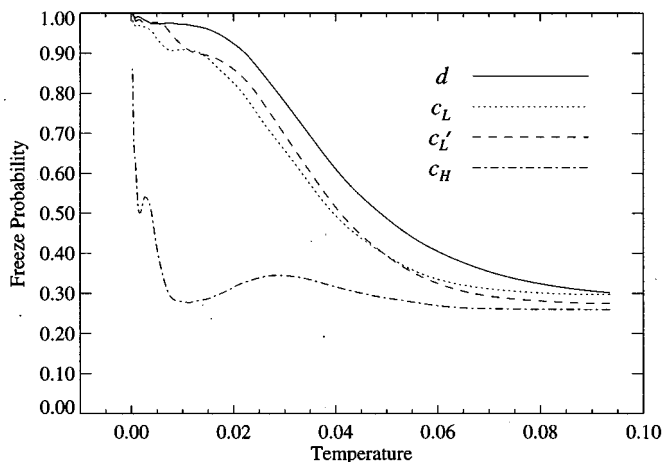


FIG. 3. Freeze probability curves. The appropriate temperatures that give all parameters a desired freeze probability can be read off the curves directly.

where  $\#(S_{ij})$  represents the number of models in the set  $\{\mathbf{m}\}$  for which the  $i$ th model component attained values within the boundaries of the  $j$ th bin. Viewed in this way, the freeze bath is seen as an instance of the technique known in Monte Carlo applications as importance sampling. The marginal distributions of the model parameters  $m_i$  are plotted in Fig. 5. The normalized histograms represent the marginal model densities, with binsizes equal to  $1/20$  of the corresponding parameter range. The true parameter values are indicated by a vertical line.

After 1000 sweeps, none of the  $\sigma_i(\mathbf{m})$ 's has converged. However, a preliminary investigation of the model set gives valuable information about the system. It can be seen that only the marginal distribution of the sediment thickness shows a single peak with a width much smaller than the search range. The position of this peak is very close to the true value, and all models with energy  $E < 0.02$  lie in a narrow region ( $\pm 2$  m) around that value. The accumulation of low energy models in this region, and the unimodal structure of the density distribution indicate a well-defined single energy minimum for this parameter.

The situation is quite different for the remaining parameters, where low energy models are widely spread over the

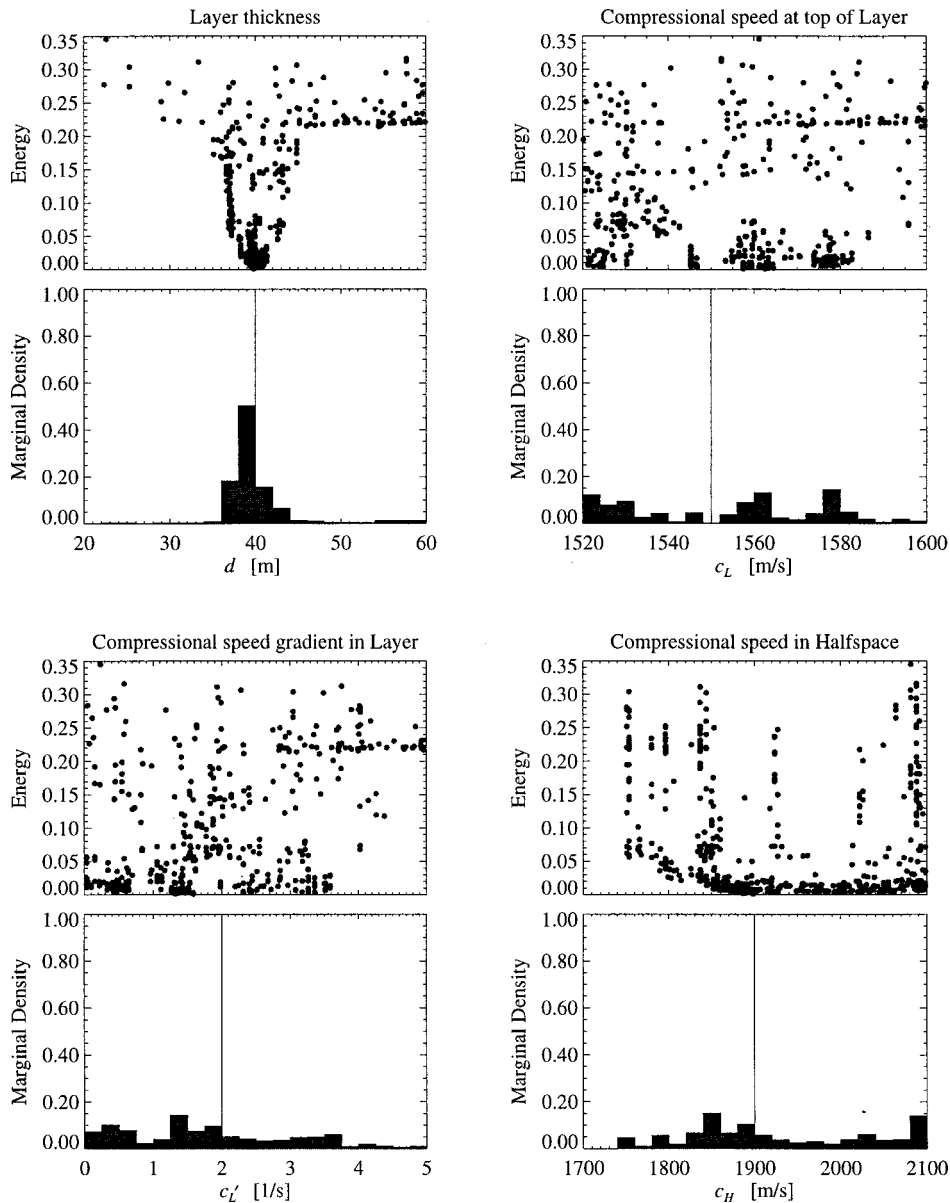


FIG. 5. Distribution of model parameters as a result of a freeze bath with 1000 sweeps. Every model is represented by a dot at the appropriate parameter value and the corresponding model energy. Thus, each of the subfigures contains 1000 dots, where identical models are plotted one upon the other.

search space. Although the sensitivity of  $c_L$  and  $c'_L$  is comparable with that for  $d$ , the marginal distributions have not converged. For each of these two parameters, regions of low energy show up as clusters. The sampled model set, for instance, contains no model with  $c_L$  close to its true value, but low energy models are found at multiple regions above and below the true value. For  $c_H$ , virtually any value above 1850 m/s can yield energies  $E < 0.02$ .

### C. Reparameterization

Although the initial inversion has not produced well behaved (i.e., converged) distributions for all parameters, this relatively short freeze bath is sufficient to estimate correlations between parameters that give valuable insight into the structure of the energy function. First, we estimate the mean model Eq. (3) and the model covariance array Eq. (4) by

$$\langle \mathbf{m} \rangle \approx \frac{1}{S} \sum_{s=1}^S \mathbf{m}_s \quad (10)$$

and

$$\text{cov}(\mathbf{m}) \approx \frac{1}{S} \sum_{s=1}^S [\mathbf{m}_s - \langle \mathbf{m} \rangle][\mathbf{m}_s - \langle \mathbf{m} \rangle]^T, \quad (11)$$

respectively. Again, an integration over a density distribution is replaced by a summation and evaluated by importance sampling. Since the model parameters have different units, we calculate Eqs. (10) and (11) in normalized model coordinates. The normalization was done by transforming the minimal and maximal parameter bounds to the dimensionless values of 0 and 1, respectively. The correlation coefficients,

$$\text{cor}(\mathbf{m})_{ij} = \frac{\text{cov}(\mathbf{m})_{ij}}{\sqrt{\text{cov}(\mathbf{m})_{ii} \cdot \text{cov}(\mathbf{m})_{jj}}}, \quad (12)$$

are measures of the interdependence between the values of individual model parameters. A small absolute correlation coefficient indicates that the corresponding parameters are relatively independent from each other, whereas a high coef-

TABLE III. Parameter correlation coefficients.

	$d$	$c_L$	$c'_L$	$c_H$
$d$	+1.00	+0.18	+0.23	-0.01
$c_L$	+0.18	+1.00	-0.63	+0.04
$c'_L$	+0.23	-0.63	+1.00	-0.10
$c_H$	-0.01	+0.04	-0.10	+1.00

ficient indicates a strong dependence. The calculated values of  $\text{cor}(\mathbf{m})$  are given in Table III. The highest correlation is observed between the compressional speed gradient and the compressional speed at the top of the layer ( $|\text{cor}(\mathbf{m})| = 0.63$ ) and moderate correlation between these parameters and the layer thickness ( $\text{cor}(\mathbf{m}) \approx 0.2$ ). The compressional speed in the half-space shows only a small dependence on any of the other parameters.

In the case of a high correlation coefficient, low energy models are mostly located along a line that is obliquely oriented to the model coordinate axes, and the parameters are said to be coupled. Since the heat bath algorithm determines a new model by sweeping along lines parallel to the coordinate axes, the navigation along an oblique low energy valley is very inefficient. This motivates a change of model components  $m_i$  to a set of new variables  $m'_i$ , whose covariance array is diagonal, similar to the approach of Collins and Fishman.<sup>17</sup> Such a set is obtained by the orthogonal transformation

$$\mathbf{m}' = \mathbf{A}^T \mathbf{m}, \tag{13}$$

where the columns of the transformation matrix  $\mathbf{A}$  are identical with the eigenvectors  $\mathbf{a}_i$  of the model covariance matrix. The elements  $m'_i$  are identified as the coefficients of these eigenvectors.  $\mathbf{A}$  can be found by the singular value decomposition of the covariance matrix

$$\text{cov}(\mathbf{m}) = \mathbf{A} \mathbf{\Lambda} \mathbf{A}^T. \tag{14}$$

Here,  $\mathbf{\Lambda}$  is a diagonal matrix with elements  $\lambda_i$  that are equal to the eigenvalue of the corresponding eigenvector  $\mathbf{a}_i$ .

Freeze probability curves for the new set of transformed or rotated model parameters were obtained using the procedure described in Sec. I. The curves are shown in Fig. 6. In

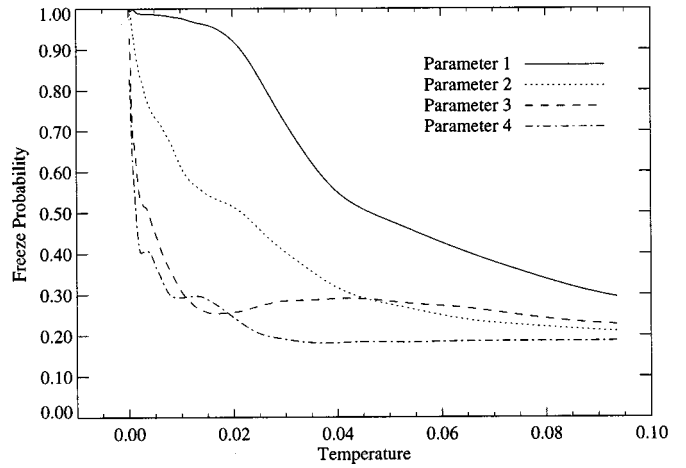


FIG. 6. Freeze probability curves of the rotated parameters  $\mathbf{m}' = \mathbf{A}^T \mathbf{m}$ .

the new parameterization, the dominant components freeze out one after the other, indicating that they are indeed less dependent.

The freeze bath was repeated using the same control parameters of 1000 sweeps and a freeze probability of  $P_* \approx 0.65$ . The energy distribution of the sampled model set is plotted in Fig. 7. Compared to the distribution that was obtained by the initial freeze bath, many more low energy models were found. The freeze bath with the original parameters yielded around 580 different low energy models, whereas the freeze bath with rotated coordinates yielded over 800 different models. The marginal model densities for the rotated coordinates are shown in Fig. 8. It can be seen that the distribution of parameters 1 and 2 have a well-defined single maximum with a narrow peak close to the true value. A representation of the model set in the original coordinate system is obtained by the appropriate inverse transformation

$$\mathbf{m} = \mathbf{A} \mathbf{m}', \tag{15}$$

and the model distributions are shown in Fig. 9. As with the initial freeze bath, only one parameter (the layer thickness) can be resolved. However, a comparison with Fig. 5 reveals mainly differences for the highly correlated parameters  $c_L$  and  $c'_L$ . Whereas the initial freeze bath left many low energy

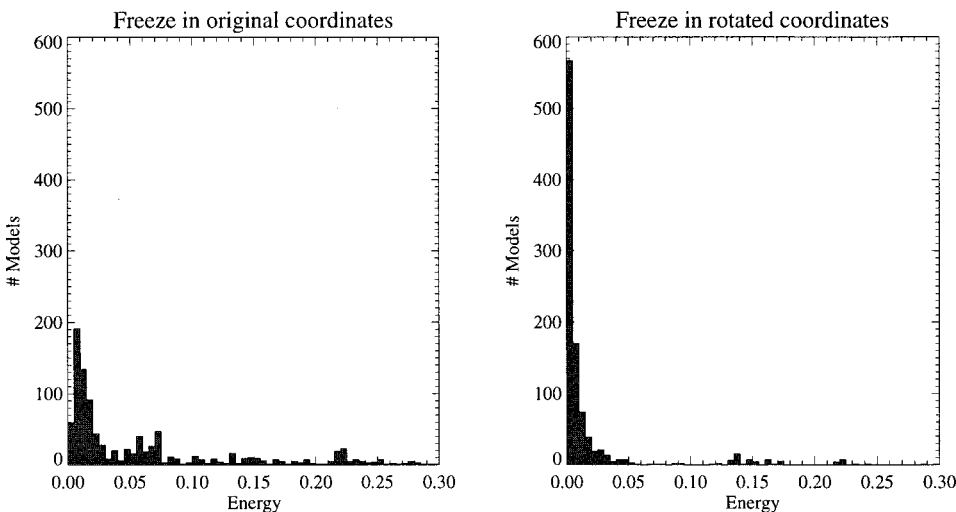


FIG. 7. Histogram of energy values obtained by the freeze bath in original coordinates (left), and obtained by the freeze bath in rotated coordinates (right).

configurations undetected, the second freeze was successful in finding low energy states throughout widespread parts of the search space. This result implies a faster convergence of the marginal parameter density distributions, and therefore a more efficient sampling, if the heat bath algorithm is applied to decoupled parameters.

At this point it is worthwhile to note that the model distributions were obtained using the specific freeze probability of  $P_* \approx 65\%$ . The choice of this value was based on the freeze probability of the most sensitive parameter. In order to estimate the degree to which statistical measures depend on the choice of a particular freeze probability, we repeated the freeze bath of rotated parameters using values of  $P_* \approx 45\%$  and  $P_* \approx 85\%$ . Table IV shows the estimated mean and standard deviation of all parameters for three different freeze probabilities. While the mean model seems to be a relatively robust estimate, the standard deviation varies significantly for different values of  $P_*$ . Generally, the standard deviation decreases with increasing freeze probability. This effect is most dominant for the sediment thickness and

the compressional speed in the half-space, which were seen to be relatively uncorrelated with any other model component. A more useful and robust estimate can be obtained by considering only models with a corresponding energy lower than a certain threshold. Discarding models with higher energy can be justified because in practice, high energy models are usually not of much interest. They merely represent temporary escapes from local minima which are necessary to assure searching over a wide range of model configurations. Ideally, the energy threshold should represent an estimate of the largest mismatch that is deemed an acceptable fit to the data, based on experimental constraints such as signal-to-noise ratio. As an example, we calculated the mean and the standard deviation of all models with  $E < 0.1$ . The results are shown in Table V, and it can be seen that the error estimate of this subset is relatively independent of the particular choice of  $P_*$ . A plot of the compressional sound speed profile estimated with  $P_* \approx 65\%$  is given in Fig. 10. The dashed line represents the true profile and the continuous line represents the mean of the models with energy  $E < 0.1$  (estimated

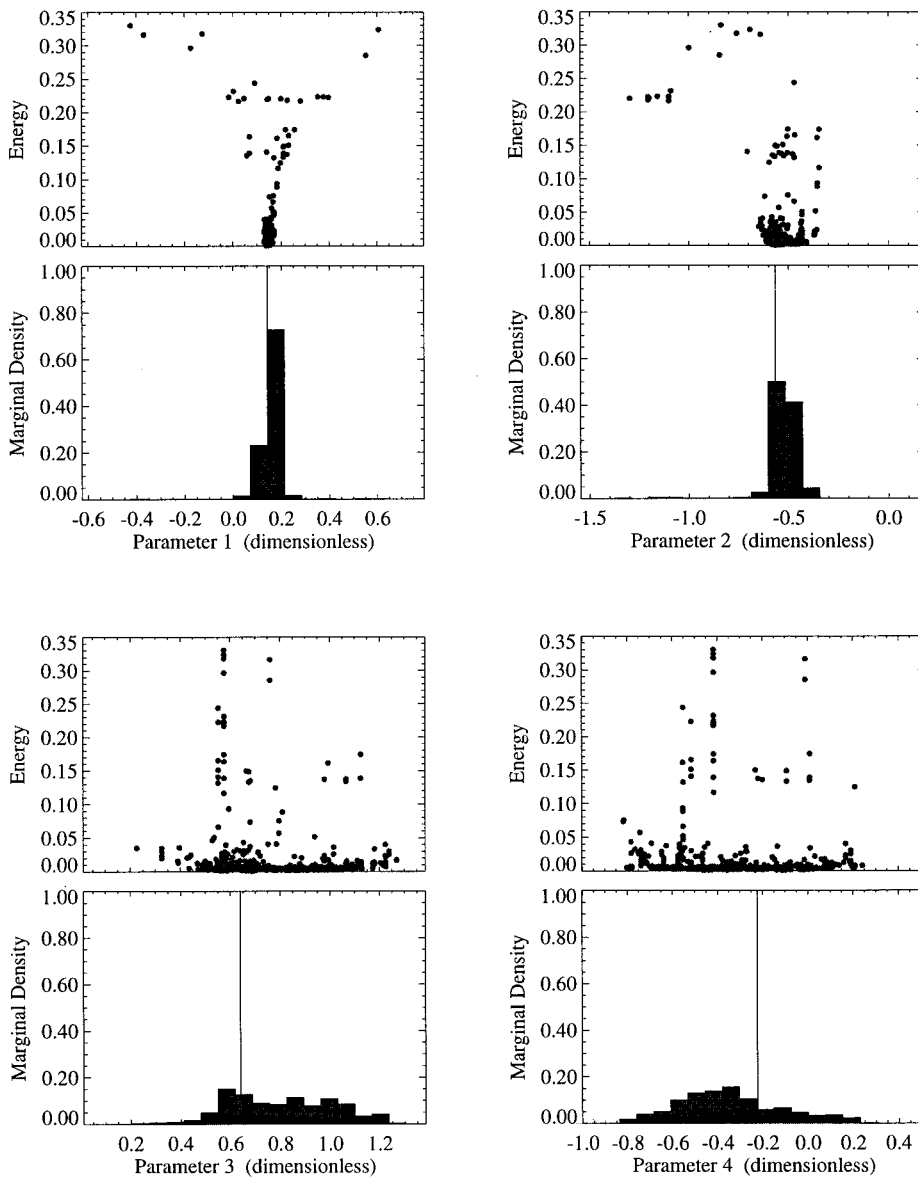


FIG. 8. Distribution of rotated model parameters as a result of heat bath sampling along eigenvectors of the model covariance. The true values are indicated by a vertical line.



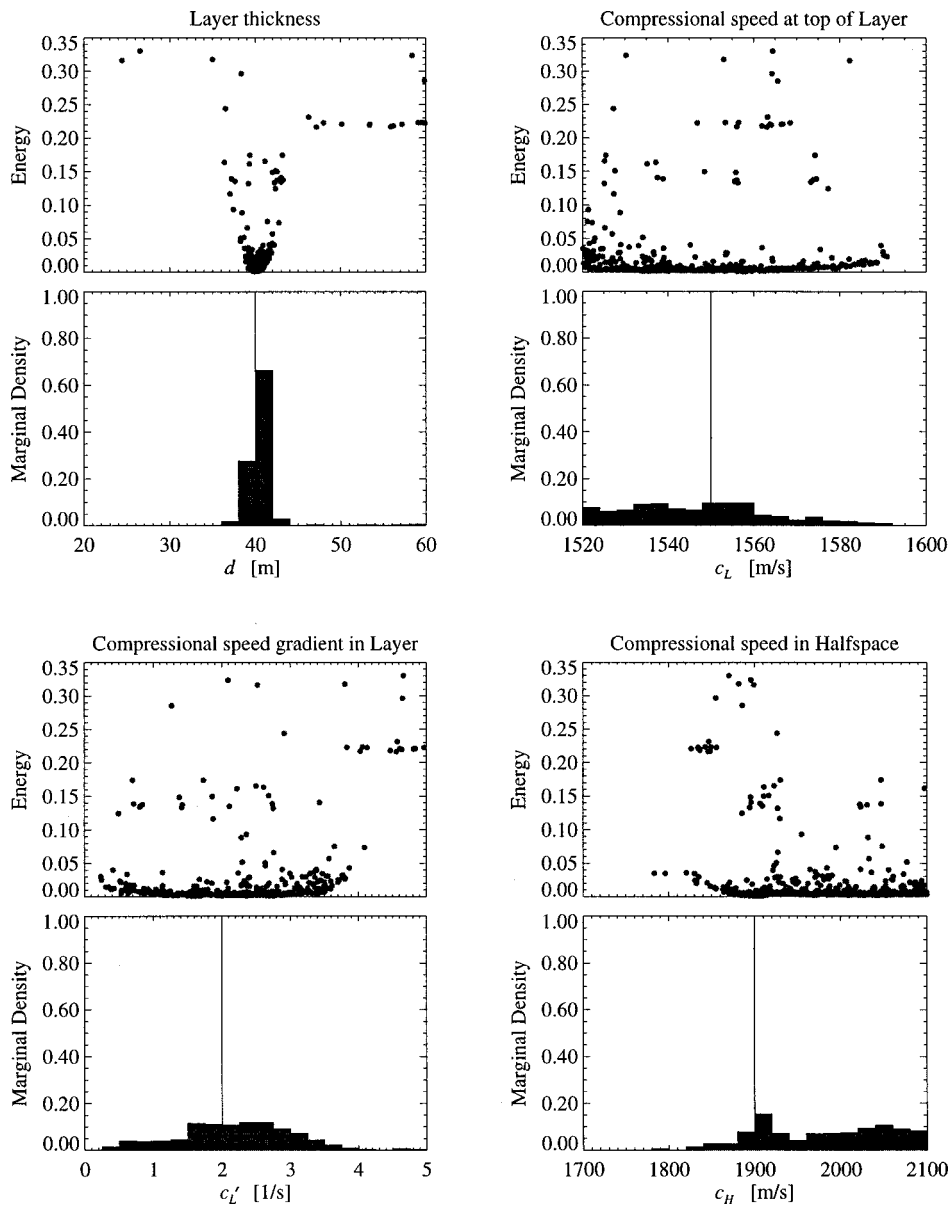


FIG. 9. The model set for the freeze in the rotated coordinate system, expressed in the original coordinates.

with  $P_* \approx 65\%$ ). The shaded area covers all parameter values within the corresponding standard deviation. The fact that the true value for the compressional speed in the halfspace lies outside the standard deviation interval shows that the average of many good models is not necessarily the best model.

### III. INVERSION OF HARO STRAIT DATA

#### A. Haro Strait experiment

The location of the three VLAs and the broadband light bulb sound sources deployed in the Haro Strait experiment

TABLE IV. Dependence of the estimated model mean and the standard deviation on the freeze probability.

	$d$ [m]	$c_L$ [m/s]	$c'_L$ [1/s]	$c_H$ [m/s]
$P_* \approx 45\%$	$40.8 \pm 4.1$	$1554 \pm 23$	$2.1 \pm 1.1$	$1936 \pm 116$
$P_* \approx 65\%$	$40.4 \pm 2.0$	$1547 \pm 16$	$2.2 \pm 0.8$	$1979 \pm 73$
$P_* \approx 85\%$	$40.2 \pm 0.5$	$1541 \pm 18$	$2.2 \pm 0.8$	$1989 \pm 76$

are shown in Fig. 11. The inversion method was applied to data recorded at the NW-VLA in order to estimate the local geoacoustic properties in the vicinity of the array. The VLA consisted of 16 hydrophones spaced at 6.25 m, with a total aperture of 100 m. The top phone was at a depth of about 30 m. The NW-array was selected because the bathymetry at the site was generally flat over an extended area that included the locations of several light bulb deployments (Fig. 11). The average water depth was 200 m, and the water column sound speed profile measured at the site was essentially isospeed with a value of 1482.5 m/s. The array geometry was determined using an inversion method based on minimizing ray

TABLE V. Estimated model mean and the standard deviation of all models with  $E < 0.1$ .

	$d$ [m]	$c_L$ [m/s]	$c'_L$ [1/s]	$c_H$ [m/s]
$P_* \approx 45\%$	$40.4 \pm 0.8$	$1553 \pm 23$	$2.0 \pm 1.1$	$1970 \pm 81$
$P_* \approx 65\%$	$40.2 \pm 0.6$	$1546 \pm 16$	$2.2 \pm 0.8$	$1983 \pm 72$
$P_* \approx 85\%$	$40.1 \pm 0.4$	$1546 \pm 18$	$2.2 \pm 0.8$	$1989 \pm 76$

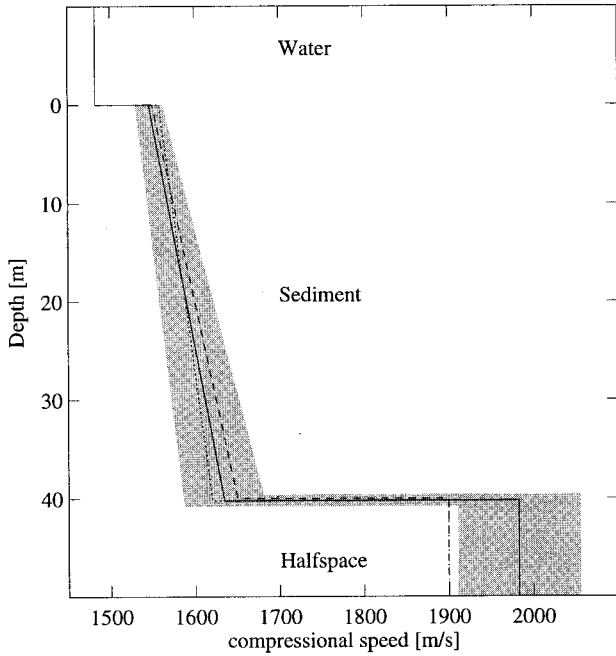


FIG. 10. Best estimated compressional sound speed profile (dotted line).

travel time errors for the direct, the surface-reflected, and the first bottom-reflected paths.<sup>16</sup> The geoacoustic model for the experimental data inversion is the same as that used for the simulation. We inverted for the same four parameters, using the same fixed parameter values and parameter bounds as in Tables I and II, respectively.

The broadband data set used in the inversion consisted of only the initial component of the first bottom reflection signal for the light bulb indicated by the arrow in Fig. 11. The depth of this shot was 61 m, and the range to the bottommost receiver was 490 m. The average spectral signal to noise ratio was estimated as  $>14$  dB within the cutoff frequencies of the applied bandpass filter ( $f_{low}=200$  Hz,  $f_{high}$

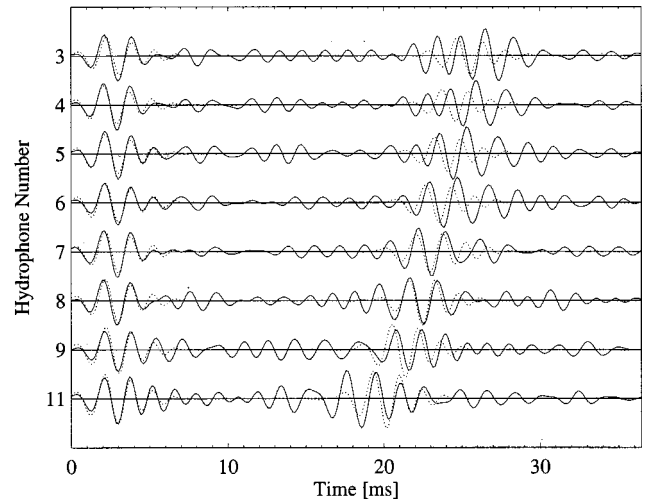


FIG. 12. Measured data (solid line) and best replica data (dotted line).

$=800$  Hz). The filtered first bottom-reflected signals are shown in Fig. 12 for hydrophones 3–9 and 11 that were located in the upper portion of the water column. These data were uncontaminated by surface-reflected paths. The time series contain two main signals, a sea floor reflection and a deeper subbottom signal delayed by about 15–25 ms. The sea floor reflection is a simple waveform indicative of reflection from a single interface, whereas the subbottom signal is dispersed in time, suggesting that the reflection may be from a system of closely spaced layers.

### B. Inversion results

Since the environmental model and the source/receiver geometry of the simulated test case were chosen in close correspondence to the actual Haro Strait data set, we assumed a similar model covariance for the real case. The same set of uncoupled model components was used and all calculations were done directly with this set. A freeze bath

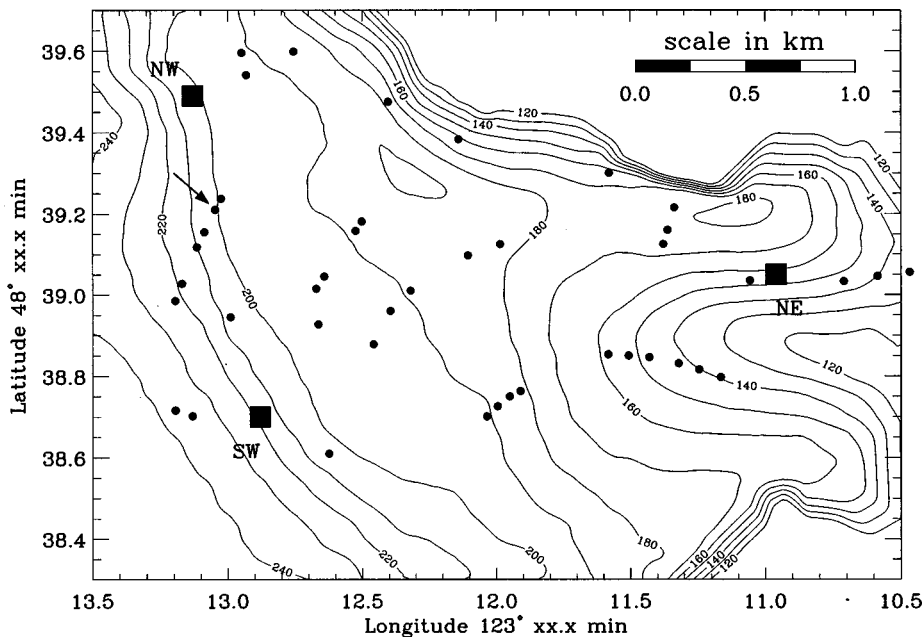


FIG. 11. Contour plot of the local bathymetry (depth in m). Locations of the light bulbs deployed in the experiment are indicated by the circles. Vertical line arrays were moored at positions indicated by the squares at NW, SW, and NE. The inversion method presented in this paper uses the acoustic pressure data from the light bulb that is indicated by the arrow, recorded at the NW-array.

with the rotated parameters was performed using 3000 heat bath iterations, and the set of models obtained was transformed back into the original parameters. The lowest encountered energy was  $E=0.145$ , corresponding to an average waveform correlation of 0.71. The replica data for this model are shown in Fig. 12. While the initial sea floor reflection is in good agreement with the measured reflection for all receivers, the data fit of the second reflection varies throughout the array.

The marginal density functions of the original model components and the distribution of energy versus parameter values are presented in Fig. 13. It can be seen that the main characteristics found in the simulation case also appear for the real data case. The most constrained parameter is the depth  $d$  of the sediment layer.  $c_L$  and  $c'_L$  cannot be resolved individually due to the strong coupling, and  $c_H$  is the least constrained parameter.

An interesting characteristic of the energy distribution is the high density of models with corresponding  $E \approx 0.16 \pm 0.01$  and  $E \approx 0.19 \pm 0.01$ , as indicated by the accumulation

of dots in Fig. 13. These regions with frequently sampled energies are separated by a band that contains considerably fewer models. The band is most notable in the plots for  $c_H$ ,  $c'_L$ , and  $c_L$ , since these parameters were sampled within a wide range between their allowed bounds. It is most likely that these separated regions of low energy belong to models close to the global minimum and to models close to suboptimal minima, respectively. This hypothesis is supported by the sensitivity study<sup>16</sup> which was carried out using the synthetic data. Suboptimal energy sidelobes were seen to occur for  $d$ ,  $c_L$ , and  $c'_L$ .

The distribution for  $c_H$  that shows a much greater density of models at large parameter values can be understood on the basis of the sensitivity study.<sup>16</sup> The sensitivity of this parameter is asymmetric, with significantly greater sensitivity for lower values ( $< 1850$  m/s). At higher values, the sensitivity is very low and nearly constant over the range of values. In light of this result, a more realistic estimate of this parameter would be a value around 1800 m/s, near the values where the distribution begins to flatten.

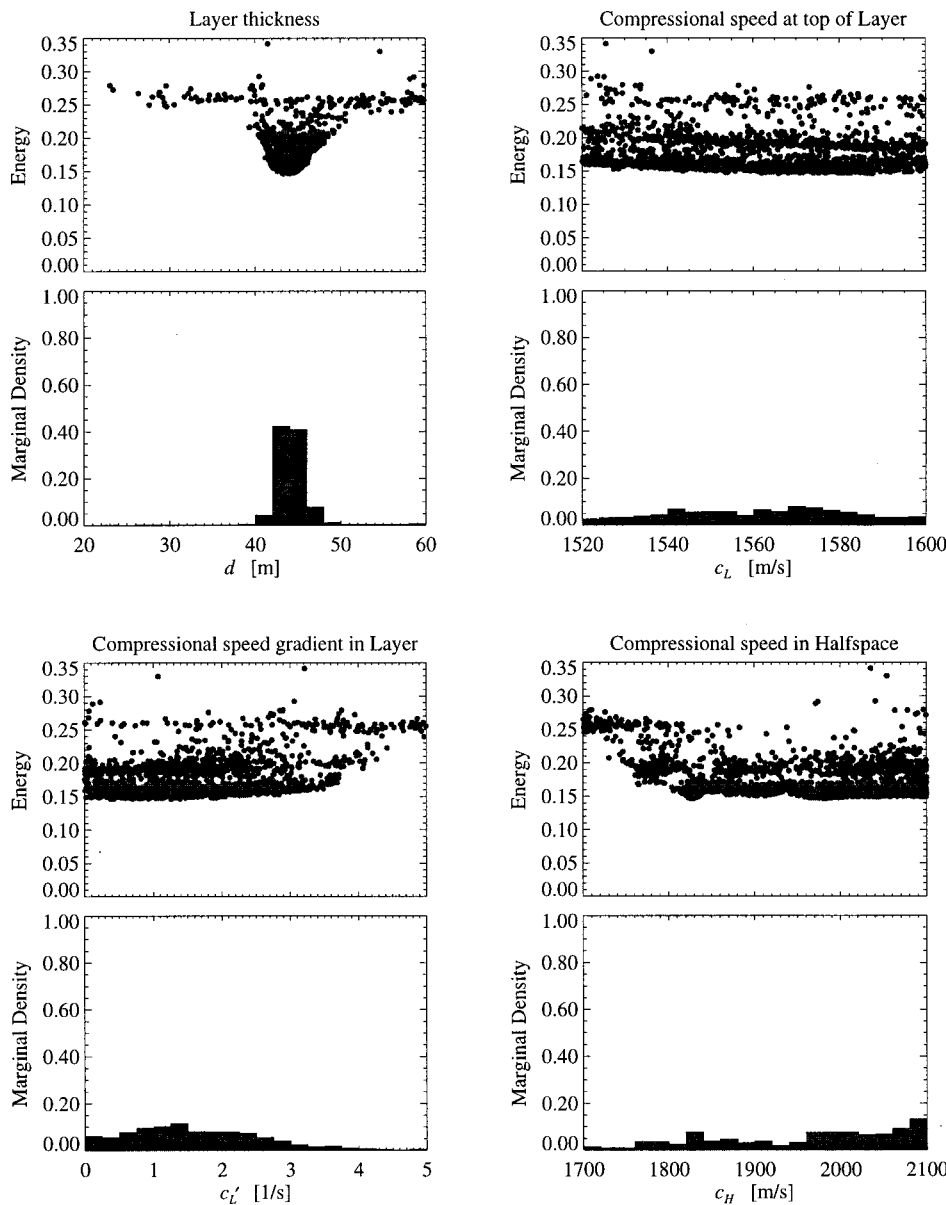


FIG. 13. Distribution of model parameters as a result of a freeze bath with 3000 sweeps.

TABLE VI. Model mean and standard deviation.

$d$ [m]	$c_L$ [m/s]	$c'_L$ [1/s]	$c_H$ [m/s]
$43.7 \pm 1.0$	$1556 \pm 18$	$1.7 \pm 0.9$	$1996 \pm 93$

The accumulation of models with  $E \approx 0.25$  for high values of  $c'_L$  and low values of  $c_H$  can be explained by the fact that these models represent sediment layers with very small acoustic impedance change at the lower interface. As a result, the reflection of that interface is weak and the modeled data effectively contains only the initial bottom reflection, leading to a maximum average correlation of approximately 0.5, i.e., energies  $E \approx 0.25$ .

In order to estimate the extent of the main low energy region, we calculate the mean and the standard deviation of all models with  $E < 0.18$ , i.e., the models that belong to the central low energy lobe. The values are listed in Table VI and the corresponding sound speed profile is shown in Fig. 14. Finally, Fig. 15 shows the distributions of parameter values in two-dimensional subspaces, where the overplotted ellipses represent the 95% confidence area of the corresponding two-dimensional model distributions.

### C. Discussion

The estimated velocity profile is consistent with the local geology that is known to consist of shallow marine sequences of consolidated sand and mudstone layers.<sup>16</sup> In particular, the sea floor velocity is consistent with values expected for the clay–fine sand sediments that were observed in grab samples taken near the NW array.<sup>7</sup> The freeze bath inversion shows clearly which parameters are well constrained by the experimental data.

The correlation of 70% is an acceptable level for waveform matching with experimental data. Propagation model errors can account for part of the mismatch. Although we have assumed a range dependent environment, the bathymetry increases monotonically with a slope less than  $0.5^\circ$  toward the array. Errors also exist in the knowledge of the experimental geometry. On the basis of the acoustic ranging analysis, the absolute range is known to within  $\pm 4$  m and the array and source depths to  $\pm 2.5$  m. These bounds may appear significant for the light bulb frequency band. However, the inversion results do not appear to be greatly sensitive to the error; an inversion using a geometry shifted by about one wavelength in range generated essentially the same estimates of the geoacoustic parameters.<sup>16</sup> The high resolution bathymetric survey carried out during the experiment indicated that the ocean bottom was smooth near the NW array,<sup>7</sup> so the effect of acoustic scattering was assumed to be negligible at the relatively low frequency band of the light bulb signal.<sup>7</sup>

A more significant source of error lies in the form of the geoacoustic model itself. For real data, the lowest energy model merely gives a best fit for a particular parameterization of the bottom. The particular model with a single sediment layer over a half-space was based on the dominant features of the data, and there was no attempt to model the finer structure at the subbottom interface that is suggested by

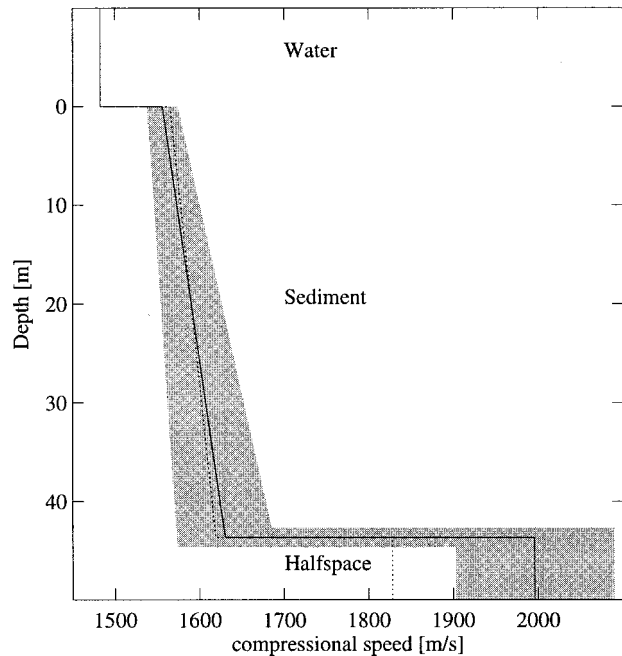


FIG. 14. Best estimated compressional sound speed profile (dotted line). The continuous line represents the mean model and the shaded area covers the standard deviation interval.

the dispersed subbottom reflection signal. Due to this simplification, the best model configuration is not expected to yield an energy value  $E = 0$ , even in the case of perfect measurements (no instrumental noise). Furthermore, due to additional modeling errors as mentioned above, the model with closest resemblance to the true environment may not be given by the model at the global minimum, but by a model that yields higher energy. Viewed in this way, the ability of the freeze bath method to detect a variety of different low energy models becomes an important characteristic for realistic applications.

It is of course possible to use other techniques such as SA to invert the Haro Strait data. In principle, a properly operated SA algorithm would likely converge to a value near one of the minima explored during the freeze bath search at constant temperature. The advantage of using the freeze bath lies in its ability to provide a measure of the confidence limits in the estimated values.

### IV. SUMMARY

We have described a statistical freeze bath method for sampling the model parameter space in order to obtain a solution of the inverse problem in terms of a representation of the *a posteriori* distribution. The method is based on a modified heat bath algorithm that samples on a fuzzy grid in order to access to the complete range of possible parameter values. Contrary to simulated annealing optimization, our implementation of the freeze bath samples the multidimensional model parameter space at constant freeze probability, corresponding to fixed temperatures for each parameter. The result of the inversion is a distribution of good models that fit the data well. Using a simple geoacoustic model that simulates the Haro Strait environment, we have shown that the



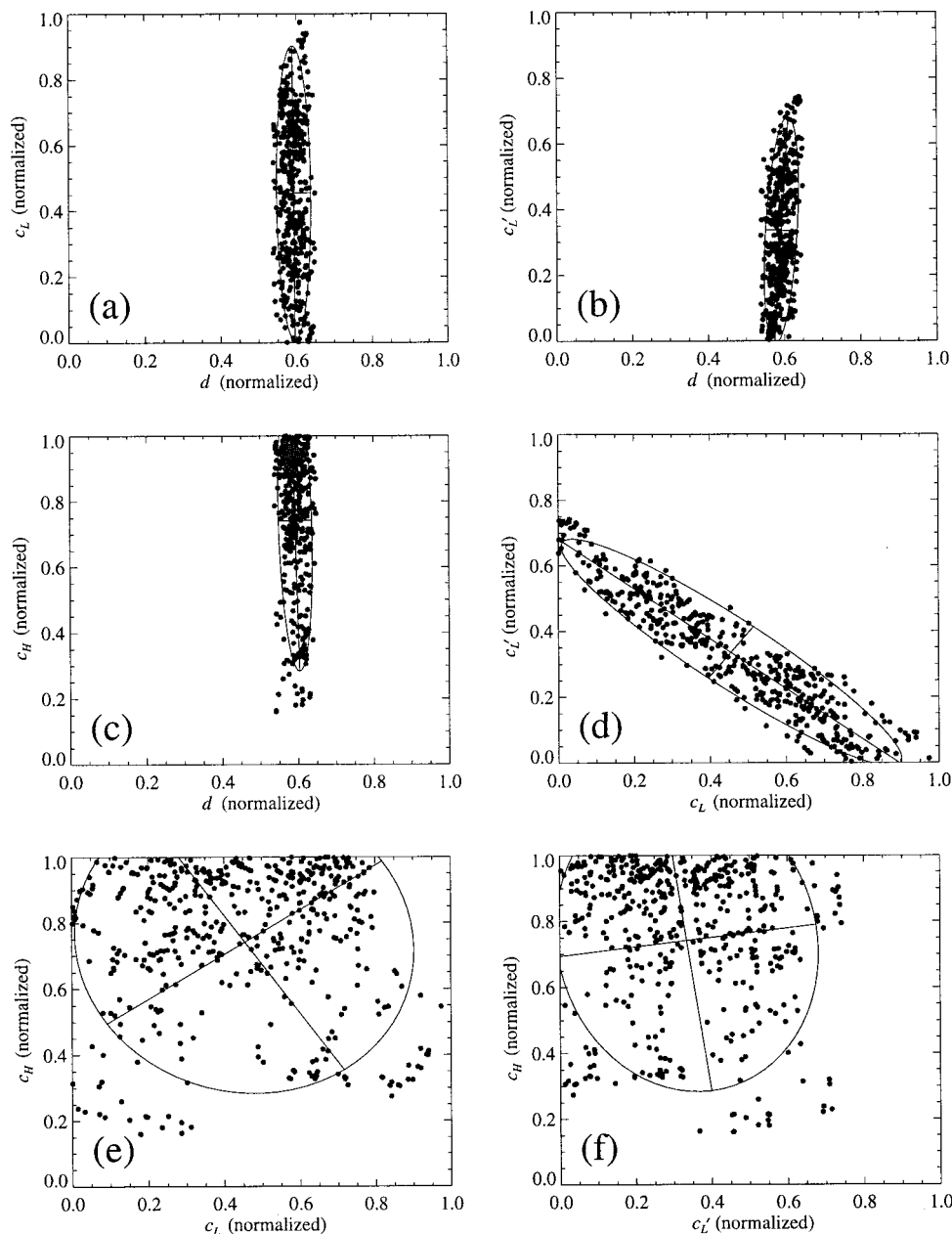


FIG. 15. Model parameters represented in normalized, two-dimensional subspaces. Only models with  $E < 0.18$  are considered.

freeze bath provides distributions that indicate how well the model parameters are constrained by the data, and the degree of correlation between parameters. We have also shown that the efficiency of the search process is greatly improved by reparameterizing to a set of new parameters based on the eigenvalues of the model covariance matrix.

The inversion method was applied to broadband data from the Haro Strait geoacoustic tomography experiment. For this inversion, the cost function was based on the correlation of modeled and measured waveforms reflected from the ocean bottom. The estimated sound speed profile was consistent with the known geology of the region and with limited ground truth information obtained during the experiment. More important, the freeze bath provided objective estimates of the uncertainty of the parameter values.

#### ACKNOWLEDGMENT

This work is supported by the ONR Ocean Acoustics Team, Code 321 OA.

- <sup>1</sup>M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean bottom properties," *J. Acoust. Soc. Am.* **92**, 2770–2883 (1992).
- <sup>2</sup>N. R. Chapman and C. E. Lindsay, "Matched field inversion for geoacoustic model parameters in shallow water," *IEEE J. Ocean Eng.* **21**, 347–355 (1996).
- <sup>3</sup>P. Gerstoft, "Inversion of seismo-acoustic data using genetic algorithms and *a posteriori* probability distributions," *J. Acoust. Soc. Am.* **95**, 770–782 (1994).
- <sup>4</sup>J. P. Hermand and P. Gerstoft, "Inversion of broadband multitone acoustic data from the Yellow-Shark experiments," *IEEE J. Ocean Eng.* **21**, 324–346 (1996).
- <sup>5</sup>A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier Science, B.V., Amsterdam, 1987).
- <sup>6</sup>A. Basu and L. N. Frazer, "Rapid determination of the critical temperature in simulated annealing inversion," *Science* **249**, 1409–1412 (1990).
- <sup>7</sup>N. R. Chapman, L. Jäschke, M. McDonald, H. Schmidt, and M. Johnson, "Low frequency geoacoustic tomography experiments using light bulb sound sources in the Haro Strait sea trial," *Oceans '97 MTS/IEEE, Conference Proceedings Vol. 2*, pp. 763–768 (1997).
- <sup>8</sup>P. W. Cary and C. H. Chapman, "Automatic 1-D waveform inversion of marine seismic reflection data," *Geophys. J. Int.* **93**, 527–546 (1988).
- <sup>9</sup>M. K. Sen and P. L. Stoffa, "Bayesian inference, Gibbs' sampler and

- uncertainty estimation in geophysical inversion,” *Geophys. Prospect.* **44**, 313–350 (1996).
- <sup>10</sup>P. Gerstoft and C. F. Mecklenbrauker, “Ocean acoustic inversion with estimation of *a posteriori* probability distributions,” *J. Acoust. Soc. Am.* **104**, 808–820 (1998).
- <sup>11</sup>D. H. Rothman, “Nonlinear inversion, statistical mechanics, and residual statics estimation,” *Geophysics* **50**, 2784–2796 (1985).
- <sup>12</sup>E. K. Westwood and P. J. Vidmar, “Eigenray finding and time series simulation in a layered-bottom ocean,” *J. Acoust. Soc. Am.* **81**, 912–924 (1987).
- <sup>13</sup>E. K. Westwood and C. T. Tindle, “Shallow water time-series simulation using theory,” *J. Acoust. Soc. Am.* **81**, 1752–1761 (1987).
- <sup>14</sup>M. K. Sen and P. L. Stoffa, “Nonlinear one-dimensional seismic waveform inversion using simulated annealing,” *Geophysics* **56**, 1624–1638 (1991).
- <sup>15</sup>E. L. Hamilton, “Geoacoustic modeling of the sea floor,” *J. Acoust. Soc. Am.* **68**, 1313–1340 (1980).
- <sup>16</sup>L. Jaschke, “Geophysical inversion by the freeze bath method with an application to geoacoustic ocean bottom parameter estimation,” M. Sc. Thesis, University of Victoria (1997).
- <sup>17</sup>M. D. Collins and L. Fishman, “Efficient navigation of parameter landscapes,” *J. Acoust. Soc. Am.* **98**, 1637–1644 (1995).

# A composite model of the auditory periphery for simulating responses to complex sounds

Arnaud Robert<sup>a)</sup>

*CIRC Group, Swiss Federal Institute of Technology–Lausanne, 1015 Lausanne, Switzerland*

Jan L. Eriksson<sup>b)</sup>

*Institut de Physiologie, Université de Lausanne, Rue du Bugnon 7, 1005 Lausanne, Switzerland*

(Received 6 April 1998; revised 3 May 1999; accepted 7 June 1999)

This paper presents a phenomenological model of the cochlea. It consists of a bank of nonlinear time-varying parallel filters and an active distributed feedback. Realistic filter shapes are obtained with the all-pole gamma-tone filter (APGF), which provides both a good approximation of the far more complex wave propagation or cochlear mechanics models and a very simple implementation. Special care has been taken in modeling nonlinear properties in order to mimic the responses of the cochlea to complex stimuli. As a result, the model reproduces several observed phenomena including compression, two-tone suppression, and suppression of tones by noise. The distributed feedback, based on physiological evidence from outer hair cell (OHC) functioning, controls the damping parameter of the APGF and provides good modeling of both low-side and high-side suppression. Responses to more complex stimuli as well as a study of the model's parameters are also presented. Areas of application of this type of model include understanding of signal coding in the cochlea and auditory nerve, development of hearing aids, speech analysis, as well as input to neural models of higher auditory centers. © 1999 Acoustical Society of America.

[S0001-4966(99)04609-3]

PACS numbers: 43.64.Bt, 43.64.Nf [RDF]

## INTRODUCTION

This paper presents a simple and easy-to-implement model of the peripheral auditory system that reproduces many of the phenomena observed in electrophysiological recordings involving tones, two-tones, and tone-noise combinations. It should help explain information encoding and processing by the higher auditory centers, and also provide a physiologically plausible front end for speech analysis. This requires that the model achieves a nonlinear time-frequency decomposition of incoming sounds.

The peripheral auditory system (PAS), composed of external, middle, and inner ear (cochlea) including inner hair cells (IHCs) and auditory-nerve fibers (ANFs), achieves transduction from an aerial sound wave into neural code. This neural code is then transmitted to higher auditory centers where complex sound processing is done. The dynamic behavior of the basilar membrane (BM), part of the cochlea, is influenced by the outer hair cells (OHCs), which are believed to be involved in a complex active feedback mechanism which enhances the passive BM filtering properties and leads to BM nonlinearities such as compression of dynamic range and suppression (Sokolowski *et al.*, 1989; Patuzzi *et al.*, 1989; Geisler, 1990). These nonlinearities, first observed in auditory-nerve responses, are now known to already exist in the BM responses. As studies on auditory-nerve responses have reported (e.g., Geisler and Sinex, 1980; Costalupes *et al.*, 1984; Palmer and Evans, 1982), the presence of broadband noise affects responses of cells of the

auditory system to stimuli such as pure tones or vowels in sometimes complex ways. Lateral interactions, occurring when a second tone or noise is added to a pure-tone stimulus, and compression are likely to play an important role in improving the signal-to-noise ratio (Ruggero *et al.*, 1992).

The mechanical frequency analysis achieved by the cochlea has been the subject of extensive quantitative and qualitative characterization over several decades. These studies, added to many recent physiological observations, provide ample information with which to develop a PAS model. There exist several models of the mechanical tuning of the basilar membrane. These incorporate biophysically based models that include various degrees of structural details and attempt to explain how the cochlea works. So-called “phenomenological” models, on the other hand, seek to reproduce cochlear-like responses without explicit guidance from detailed anatomical and physiological data. Our model fits in the second category.

The filtering process is achieved in the model with a filter bank of all-pole gamma-tone filters (APGFs), introduced by Lyon (1996, 1997). They are derived from the gamma-tone filters, known for providing an excellent fit to the impulse response of ANFs (de Boer, 1975; Aertsen and Johannesma, 1980). The use of APGFs simplifies the implementation of a feedback process which may be extended from a simple local feedback in a single filter of the filter bank (Carney, 1993; Giguère and Wooland, 1994) to a distributed feedback among neighboring filters. In addition, the model includes many other nonlinear properties found in the later stages of the PAS, such as saturation, adaptation, and refractoriness. These elements of our model are similar to

<sup>a)</sup>Electronic mail: Arnaud.Robert@epfl.ch

<sup>b)</sup>Electronic mail: Jan.Eriksson@iphysiol.unil.ch

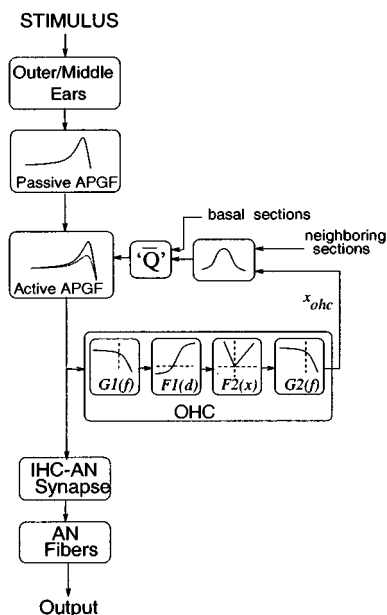


FIG. 1. Model of one local peripheral section. It includes outer/middle ears, BM, and IHC-AN synapse models. The feedback is distributed among neighboring sections. In the presented simulation results, the peripheral model included 120 of such sections.

those in other composite models (Meddis, 1986, 1988; Hewitt and Meddis, 1991).

As a result, the cochlear model reproduces important nonlinear behaviors: (1) compression, with associated changes in the filters' bandwidth for different stimulus sound levels (SPL), (2) suppression, believed to contribute to the masking phenomenon in psychoacoustics and to play a role in perception/coding of speech or other complex signals by our hearing system (Delgutte, 1990), and (3) the shift in rate-intensity functions when noise is present. The model responses of auditory-nerve fibers to tones, noise, and tones in noisy background are also given.

## I. COMPOSITE MODEL

### A. Overview

In the auditory system, the incoming sound is first processed by the external and middle ears and then transferred to the cochlea, where it is transformed into a BM displacement. The latter is finally transformed into action potentials (spikes) in the ANFs. External and middle ears are modeled as a passive bandpass filter. The cochlea, reduced to a BM model with feedback, is modeled as a set of parallel sections, each representing a specific location along the BM—the characteristic frequency (CF). The filtering process in each local BM model is carried out by a cascade of two bandpass filters: (a) one passive, with fixed parameters, and (b) one active, with time-varying properties. A feedback process achieves the tuning of the active filter's parameters and takes into account responses from neighboring sections. The output of each cochlear section feeds an IHC-AN synapse model, and spike trains are finally generated. The general scheme of one cochlear (local) section is shown in Fig. 1. Simulation parameters are given in Table I and II.

TABLE I. Values of the model parameters.

Number of BM sections	120	
Sampling frequency $f_s$	120 kHz	
Function $G1$		
$N$	4	
$w'_n$	$w_n/1.25$	
$Q$	0.8	
$c1$	2.915	
$c2$	-0.534	
$c3$	-0.4529	
$c4$	-0.666	
$c5$	0.4848	
$G2$ cutoff frequency	1000 Hz	
Basal integration	40 sections	
IHC model		
$V0$	-2.32e-2	ANL
$k0$	2.39e-4	0
$k1$	0.1134	7.33e-5
$k2$	2.34e-4	3.66e-4
		8.05e-4

### B. Outer/middle ears

The sound wave incident upon the ear and transferred to the input of the cochlea is mostly passive. Therefore, in the model, this complex is reduced to a single linear transfer function implemented as a first-order Butterworth bandpass filter with low- and high-cutoff frequencies of 1.4 and 20 kHz, respectively.

### C. Traveling wave delay

The traveling wave behavior represents wave propagation along an active cochlear partition. In the present model, each section's input to the BM model was time delayed by a quantity  $\tau$  that encompasses several physiological factors, namely acoustical, traveling wave, and synaptic delays. The delay depends on the section's CF (expressed in Hz) according to

$$\tau[s] = \frac{1}{4 \cdot CF} \cdot \left( 1 + \tanh \frac{\log(600/CF)}{1.4} \right).$$

In developing this formula, the delays already introduced by the filter bank were taken into account; thus its apparent complexity.

### D. Time-varying filters

In the design of the BM filtering, we were concerned with obtaining different filtering properties in response to different stimulus waveforms and levels. At low stimulus levels, the filters must be sharply tuned (narrow-band filter) and, as level increases, the tuning should decrease and a shift of CF towards lower frequencies appear. The use of two filters in each local BM section reproduces these observations.

The all-pole gamma-tone filter (APGF) was chosen as a basis of the filter bank. It is derived by discarding zeros from the well-known gamma-tone filter. Advantages of the APGF are: (a) its simplicity—small number of parameters, (b) a more controlled behavior of the tuning-curve tail—always a



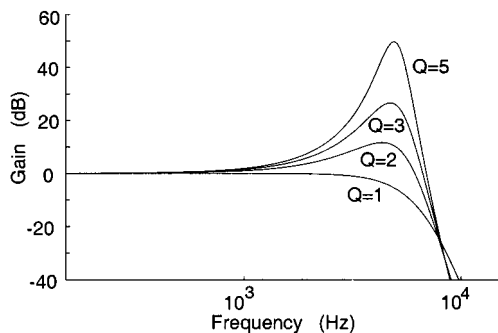


FIG. 2. Gain responses of an APGF tuned at 5 kHz, for different quality factor ( $Q$ ) values.

unitary gain at dc, and (c) easy modeling of level-dependent gain, bandwidth, asymmetry, and center-frequency shift. Lyon (1996, 1997) has made a thorough analysis of these filters and has derived equations which greatly facilitate their use. Relations between the gain  $H$ , the natural frequency  $w_n$ , and the quality factor  $Q$  are given in Eqs. (1), (2), and (3) for an  $N$ th-order APGF. The response of one APGF tuned at 5 kHz is shown in Fig. 2 for different values of  $Q$ .

$$H(s) = \frac{1}{\left(1 + \left(\frac{s}{w_n \cdot Q}\right) + \left(\frac{s}{w_n}\right)^2\right)^N}, \quad (1)$$

$$|H(f)| = \frac{1}{\left(1 + \left(\frac{1}{Q^2} - 2\right) \cdot \left(\frac{2\pi f}{w_n}\right)^2 + \left(\frac{2\pi f}{w_n}\right)^4\right)^{N/2}}, \quad (2)$$

$$H_{\max} = \left(\frac{Q}{\sqrt{1 - \frac{1}{4 \cdot Q^2}}}\right)^N. \quad (3)$$

The decrease of tuning properties and the CF shift as stimulus level increases was successfully modeled by a cascade of two APGFs in each BM section, one *passive* and one *active*. The passive filter is an order 2 APGF with constant parameters and center frequency set below the section's CF. The active filter is an order 4 APGF with a fixed natural frequency  $w_n$  and a damping parameter  $Q$  that is set by the feedback process, and therefore time varying and dependent on stimulus waveform. Initial CF and bandwidth, which are the starting values of all simulation runs, are set according to physiological data from the cat (Greenwood, 1961, 1990; Liberman, 1982). Responses of the cascade combination are shown in Fig. 3 for different stimulus levels, varying from 0 to 80 dB in 20-dB steps. High-order filtering, due to the cascade, leads to sharp tuning properties at low stimulus levels, where the effect of the passive filter is small. The CF shift with increasing level results from both the passive filter's cutoff frequency set at below CF and the effective disappearance of the effects of the active filter due to the decrease of its damping parameter at high levels.

### E. Feedback control of quality factor

The goal of the feedback process is to achieve modeling of the BM nonlinear behavior by continuous modification of

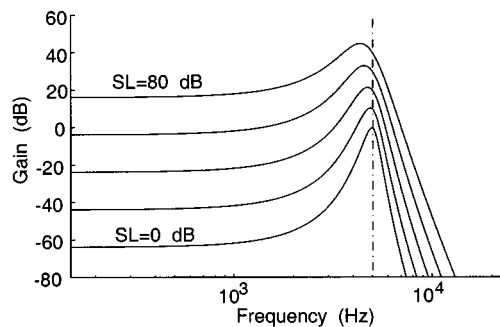


FIG. 3. Gain responses of the active and passive APGFs cascade tuned at 5 kHz. Stimuli consisted of 5-kHz tones at different levels (SPL), from 0–80 dB in 20-dB steps. Decrease in tuning sharpness and shift of center frequency with increasing input level are shown.

the active filter's parameters, in a way dependent on the stimulus temporal and spectral properties. The feedback is based, in part, on physiological data of OHC functioning (Dallos 1985, 1986; Russell *et al.*, 1986; Geisler, 1990). The principle is similar to the one used in Carney (1993), except that the filtering stage is simpler and, in addition to a functional OHC model, we included a distributed form of feedback that updates the filtering properties. By distributed, we mean that the feedback is not local to its corresponding BM section but also affects neighboring sections.

The feedback loop, illustrated in Fig. 1, is composed of (1) an OHC model consisting of (a) a low-pass filter  $G1(f)$ , (b) an asymmetric saturation, (c) a rectification function, and (d) a smoothing filter  $G2$ ; (2) a Gaussian summation of outputs from neighboring sections, and (3) a computation of an *effective* quality factor by summing effects of neighboring basal (high-frequency region) sections.

### 1. OHC model

The low-pass filter ( $G1$ ) has a cutoff frequency set below its section's characteristic frequency. As a consequence, for tones at CF, the feedback threshold is higher than the BM excitation threshold, and the suppression area is modified because of reduction of the suppressive effect of higher than CF frequencies. Without  $G1$ , the suppression area would mostly be comprised of frequencies higher than CF and the suppression strength would be much too important.  $G1$  was implemented as an order 4 APGF with a low quality factor ( $Q=0.8$ ) and a natural frequency of  $w'_n = w_n/1.25$ , where  $w_n$  is the natural frequency of the local BM section considered. The value of the ratio between  $w_n$  and  $w'_n$  is rather arbitrary; any increase of its value will increase the suppression threshold and shift the suppression area towards frequencies lower than CF.

In order to correctly model the compressive BM nonlinearity, the feedback must have no effect at low SPLs, increasing effect for medium to high SPLs—approximately from 30 to 80 dB—and saturate for even higher levels. This is taken care of, in the OHC model, by an asymmetric saturation followed by a rectification function. These were inspired by the observed saturation of the OHC membrane potential as a function of BM displacement, and observations that deflections of hair bundles in both directions do occur.

The saturation function is similar to the ones proposed in (Mountain and Hubbard, 1996; Russell *et al.*, 1986), and is given by

$$F_1(d) = \frac{c_1}{1 + e^{(d+c_2)/c_3} \cdot (1 + e^{(d+c_4)/c_5})^{-c_6}}, \quad (4)$$

where  $d$  is the BM displacement and the set of parameters  $c_{1, \dots, 5}$  are given in Table I. The parameter  $c_6$  is adjusted so that  $F_1(0) = 0$ . The rectification function is given by

$$F_2(y) = \begin{cases} \frac{r}{\min(F_1)} \cdot y & \text{if } y < 0, \\ \frac{r}{\max(F_1)} \cdot y & \text{if } y > 0, \end{cases} \quad (5)$$

where  $y = F_1(d)$ . The parameters  $r$  is a scaling factor, adjusted to produce the desired compression. It is greater for those sections with high gains (high initial quality factor):  $r = r_0 \cdot Q_{i,0}$  ( $r_0 = 0.862$ ).

The output of the rectifier is smoothed using a first-order Butterworth low-pass filter ( $G_2$ ) at cutoff frequency of 1000 Hz. Such a value is necessary as fast variations are seen at the OHC's output; it is also supported by physiological data (Dallos, 1985) and is in accordance with the measured time constant, in the range of 1 ms.

Because many questions remain unanswered regarding the cochlea's active mechanisms, it is still premature to speak of a realistic modeling of the OHCs and their role in the feedback (Hudspeth, 1997). Our model should be considered as a "black box," whose details are based on, but do not necessarily correspond to, physiological processes.

## 2. Gaussian average

A measure of activity of neighboring sections is introduced by computing a local average of OHC outputs,  $x_{\text{ohc}}$ , according to

$$\bar{x}_{\text{ohc}}(i, t) = \sum_{k=i_{\text{base}}}^{i_{\text{apex}}} e^{-((k-i)/3)^2} \cdot x_{\text{ohc}}(k, t). \quad (6)$$

The Gaussian weighting function is defined with respect to the channel index. The index standard deviation of 3 corresponds to a distance standard deviation of 0.33 mm for the filter density used in these simulations. This smoothing of OHC outputs may be said to reflect either longitudinal coupling or the effect of efferent feedback from the brainstem involving neurons with less-sharp tuning than afferent fibers.

The signal  $\bar{x}_{\text{ohc}}$  represents the reduction factor, in dB, of the active filter's gain. We now compute an intermediate quality factor by first converting  $\bar{x}_{\text{ohc}}$  in sound-level units, then deducing the corresponding maximal gain  $\bar{H}_{\text{max}}$ , and finally computing a quality factor,  $\bar{q}$ , corresponding to an averaged activity in neighboring channels. Hence, we have

$$\begin{aligned} \bar{x}_{\text{pohc}} &= 10^{-\bar{x}_{\text{ohc}}/(10 \cdot N)}, \\ \bar{H}_{\text{max}} &= \max(1.0, H_{\text{max},0} \cdot \bar{x}_{\text{ohc}}). \end{aligned} \quad (7)$$

The effective quality factor of the section,  $\bar{q}$  is deduced from  $\bar{H}_{\text{max}}$  using Eq. (3).

## 3. Basal summation

If we look at the BM response as a traveling wave, it is appropriate to consider the output of one BM section as the input to the next. If the gain of a section basal to section  $i$  is reduced, responses to frequency components at  $f = CF_i$  are influenced and this is measured in our model by a gain modification factor  $\delta_k$ . The latter is computed as the logarithmic difference between the gain  $H_k$  evaluated at frequency  $f = CF_i$  for a quality factor  $Q = \bar{q}_k$ , and the gain  $H_k$  evaluated at the same frequency but for its initial quality factor  $Q_{k,0}$ . We therefore have

$$\delta_k(\text{dB}) = \log\left(\frac{H_k(CF_i; \bar{q}_k, t)}{H_k(CF_i; Q_{k,0})}\right).$$

To model the effect of the traveling wave along the BM partition at section  $i$ , we add the gain modification factors for a number of *basal* sections, represented by a set  $S$ . In our implementation, the set  $S$  was limited to a maximum of 1/3 of the total number of sections, namely 40 basal sections. The effective gain computed at the output of the feedback process is thus

$$\log(H'_{\text{max},i}) = \sum_{k \in S} \delta_k. \quad (8)$$

We deduce the *effective* quality factor of the BM section,  $Q'_i$ , using the procedure described by Eq. (7). The parameter  $Q'_i$  will determine the active filter's properties—its bandwidth and gain.

## F. IHC-AN synapse

The transformation of the mechanical basilar membrane filtering into a firing pattern in the auditory nerve (AN) is accomplished by the inner hair cells (IHCs). Each local BM section's response excites an ensemble of IHCs. The IHC-AN synapse model ensures the transformation from release probability of neurotransmitters to electrical impulses (spikes) occurrence. Three important properties of IHC are modeled here: half-wave rectification, low-pass filtering, and adaptation, i.e., a reduction in firing activity with time.

Meddis (1986, 1988) has developed a model of inner hair cell/auditory-nerve synapse that reproduces several characteristics of auditory-nerve firing patterns. It is based on multiple stores of transmitter substance, one of which contains a time-varying quantity available for immediate release. We use the same basic three-reservoirs model, but with a modification to the permeability factor. Briefly, permeability  $k(t)$  is an instantaneous nonlinear function of input signal amplitude—here, filtered BM displacement. The amount of transmitter released into the synaptic cleft in an interval  $dt$  is defined as  $k(t)q(t)dt$ , where  $q(t)$  represents the amount of transmitter available for release. In our case,  $k(t)$  is derived from the basilar-membrane displacement in two stages, first by transforming the output of the BM model into a hair cell membrane potential,  $V_{\text{IHC}}$ , and then making the membrane permeability a function of  $V_{\text{IHC}}$ . Following Russell and colleagues (Russell *et al.*, 1986),  $V_{\text{IHC}}$  is derived from the BM displacement via paired rectangular hyperbola similar to that

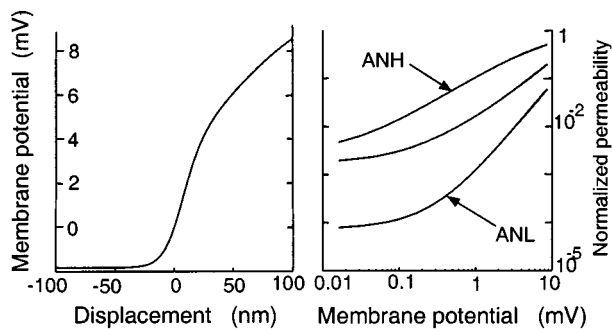


FIG. 4. Left: IHC membrane potential as a function of hair bundle displacement. Right: Transmitter normalized permeability as a function of membrane potential for both ANH and ANL fibers.

of Eq. (4), fitted to their data (Fig. 4, left). The output of this function is then low-pass filtered (first-order Butterworth) with a cutoff frequency of 2 kHz. The result is that, for high frequencies, the ac component of the membrane potential is substantially reduced compared to the dc component and is virtually absent above 5 kHz. This is computationally convenient as the sampling frequency can be reduced. Then,  $k(t)$  is derived from  $V_{IHC}(t)$ , using the function

$$k(t) = k_0 + k_1 \cdot \max(0, V_{IHC}(t) - V_0) + k_2 \cdot (\max(0, V_{IHC}(t) - V_0))^2.$$

The quadratic term is needed to model low spontaneous rate IHC/AN synapses (see Geisler and Greenberg, 1986 for a discussion on this topic). Parameters were selected to produce two IHC/AN synapse permeability functions, shown in Fig. 4 (right). This enables the modeling of auditory-nerve fibers with different characteristics, as shown in Table II, where parameters for fibers of low and high spontaneous rates (SR) are given.

### G. Spike generation

The output of the IHC model provides a time-varying function for probabilistic generation of excitatory postsynaptic potentials (EPSPs) in auditory-nerve fibers. We used a uniform random number generator. The threshold was set so that an EPSP would generate a spike with 100% probability except when the cell was in a refractory state. Spiking activity is blocked for a fixed duration to simulate the spike duration and the absolute refractory period. Influence of occurrence of a discharge on the probability of subsequent discharges is modeled by including a relative refractory period (decreasing exponential). In the simulations, the absolute and relative refractory periods were respectively fixed to  $\tau_{abs} = 0.7$  ms and  $\tau_{rel} = 10$  ms. In order to represent the en-

TABLE II. Characteristics of the two types of AN fibers used in the simulations.

Type	Spont. rate (spike/s)	Threshold (dB)	Saturation (dB)
ANH	65	5	30
ANL	2	30	>90

semble of the ANFs population, we modeled high and medium/low spontaneous rate fibers, differing by their IHC-AN synapse parameters.

In simulations involving long-term adaptation—of the order of a second or more—known to take place in the auditory nerve, the output of the IHC, which determines the probability of firing, is reduced by a factor computed on the basis of mean firing prior to the signal onset. This factor decreases linearly from 1 for low rates (less than 80 spikes/s) to 0.6 at the maximum rate (approx. 250 spikes/s). As an example, a rate of 200 spikes/s would be reduced to 152 spikes/s.

### H. Implementation

In all the simulations reported below, the basilar membrane was divided into 120 sections. The distribution of CF along frequency axis and the bandwidth of the active filter were set according to cat data (Greenwood, 1961, 1990; Liberman, 1982). The parameters values for our simulations are reported in Tables I and II, except for the IHC-AN synapse model, which used the value defined in Meddis, 1988. The APGFs were implemented using the bilinear transform  $s = 2/\Delta t \cdot (1 - z^{-1}) / (1 + z^{-1})$ . The sampling frequency was set to  $f_s = 1/\Delta t = 120$  kHz.

## II. RESPONSES OF THE MODEL

### A. Stimuli

Stimuli consisted of pure tones presented alone or with background noise. Pure tones of different frequencies and intensities were generated to determine the response areas of the model neurons. The same noise signal was used for all experiments (“frozen” noise, low-pass filtered at 20 kHz). The duration of all simulation runs is 250 ms and a root-mean-square (rms) of 200 corresponds to 40 dB. In the case of pure-tone stimuli, the simulation run consisted of 100 ms of silence, followed by the stimulus during 100 ms, and finally 50 ms of silence again. The stimulus had an on/off ramp of 2.5 ms. Responses and intensities, in particular BM amplitude, are all expressed in dB, calculated on the basis rms amplitude of the data.

When simulating responses to mixed tone/noise stimuli, white noise was added at intensities which resulted in a range of signal-to-noise ratios (SNR). The latter were computed integrating the signal energy over all frequencies. The noise onset was at the beginning of the simulation run, with an onset ramp of 40 ms.

### B. Filtering properties

Values of the effective quality factor ( $Q'$ ) and the bandwidth of the cascade of passive and active APGFs are given in Fig. 5 for a 5-kHz pure tone at different SPLs. The tuning sharpening resulting from the feedback process occurs as the bandwidth of the cascade, large at high input levels, decreases for low input levels.

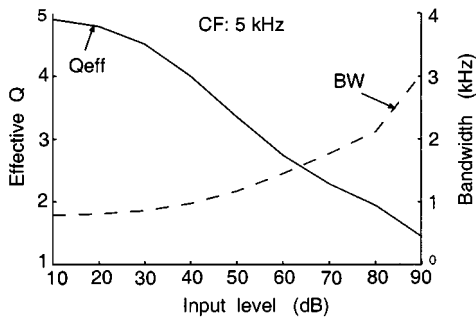


FIG. 5. Values of effective  $Q$  ( $Q_{\text{eff}}$ ) and bandwidth (BW) of the cascade of APGFs. Stimuli consisted of a 5-kHz tone at different levels.

### C. BM tuning curves

The locus of frequency–intensity combinations of tones that cause a just measurable increase—either in BM displacement or AN firing rate—is known as the *threshold tuning curve*. The tonal frequency at which responses are elicited for the lowest levels is the characteristic frequency (CF). At low stimulus intensities, an increase in amplitude is only evoked by a narrow range of stimulus frequencies, while at higher levels a more widespread set of frequencies elicits such an increase. The shapes (symmetry, bandwidth) of the tuning curves change as a function of CF and the best frequency shifts slightly as a function of sound level. Overall, the auditory filters resemble an array of overlapping asymmetric bandpass filters. The model tuning curves, shown in Fig. 6, were obtained by stimulating each BM section by pure tones of varying frequencies and levels.

### D. Compression

#### 1. Responses to tones

Compression is the phenomenon by which the range of intensities of the input signal is reduced to a smaller response range at the cochlear level. Here, compression results from time-varying modification of the APGF gain as a function of sound level and nonlinearity in the feedback loop. It was measured by recording the BM displacement as a function of stimulus levels. The strong compressive nonlinear behavior—up to 40 dB—for frequencies near the CF can be divided into three regions: (1) at low stimulus intensity the function is linear, (2) at moderate SPLs, it displays a power-law form with a flattened slope, and (3) at higher intensities it becomes linear again. The transition from region 1 to region 2 is thought to be a consequence of the nonlinearity of

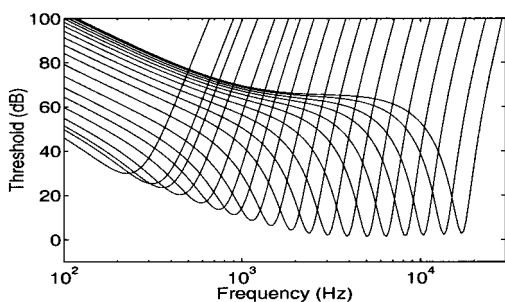


FIG. 6. Model threshold tuning curves at different values of CF.

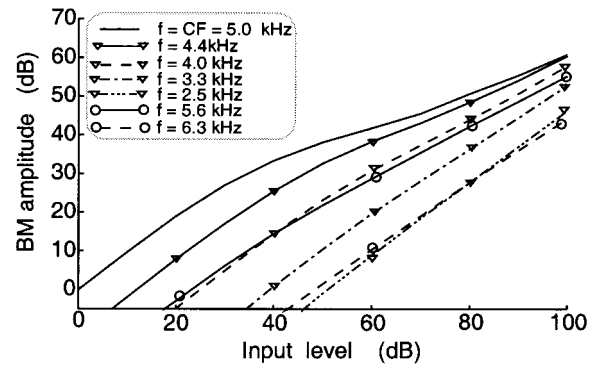


FIG. 7. Compression in the BM shown in the responses to a 5-kHz tone at different input levels. Responses recorded at different CFs, both below and above 5 kHz.

OHC current transduction reducing the drive to the active process, while transition from region 2 to region 3 would be a consequence of the saturation of OHC currents.

In Fig. 7, responses to a pure tone at 5 kHz for varying stimulus levels (from 0 to 100 dB) recorded at a section with the CF equal to 5 kHz are given. The model responses show strong compressive behavior and the three regions described above are clearly observed. Figure 7 also provides results for tones frequencies both lower and higher than the CF. The compressive behavior vanishes for tone frequency more distant from the CF value.

#### 2. BM response with masking by broadband noise

A signal with a broader bandwidth than a pure tone will affect a wider portion of the basilar membrane. Due to the weighted summation at the output of the OHC model from neighboring sections, the damping factor is augmented and the input–output function for a broadband signal will be shifted compared to that of a CF tone.

Responses to a 5-kHz tone in a white-noise background are shown in Fig. 8 for different tone and noise levels. At low-tone SPLs, the BM responses are mainly due to presence of noise. Eventually, the response of the combination (tone–noise) equals that of a tone presented alone. Beyond that point, the tone/noise combination exhibits a shifted version of the tone-alone compressive behavior, although the shift is not great.

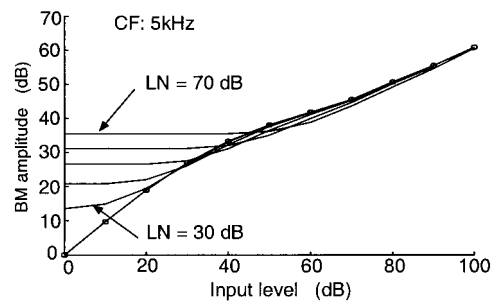


FIG. 8. BM compression in presence of noise. Stimuli consisted of a 5-kHz tone in presence of wideband noise at different levels.



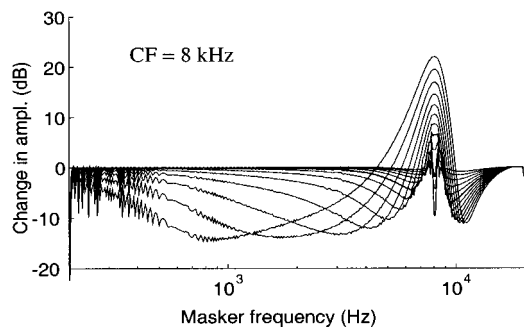


FIG. 9. Suppression behavior measured as the changes in BM amplitude due to presence of masker, relative to BM response for the single tone, as a function of masker frequency. Probe was set at 8 kHz and 40 dB. Masker tones had different frequencies and levels (30–90 dB). Responses recorded at CF=8 kHz.

## E. Suppression

### 1. Two-tone suppression

Two-tone rate suppression (2TS) is a nonlinear property of the cochlea by which the BM displacement or the neuron firing rate in the region most sensitive to a probe tone is reduced by the addition of a second (suppressor) tone at a different frequency. Rate suppression can attain large values: 40–50-dB decreases in input signal level have been recorded in experiments (Delgutte, 1990). The suppression phenomenon, like compression, is attributed to the saturation of the active mechanism in the cochlea, as reported in many recent studies (Geisler, 1990; Cheatham and Dallos, 1990; Ruggero *et al.*, 1992; Geisler and Nuttall, 1997). Although suppression has been extensively observed and studied, its functioning remains incompletely understood (see Geisler and Nuttall, 1997, for a good review). In our model, three stages of the feedback process contribute to modeling suppressive behavior: (1) the saturating nonlinearity, (2) the presence of the filter  $G_1$ , leading to a suppression threshold that is higher than the BM excitation threshold, and (3) the influence from neighboring sections, resulting from Gaussian and basal integrations. Taken together, they should allow good representation of the many suppression phenomena.

### 2. BM response with masking by a pure tone

We assessed the suppression behavior of our model by first running simulations with two tones. The stimuli con-

sisted of an 8-kHz probe tone at 40 dB and masker tones at varying frequencies (200 Hz–20 kHz) and intensities (30 to 90 dB in 5-dB steps). Responses recorded at CF=8 kHz are presented in Fig. 9. The  $x$ -coordinates represent the masker frequency, while the  $y$ -coordinates indicate the difference of responses (BM amplitude) between the probe alone and the combination of tones. Negative values indicate suppression. The model exhibits both low- and high-side suppression, the low and high adjectives referring to the spectral position of the masker with respect to the probe. Suppression occurs only in a limited neighborhood of CF, and is strongest near CF. Suppression is also clearly asymmetrical, in accordance with reported physiological experiments. This asymmetry is more pronounced for low and medium masker intensities for which high-side suppression is dominant. Finally, the width of the suppression area increases with masker level, especially for low-side suppression.

Suppression can also be illustrated in the BM displacement measures, as done in a number of studies. For comparison purposes, we also present the suppression behavior in that form, in Fig. 10 (left). The probe and masker frequencies were fixed, and set, respectively, to 8 and 10 kHz. The probe and masker levels varied respectively from 20 to 90 dB and from 20 to 50 dB, both in 10-dB steps. Figure 10 (right) shows physiological recordings found in Ruggero *et al.*, 1992. There is a strong agreement between our simulation results and their data, as well as results from other models (Goldstein, 1995).

### 3. Low-side phase suppression

It is known that low-frequency suppressors affect responses to high-frequency probe tones in a phasic manner: the CF envelope is modulated at the suppressor frequency, the strength of modulation being related to the suppressor's level. Two phases of suppression occur: one when the BM has greatest displacement toward scalar tympani, and one 180 deg out-of-phase with the first one. The degree of suppression in the two phases is not always equal (Geisler and Nuttall, 1997; Cai and Geisler, 1996; Temchin *et al.*, 1997). Figure 11 shows the effect of suppression of a 40-dB, 8-kHz tone by a low-frequency masker at 80 dB. The masker was frequency modulated between 200 (start) and 300 Hz (end) over the time interval shown. The asymmetry of the function

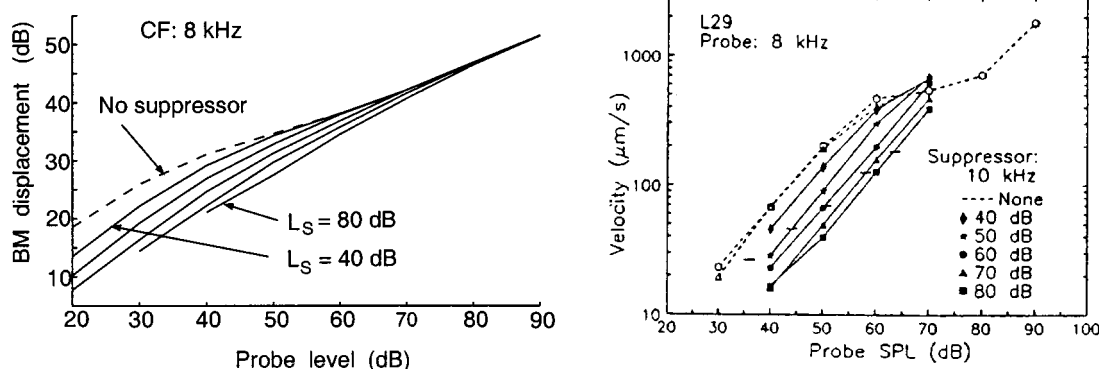


FIG. 10. BM suppression as seen in CF input–output functions for an 8-kHz probe and a 10-kHz suppressor at different levels ( $L_s$ ). Left: Model simulation results. Right: Physiological data, from Ruggero *et al.*, 1992.

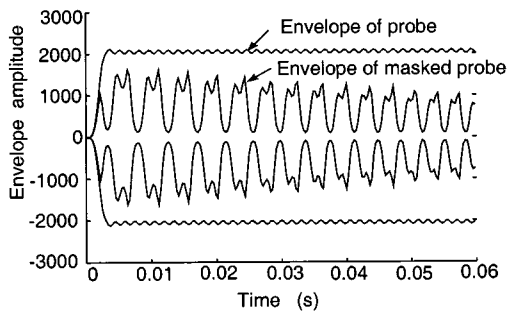


FIG. 11. Modulation effect of low-side suppression. Envelope of 8 kHz, 40-dB pure tone alone and masked tone. The masker, with level of 80 dB, was frequency modulated between 200 (start) and 300 Hz (end) over the shown time interval.

$F1$  in the model enables the simulation of the modulation phenomenon and the two phases of suppression.

#### 4. Responses to notch noises

To emphasize the role of basal-section integration in modeling suppression, we provide BM responses to notch noises (NoN). The spectral representation of a NoN can be seen as a wideband noise with level (NL) in which a notch (hole) is made. The presented NoNs level and width were chosen, respectively, in the range 20–80 dB and 500–6000 Hz, centered around 5 kHz. The NL was set at 30 dB above the notch floor. We simulated BM responses to this family of NoNs and normalized them by the BM response to a wideband noise at NL-30 dB. Figure 12 presents the normalized BM responses recorded at a section with a CF of 5 kHz for the different parameter values. Null relative gains are attained when the rms response of the NoN is equal to the rms response of a wideband noise, while negative values reflect that suppression occurred. A few comments can be made on the figure. First, when the notch width is smaller than the effective width of the section's bandpass filter, the relative gain is greater than 0 since there is more energy in the filter than if wideband noise was presented alone. When the notch width increases to a value where the notch is actually greater than the section's bandwidth, suppression by neighboring sections occurs and therefore the relative gain becomes smaller than 0. Eventually, as the width increases even more, no suppression occurs, resulting in a relative gain tending towards the value 0. Of course, as stimulus level increases,

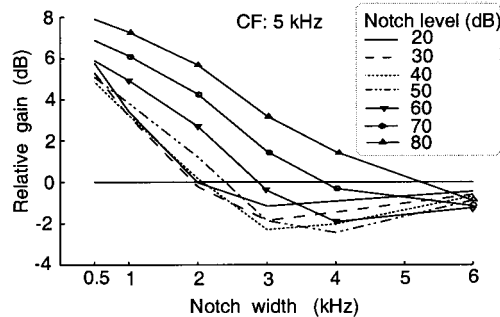


FIG. 12. Normalized (to NL-30 dB) BM responses to notch noises for different notch widths and levels.

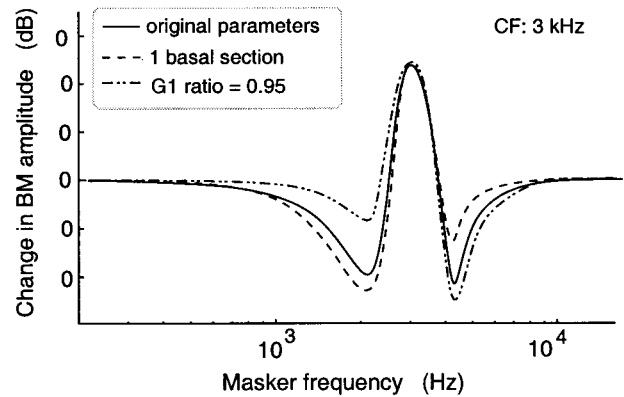


FIG. 13. Effect of model parameters on suppressive behavior. Suppression responses for the original parameter set (bold line), when the number of basal sections is reduced to one (normal) and when cutoff frequency of  $G1$  is increased ( $w_n'/w_n = 0.95$ ) are presented. The frequency and level of the probe are, respectively, 3 kHz and 40 dB. Masker level is 60 dB, with varying frequencies.

the effective bandwidth of the section increases and the point of zero-crossing is shifted towards larger values of the notch width.

#### 5. Further studies on suppression

Suppression in our model results from three main design characteristics (Sec. I). We present here results of  $2TS$  measures that help us understand and discriminate the effects of individual parameters, namely the basal integration width and the filter  $G1$  ratio  $w_n'/w_n$ . Before simulation runs could be started, we ensured that compressive behavior was not altered by readjusting parameter  $k$  and the OHC functioning point.

Suppression results are presented in Fig. 13 when only one section was considered in the basal integration and when the  $G1$  ratio was changed from 1.25 to 0.95. The probe frequency and level were set to 3 kHz and 40 dB. The masker level is 60 dB. Noticeable differences are observed in the suppressive behavior. When  $G1$ 's cutoff frequency is set higher than its section CF, high-side suppression is much stronger, and low-side suppression is significantly reduced. This is a logical consequence of the shaping by  $G1$  of the suppression area. In return, when the basal integration width is reduced to the section itself—there is no more influence of neighboring sections—high-side suppression is strongly reduced. Had we not set the compressive behavior—as stated above—prior to the simulations, the response for modified basal integration width would be higher.

#### F. Responses of ANFs

##### 1. Responses to tones and to noise

In Fig. 14 we show the rate-intensity curves and the response areas of the two groups of ANFs included in this study, recorded at CF of 5 kHz, in response to pure tones. Both onset and steady-state response rate-intensity curves (averaged over 25 simulation runs) were obtained; onset rate was computed during the first 5 ms of the response to the

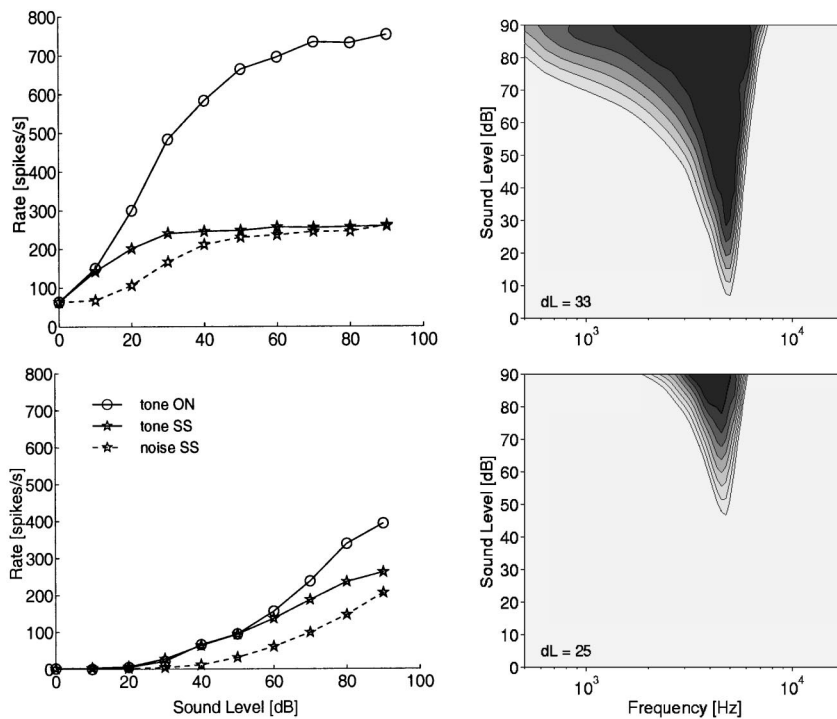


FIG. 14. Responses for ANH (top) and ANL (bottom) AN fibers (CF=5 kHz). Left: Rate-intensity curves. Onset (ON) and steady-state (SS) firing rates in response to pure tone and to noise. Right: Response areas.  $dL$  is the increment in spikes per second between two contour curves.

stimulus and the steady-state rate during the last 70 ms. Also shown are the steady-state rate-intensity responses to a noise presented alone.

The response areas show the activity evoked by tones of varying frequency and intensity; rate was calculated over the full stimulus duration (100 ms), and each stimulus was repeated ten times. As expected from biological ANFs, dynamic range of modeled fibers covers only a portion of the input intensity range, and although frequency selectivity is lost at high intensities for high-spontaneous rate fibers, it is preserved in those with low-spontaneous activity.

The poststimulus time histogram (PSTH) of one fiber's response to a simulation run, shown in Fig. 15, illustrates the effect of adaptation.

## 2. Responses to pure tones in noise background

When simultaneously presenting a broadband noise with a tone, we can obtain the masked average-rate intensity responses area (MRA), which reflects the ability of the noise to alter the response area to the tonal stimulus alone (Rhode and Greenberg, 1994).

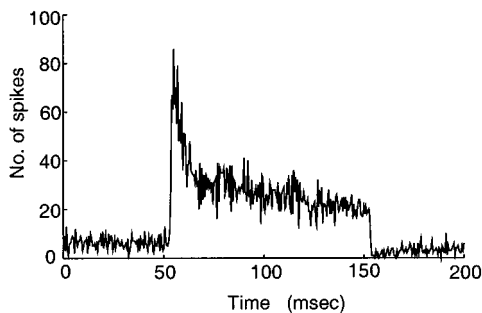


FIG. 15. PSTHs of ANH fiber (CF=5 kHz) in response to a 5-kHz tone at 50 dB.

Figure 16 (left) shows the MRAs for the two types of ANFs included in this study. The MRAs were obtained for varying 5-kHz probe levels. The level of the background noise was adjusted to evoke a rate of activity intermediate between baseline and saturation, *viz.*, 30 dB (top left) and 60 dB (bottom left). Regions of suppressed activity (decrease in mean firing rate of noise response due to tone) are obtained for both ANH (top) and ANL (bottom) fibers, but are most evident in the second case. For ANH fibers, suppression is only observed for above-CF tones. The simulation results are compared to physiological data provided in Rhode and Greenberg, 1994.

The alteration in the response of a cell when a background noise is present can also be displayed in a set of masked rate-intensity curves (MRC), which is simply a family of rate-intensity curves, each with a different background noise level. Figure 17 shows the masked rate-intensity curves for the two ANF types. The mean rate was calculated over the full stimulus duration. The first observation is the increase in baseline activity (mean rate at 0-dB tone intensity), due to the background noise. Second, due to adaptation taking place at the IHC/AN synapse, the fiber's saturation rate is reduced. These two effects combine to reduce the range of firing rates over which the fiber can respond to the tonal stimulus.

The range of firing rates is most affected in AN fibers with the lowest threshold (ANH). For a 70-dB noise level, there is virtually no increase anymore in firing rate with increased tone intensity. Compression due to increased baseline activity and reduced saturation rate does not depend on the compression observed in the BM input-output function. Adaptation of the AN fiber is related to the driven rate. With the *ad hoc* method for modeling long-term adaptation, the MRC curves are shifted downward.

The other important effect of background noise is the

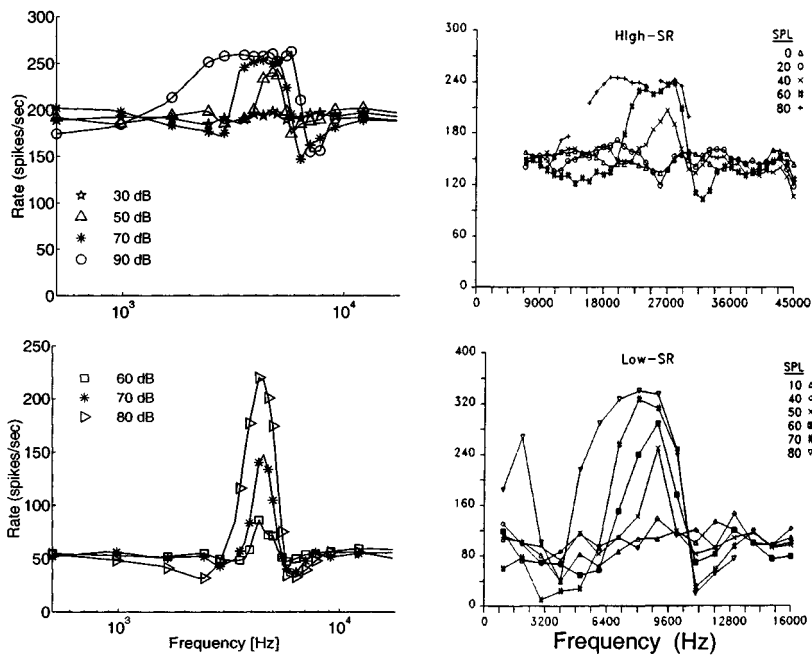


FIG. 16. Masked response areas (MRA) for ANH (top) and ANL (bottom) AN fibers (CF=5 kHz). Left: Simulation results. Right: Physiological recordings, from Rhode and Greenberg, 1994. The noise level was chosen so as to produce about half of the noise-driven rate response. Symbols pertain to the sound-pressure level of the tonal stimuli.

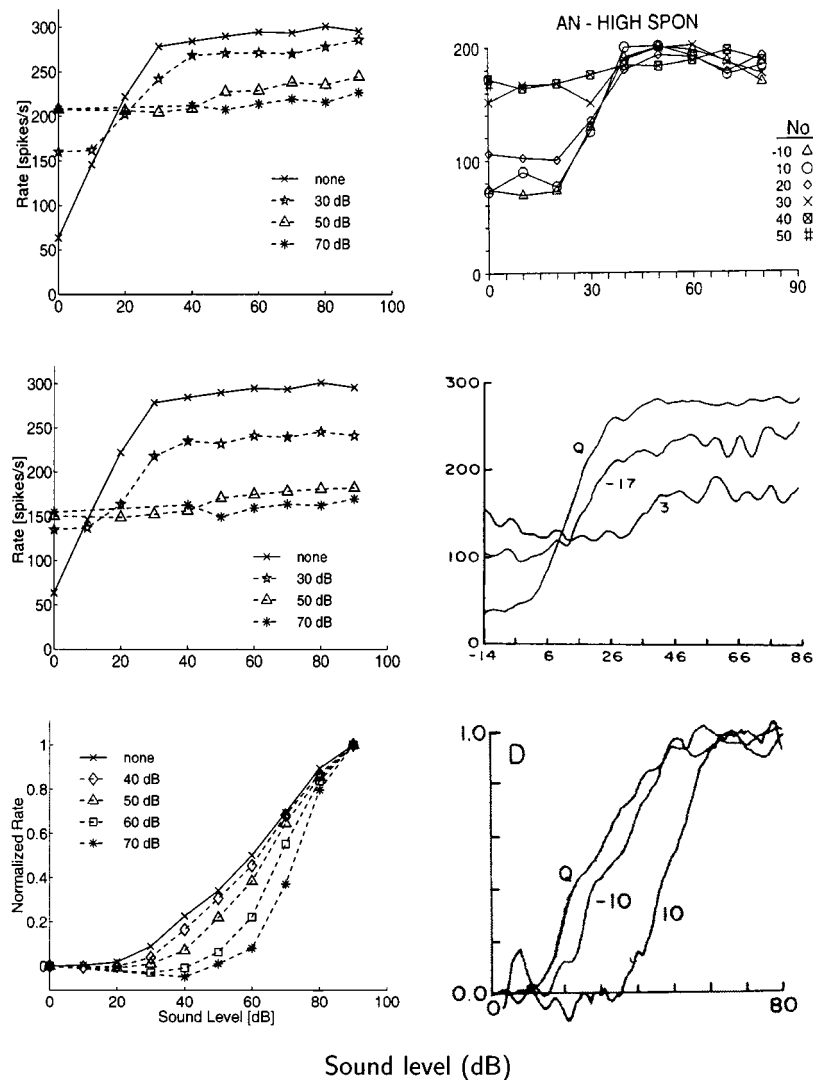


FIG. 17. Masked rate-intensity curves (MRC) of AN fibers in noisy background. Top: ANH fibers. Middle: ANH fibers with adaptation. Bottom: normalized MRC of ANI fibers. Right: Simulation results for different noise levels. Left: Physiological recordings. Top: from Rhode and Greenberg, 1994. Middle and bottom: From Costalupes *et al.*, 1984.



shift of the masked rate-intensity curve toward higher tone intensities. To better visualize this shift, the MRC curves were normalized, using the response rates at 0 and 90 dB as reference points, as done similarly in Costalupes, Young, and Gibson (1984)

$$r'(i) = \frac{r(i) - r(0 \text{ dB})}{r(90 \text{ dB}) - r(0 \text{ dB})}.$$

Normalized MRCs are shown in the bottom row of Fig. 17. For ANL fibers, the horizontal shift in masked rate-intensity curves is accompanied by an increased slope. Simultaneously gated tone and noise (not shown) produces the same effect for both ANL and ANH fibers.

### III. DISCUSSION

#### A. General summary and conclusion

We have presented a simple, easy to implement, nonlinear active model of the cochlea with distributed feedback. It produces realistic filtering responses (tuning curves) and accurately models compression and suppression phenomena. Adaptation observed at the IHC-AN synapse level is also well represented. Previous versions of the model were presented in Eriksson and Robert, 1998. The model is currently used in studying responses to complex stimuli in models of the auditory nerve and cochlear nucleus neurons, as well as in speech analysis.

#### B. Comments on parametrization

The range of parameters over which the model is a reasonable approximation to early auditory processing is given here. We qualitatively describe the role of three main model parameters, namely the filter  $G1$  cutoff ratio, the Gaussian summation, and the basal integration.

As a reminder, (1) the low-pass  $G1$  has a cutoff frequency below the section's CF, therefore setting the feedback threshold higher than the excitation threshold and shaping the suppression area by reducing the suppressive effect of higher than CF frequencies, (2) the Gaussian function sums the activity over channels neighboring the CF section, and (3) basal integration integrates activity over basal sections (to section at CF), which increases high-side suppression.

Any increase of the ratio  $w_n/w'_n$  results in even higher suppression threshold and strong suppression asymmetry emphasizing low-side suppression. At values below 1, high-side suppression becomes too large and suppression threshold too low. By substantially increasing its value, high-side suppression eventually disappears because of a very large suppression threshold in the feedback loop. With this ratio set in the range 1.1 to 1.5, one can obtain proper responses by adjusting the factor  $r$  in Eq. (5) and the basal integration width.

The Gaussian summation mainly influences the responses when broadband stimuli are presented. For example, integrating activity in lateral sections, in the case of tones with large SPL in the presence of a low-SPL noise background, allows a moderate suppression of the tone that would not occur otherwise because of SPL differences. In-

creasing the Gaussian width will help broader (wide-band) signals contribute to suppression. For pure-tone stimuli, this parameter has no effect.

Basal integration, whose width determines the level to which high-side suppression is increased, balances the effect of  $G1$  in this suppression area but plays no role in the low-side suppression region. The number of basal sections was set to 40, an empirical limit over which no noticeable difference in responses was observed. It could be reduced with proper adjustment of the other parameters. For pure-tone stimuli, this parameter has little or no effect.

It could be argued that  $G1$  and the basal integration could be combined in a single function that would shape the suppression response area as desired. However, this would require a more complex mathematical function, including asymmetry and gain control, capable of generalizing to many different stimulus waveforms and levels. We believe our approach is easier to implement and finds justification in the physiological OHCs.

The setting of the model parameters is not straightforward as a result of combining influences of individual parameter effects. We suggest that the reader adjust the basal integration to 40 neighboring sections and set the  $G1$  filter parameter next. By modifying the loop factor  $r$  [Eq. (5)], one obtains the desired compression and suppression behavior.

#### C. Model responses

The model tuning curves are realistic and in accordance with physiological data. The physiologically observed CF shift as stimulus level increases is represented in the model responses thanks to the cascade of two APGF filters in each BM section. Simulation results showed the model's compressive behavior; compression factors of up to 40 dB were simulated.

The suppression phenomenon is very well represented by our model, due to the feedback comprising (1) a saturating function, as proposed two decades ago by Zwicker (1979), (2) a low-pass filter  $G1$ , corresponding to the intuitive reasoning that the area of the BM where positive feedback is strongest lies just basally from the point of maximal response and, hence, that this is where the saturation threshold is lowest, and (3) a distributed form of feedback that has improved suppression modeling, particularly to broadband signals such as noise. This makes the model quite attractive for simulating auditory responses to complex sounds such as speech.

##### 1. Responses in background noise

The underlying mechanics of the BM and IHC/AN models accounts for most of the effects produced by tone/noise combinations observed in the auditory nerve. Short-term adaptation is accounted for by Meddis's inner hair cell model (Meddis, 1986, 1988). Long-term adaptation was modeled with an *ad hoc* method. In broadband noise background, the reduction in BM displacement amplitude in response to a pure tone was accurately simulated, in agreement with the shift in AN fibers' operating range reported in Costalupes *et al.*, 1984.

The model mimics the fact that the activity of AN fibers resulting from stimulation by broadband noise can be suppressed by pure tones both above and below the fiber's CF. The simulations of AN fibers' responses can be compared with results from Rhode and Greenberg (1992) and Costalupes and colleagues (1984, 1985). In the former study, MRAs had suppression regions similar to those obtained here, but showed little sign of suppression of pure tones by the noise. The studies conducted by Costalupes and colleagues, on the other hand, showed that the dynamic range of AN fibers was shifted toward higher intensities by the presence of background noise, and attributed this effect to two-tone suppression in the cochlea. Like Rhode and Greenberg (1992), Palmer and Evans (1982) and Geisler and Sinex (1980) also found less evidence for suppression of tonal stimuli when background noise was present. The differences could be due to variations in experimental methodologies.

Our model reproduces clear shifts in operating range for low SR fibers, while the response of the high-SR fiber saturates at moderate noise levels. This could only be achieved with a distributed form of feedback. To our knowledge, no attempts have been made to simulate this behavior. Our experience with different models indicated that they do not accurately reproduce this phenomenon.

## 2. Responses to notch noise

In their experiments on responses to notches in the auditory-nerve fibers, Poon and Brugge (1993) use a single notch width (within the filter's response area) and a triangular shape. They found that the AN fibers can encode notches well by a decrease in firing rates, relative to the response to a flat broadband noise. Our simulation results, in conditions similar to theirs, are in general agreement with their findings. However, for notch widths greater than 2 kHz (the bandwidth of the considered BM filter is approximately 1 kHz), our model exhibits suppression (relative to the response of the broadband noise at notch floor level) and therefore non-monotonic responses are observed, as shown in Fig. 12. This is not surprising, since for large notch widths the conditions are similar to those of suppression of noise by a pure tone. The observed suppression is relatively small (less than 2 dB) compared to that found in results from two-tone suppression simulations, shown in Fig. 9. Suppression is always accompanied by a broadening of the filter's bandwidth that compensates, in the case of broadband stimuli, the response decrease. Therefore, the output level cannot be derived directly and strongly depends on the model parameters and the relation between  $H$  and  $Q$  in the APGF. We are not aware of physiological recordings, for similar conditions, that could confirm or invalidate our simulation results.

## D. Comparison to other models

Because of its inherently nonlinear structure, our model differs from linear implementations (Flanagan, 1962; Patterson *et al.*, 1988; Jenison, 1991) and allows better representation of nonlinear behaviors such as compression and suppression. More recent models also make use of nonlinearities (Goldstein, 1990; Kates, 1991; Carney, 1993; Kanis and de

Boer, 1994; Giguère and Wooland, 1994; Patterson *et al.*, 1995), but the feedback effect is restricted to local BM sections, limiting for example the suppression modeling. Most models simulate the compression behavior using either the saturating feedback approach or by including analytic functions that mimic this compressive property (Goldstein, 1990). Furthermore, using the all-pole gamma-tone filter (APGF) we greatly simplify the filtering equations, allowing an easy control of its damping parameter (quality factor) while ensuring realistic filter shapes. The IHC-AN synapse model was a slightly modified version of the model introduced by Meddis (1986, 1988).

## E. Limitations

The choice of modeling the cochlea as a set of parallel filters instead of a cascade design can be debated. We settled for a parallel implementation because we could adjust each BM section more independently. Suppression phenomena is accurately modeled, but this could be improved as the low-pass filter in the feedback process hinders a fast suppression effect that seems to exist (Ruggero *et al.*, 1992). Also, in addition to the amplitude modulation of the CF response, a low-frequency suppressor also produces time-related phase modulation: phase leads in suppression response were observed, this lead increasing with the strength of amplitude modulation (Geisler and Nuttall, 1997). We are presently studying the phases of the response to verify this aspect.

## ACKNOWLEDGMENTS

We would like to thank M. Lyon for providing us with some MATLAB routines. This research was supported by the Swiss National Scientific Foundation, Grant Number 2150-045689.95.

- Aertsen, A., and Johannesma, P. (1980). "Spectro-temporal receptive fields of auditory neurons in the grassfrog. I. Characterization of tonal and natural stimuli," *Biol. Cybern.* **38**, 223–234.
- Cai, Y., and Geisler, D. (1996). "Suppression in auditory-nerve fiber of cats using low-side suppressors. III. Model results," *Hearing Res.* **96**, 126–140.
- Carney, L. H. (1993). "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am.* **93**, 401–417.
- Cheatham, M., and Dallos, P. (1990). "Comparison of low- and high-side two-tone suppression in inner hair cell and organ of Corti responses," *Hearing Res.* **50**, 193–210.
- Costalupes, J. A. (1985). "Representation of tones in noise in the responses of auditory nerve fibers in cats," *J. Neurosci.* **5**, 3261–3269.
- Costalupes, J. A., Young, E. D., and Gibson, D. J. (1984). "Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat," *J. Neurophysiol.* **51**, 1326–1344.
- Dallos, P. (1985). "Response characteristics of mammalian cochlear hair cells," *J. Neurosci.* **5**, 1591–1608.
- Dallos, P. (1986). "Neurobiology of cochlear inner and outer hair cells: Intracellular recordings," *Hearing Res.* **22**, 185–198.
- de Boer, E. (1975). "Synthetic whole-nerve action potential for the cat," *J. Acoust. Soc. Am.* **58**, 1030–1045.
- Delgutte, B. (1990). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," *Hearing Res.* **49**, 225–246.
- Eriksson, J. L., and Robert, A. (1998). "A simple nonlinear active cochlear model with distributed feedback," in *NOLTA 98 Proceedings*.
- Flanagan, J. (1962). "Computer simulation of basilar membrane motion," in *Fourth International Congress on Acoustics*.

- Geisler, C. D. (1990). "Evidence for expansive power functions in the generation of the discharges of 'low- and medium-spontaneous' auditory-nerve fibers," *Hearing Res.* **44**, 1–12.
- Geisler, C. D., and Sinex, D. G. (1980). "Responses of primary auditory fibers to combined noise and tonal stimuli," *Hearing Res.* **3**, 317–334.
- Geisler, D. C., and Greenberg, S. (1986). "A two-stage nonlinear cochlear model possesses automatic gain control," *J. Acoust. Soc. Am.* **80**, 1359–1363.
- Geisler, D. C., and Nuttall, A. (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.
- Giguère, C., and Wooland, P. C. (1994). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," *J. Acoust. Soc. Am.* **95**, 331–342.
- Goldstein, J. L. (1990). "Modeling rapid waveform compression on the basilar membrane as multiple bandpass nonlinearity filtering," *Hearing Res.* **49**, 39–60.
- Goldstein, J. L. (1995). "Relations among compression, suppression, and combinations tones in mechanical responses of the basilar membrane: data and MBPNL model," *Hearing Res.* **89**, 52–68.
- Greenwood, D. D. (1961). "Critical bandwidth and the frequency coordinates of the basilar membrane," *J. Acoust. Soc. Am.* **33**, 1344–1356.
- Greenwood, D. D. (1990). "A cochlea frequency-position function for several species-29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Hewitt, M., and Meddis, R. (1991). "An evaluation of eight computer models of mammalian inner hair-cell function," *J. Acoust. Soc. Am.* **90**, 904–917.
- Hudspeth, A. (1997). "Mechanical amplification of stimuli by hair cells," *Curr. Opin. Neurobiol.* **7**, 480–486.
- Jenison, R. L. (1991). "A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers," *J. Acoust. Soc. Am.* **90**, 773–786.
- Kanis, L.-J., and de Boer, E. (1994). "Two-tone suppression in a locally active nonlinear model of the cochlea," *J. Acoust. Soc. Am.* **96**, 2156–2165.
- Kates, J. M. (1991). "A time-domain digital cochlear model," *IEEE Trans. Signal Process.* **39**, 2573–2592.
- Liberman, M. (1982). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," *J. Acoust. Soc. Am.* **72**, 1441–1449.
- Lyon, R. F. (1996). "The all-pole gammatone filter and auditory model," in *Forum Acusticum, April '96, Antwerp, Belgium*.
- Lyon, R. F. (1997). "All-pole modes of auditory filtering," in *Diversity in Auditory Mechanics*, edited by Lewis *et al.* (World Scientific, Singapore).
- Meddis, R. (1986). "Simulation of mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.* **79**, 702–711.
- Meddis, R. (1988). "Simulation of mechanical to auditory-neural transduction: Further studies," *J. Acoust. Soc. Am.* **83**, 1056–1063.
- Mountain, D. C., and Hubbard, A. E. (1996). "Computational analysis of hair cell and auditory nerve processes," in *Auditory Computation*, edited by A. N. Popper and R. R. Fay (Springer, New York), Chap. 4.
- Palmer, A. R., and Evans, E. F. (1982). "Intensity coding in the auditory periphery of the cat: Responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise," *Hearing Res.* **7**, 305–323.
- Patterson, R. D., Allerhand, M. H., and Giguère, C. (1995). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.* **98**, 1890–1894.
- Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988). "Spiral VOS Final Report-Part A: The Auditory Filterbank," Tech. Report MRC Applied Psychology Unit, Cambridge, England.
- Patuzzi, R. B., Yates, G. K. M., and Johnstone, B. (1989). "Changes in cochlear microphonic and neural sensitivity produced by acoustic trauma," *Hearing Res.* **39**, 189–202.
- Poon, P. W., and Brugge, J. F. (1993). "Sensitivity of auditory nerve fibers to spectral notches," *J. Neurophysiol.* **70**, 655–665.
- Rhode, W. S., and Greenberg, S. (1992). "Physiology of the cochlear nuclei," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, New York), Chap. 3.
- Rhode, W. S., and Greenberg, S. (1994). "Encoding of amplitude modulation in the cochlear nucleus of the cat," *J. Neurophysiol.* **71**, 1797–1825.
- Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (1992). "Basilar membrane responses to two-tone and broadband stimuli," *Philos. Trans. R. Soc. Lond.* **336**, 307–315.
- Russell, I., Cody, A., and Richardson, G. (1986). "The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown in vitro," *Hearing Res.* **22**, 199–216.
- Sokolowski, B. H., Sachs, M. B., and Goldstein, J. L. (1989). "Auditory nerve rate-level functions for two-tone stimuli: Possible relation to basilar membrane nonlinearity," *Hearing Res.* **41**, 115–124.
- Temchin, A., Rich, N., and Ruggero, M. (1997). "Low-frequency suppression of auditory nerve responses to characteristic frequency tones," *Hearing Res.* **113**, 29–56.
- Zwicker, E. (1979). "A model describing nonlinearities in hearing by active process with saturation," *Biol. Cybern.* **35**, 243–250.

# The representation of pure tones and noise in a model of cochlear nucleus neurons

Jan L. Eriksson<sup>a)</sup>

*Institut de Physiologie, Université de Lausanne Rue du Bugnon 7, 1005 Lausanne, Switzerland*

Arnaud Robert<sup>b)</sup>

*CIRC Group, Swiss Federal Institute of Technology–Lausanne, 1015 Lausanne, Switzerland*

(Received 6 April 1998; revised 3 May 1999; accepted 7 June 1999)

The limited dynamic range of the majority of auditory-nerve fibers represents a difficulty in accounting for normal hearing capabilities over the known psychoacoustic intensity range. The presence of noise is an additional complication because it will tend to saturate these fibers, thereby considerably reducing their dynamic range, i.e., the range of mean firing rates. In this study, simulations involving a model of auditory nerve and cochlear nucleus neurons were conducted using pure-tone stimuli in the presence of noise. The main focus is on the role of inhibition in regulating the activity of cells, improving their capability to represent signals in background noise. This concerns in particular those inhibitory neurons that receive input from a wide range of auditory-nerve fibers and respond with an onset chopper pattern. A detailed model of stellate cells is used. It allows several parameters such as the number, location, and strength of inputs to be manipulated. The first part of this paper presents the model and its responses to pure-tone and noise stimuli presented separately. The model's capacity to generalize to tone/noise combinations is then tested. Responses to these stimuli are found to be qualitatively similar to neurophysiological findings. Model neurons exhibit appropriate shifts in their rate-level functions and their responses are inhibited or suppressed by tones outside their characteristic frequency. The model stellate cell is also found to display many of the temporal patterns reported in electrophysiological studies as a result of appropriate settings of certain parameters. Therefore, the model is sufficient to account for a large number of findings and should serve as a basis for predicting responses to novel stimuli, or as a building block for modeling larger networks. © 1999 Acoustical Society of America. [S0001-4966(99)04709-8]

PACS numbers: 43.64.Bt, 43.64.Qh [RDF]

## INTRODUCTION

The cochlear nucleus (CN) is the part of the brain that receives input from the auditory nerve. It is characterized by a large variety of cells, distinguishable both physiologically and morphologically (see, e.g., Rhode and Greenberg, 1992; Cant, 1992). It is very likely that this variety reflects a basic requirement for auditory signal processing since it is encountered in all species studied so far. The functional role of these multiple representations is still under investigation.

This work is part of a series of investigations involving simulations of the auditory pathway to better understand how complex sounds are represented. In a previous paper (Robert and Eriksson, 1999) a model of the auditory periphery was proposed whose output (the auditory-nerve fibers' action potentials) in response to tones in background noise corresponds to that observed in electrophysiological recordings. This paper's main focus is on the spectral representation of a subset of cells in the cochlear nucleus.

In the auditory nerve, the majority of fibers—those with high spontaneous rates and low threshold—saturates at low to medium intensities. Consequently, in these fibers, spectral features are lost at higher sound levels (Sachs and Young,

1979; Blackburn and Sachs, 1989). This led earlier investigators to reconsider the notion that sound features are represented by the rate of activity in auditory-nerve fibers innervating different parts of the basilar membrane. However, it is now well established that a certain proportion of fibers, 15 percent to 30 percent, have greater dynamic ranges and saturate only at high sound levels, if at all. A robust rate-place representation of spectral features could be achieved if input rates on a target neuron were weighed such that the neuron would “listen” to high-spontaneous rate fibers at low sound levels and medium/low-spontaneous rate fibers at higher sound levels (Winslow, Barta, and Sachs, 1987). Inhibitory circuits are likely to play a fundamental role in any such mechanism. Whether an invariant rate/place representation is achieved primarily in the cochlear nucleus or progressively in consecutive stages is unclear, although there is evidence that indicates that spectral features are refined throughout the ascending auditory pathway (see, e.g., Zhang, Suga, and Yan, 1997; Schreiner and Langer, 1997).

The presence of broadband noise affects responses of cells of the auditory system to stimuli such as pure tones or vowels in sometimes complex ways. Most studies so far have focused on its effect on auditory-nerve responses (e.g., Geisler and Sinex, 1980; Palmer and Evans, 1982; Costaloupes, 1985). There have also been studies done in higher brain areas (e.g., Gibson, Young, and Costaloupes, 1985;

<sup>a)</sup>Electronic mail: Jan.Eriksson@iphysiol.unil.ch

<sup>b)</sup>Electronic mail: Arnaud.Robert@epfl.ch



Phillips and Cynader, 1985; Rees and Palmer, 1988; Rhode and Greenberg, 1994). Globally, it can be said that rate–place representation of sounds in background noise is severely compromised in the auditory nerve, particularly in the high-spontaneous rate fibers which saturate at low stimulus levels. In cochlear nucleus neurons, as in auditory-nerve fibers, there is a compression of the range of firing rates (the dynamic range) that the cell is capable of producing and often a shift toward higher intensities. Finally, in many cases, lateral interactions beginning in the cochlea and extending into the higher brain areas, in the form of inhibitory circuits, are “exposed” when noise is added to, for example, a pure tone (Rhode and Greenberg, 1994). These mechanisms are likely to play a crucial role in improving the signal-to-noise ratio.

Neurons of the cochlear nucleus are often characterized according to two classification schemes, the poststimulus time histogram (PSTH) scheme and the excitatory–inhibitory response area (EIRA) scheme (see, e.g., Rhode and Greenberg, 1992). The EIRA scheme takes the average rate of firing in response to pure tones of different frequencies and intensities into account. Type I responses are those found in auditory-nerve fibers and primary-like neurons. Neurons with low-spontaneous rate and weak or no response to noise are labeled as type II. Type III neurons are spontaneously active and exhibit inhibitory sidebands, and type I/III are similar but with little or no spontaneous activity. Finally, type IV and V neurons are found primarily in the dorsal cochlear nucleus. These have nonmonotonic response curves and are always inhibited at medium/high intensities. The PSTH scheme characterizes the temporal response of the cell and generally includes the interspike interval histogram and the coefficient of variation. The main categories found in the cochlear nucleus cells are the primary-like, chopper, onset, and pauser/buildup patterns.

The type of output neuron which is the main focus of this study is the multipolar, or stellate, cell of the ventral cochlear nucleus. Axons of stellate cells of the cochlear nuclei form a direct (monosynaptic) pathway to the inferior colliculus (e.g., Adams, 1979) and are therefore prime candidates for conveying signal information to higher brain centers. The model features two sets of inhibitory interneurons that have been mentioned by various authors as potential sources of inhibition in the cochlear nucleus. The synaptic strength of the inhibitory-to-principal neuron connections constitute the main variable of the simulation experiments.

The main aim of this study has been to try to reproduce several known response characteristics of certain cochlear nucleus neurons. Since there is now a large body of physiological data available with which comparisons can be made, it is hoped that this can provide useful constraints for models. Unlike previous modeling studies, this one takes a more general approach by considering several aspects (the EIRA/PSTH schemes) in parallel.

This paper is divided into several parts. After describing the model and its component parts, the standard features of the different model neurons are presented in the form of response areas and rate-level curves. The next section examines how these neurons behave in a tone/noise context. Here

also, masked response areas (MRAs) and rate level curves (MRCs) are used as representations to allow comparison with physiological data. The two main sources used as references (Gibson, Young, and Costaloupes, 1985; Rhode and Greenberg, 1994) differ in the experimental manipulation used (continuous noise vs simultaneous noise/stimulus onset). In the first, the emphasis is on the shift in rate-level functions of auditory nerve (AN) and cochlear nucleus units produced by a background noise. The main conclusion is that the dynamic range adjustment seen in CN neurons is similar to that of AN fibers. This shift is of the order of 1 decibel per decibel increase in noise. In the second, the emphasis is more on lateral suppression. There it is argued that much of this suppression and the accompanying dynamic range shift originate in the cochlear nucleus.

The previous results are extended in the next section, in which the model stellate cells are examined in terms of their capacity to reliably transmit information in the presence of background noise using signal-detection methods. As has been shown experimentally, ventral cochlear nucleus (VCN) chopper units can enhance spectral contrasts, relative to the auditory nerve. In their work on vowel representation in the auditory nerve and VCN, May and colleagues (see, e.g., May *et al.*, 1996; May, Prell, and Sachs, 1998) used the same method to show that this is also true in a statistical sense.

The next section of the paper examines the model’s response to a different type of stimulus. Notch noise has been used to study responses of AN fibers and, more extensively, for dorsal cochlear nucleus (DCN) neurons. There is now evidence that the DCN and the principal cells of the DCN in particular receive excitatory input from the cochlear nucleus itself (Zhang and Oertel, 1994; Doucet and Ryugo, 1997). Hence, the model can serve to make predictions regarding this input to the DCN.

Finally, the PSTH patterns that the model generates as a result of different connection properties are presented. Because stellate cells are known to exhibit different response patterns, it is of interest to understand how these may be generated, particularly with regard to the role of inhibition. As in previous experimental reports (Blackburn and Sachs, 1989, 1992) and modeling studies (Banks and Sachs, 1991; Hewitt and Meddis, 1993), regularity analysis serves as the main analytic tool. Stellate cells exhibit a *chopper* type of activity when stimulated, with a pattern that is in part dependent on the neuron itself and in part dependent on the stimulus condition. The modeling studies have shown how certain factors such as synaptic input can influence the response patterns. This section shows how the specific components used in the model used for this study can explain some of these features, which complements the previous studies.

## I. COMPOSITE MODEL

The composite model consists of an auditory periphery part and a cochlear nucleus part. In the auditory periphery [Fig. 1(A)], the outer/middle ear is represented by a linear bandpass filter. Following this, an array of filters simulates the activity along the basilar membrane (BM) in response to the sound. Data from the cat have been used to set parameters such as filter bandwidth and the distribution of “best”

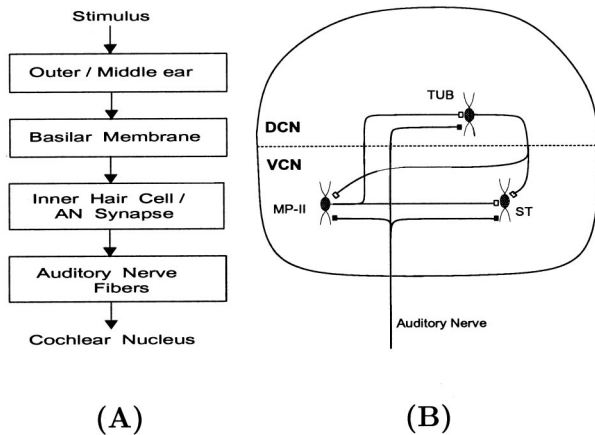


FIG. 1. Schematic diagram of the auditory periphery (A) and cochlear nucleus (B) models. (B) represents connectivity in a single isofrequency sheet. Filled and open squares represent excitatory and inhibitory synapses, respectively.

frequencies. The output of these filters is transformed into spike discharges by an inner hair cell (IHC) and auditory-nerve (AN) model.

The different cochlear nucleus (CN) neuron types modeled and their interconnections are shown in Fig. 1(B). The principal neurons that are simulated are the stellate cells (ST) of the ventral cochlear nucleus (VCN), which have been associated with the chopper response characteristics observed in electrophysiological recordings, and probably correspond to the multipolar type I cells in the Nissl preparation (Cant, 1981; Rouiller and Ryugo, 1984; Smith and Rhode, 1989). Stellate cells are known to receive inhibitory input (Smith and Rhode, 1989; Liberman, 1991, 1993). It is probable that collaterals of tuberculoventral cells (TUB) in the deep DCN contact cells in both the ventral and dorsal cochlear nucleus (Wickesberg and Oertel, 1988; Voigt and Young, 1990). Another putative source of inhibitory input is the multipolar type II (MP-II) cells, which display an onset chopper-type response and appear to contain pleomorphic vesicles in the terminals of their axon collaterals (Smith and Rhode, 1989). Onset choppers are thought to receive input from a wide range of frequencies and respond vigorously to noise (Palmer, Jiang, and Marshall, 1996), and could therefore constitute the proposed wideband inhibitory input (WBI, see Nelken and Young, 1994) onto tuberculoventral and fusiform cells (Oertel *et al.*, 1990; Zhang and Oertel, 1993, 1994) as well as cells in the VCN (Smith and Rhode, 1989).

Model cells are labeled using morphological terms for presentation purposes. A survey of the existing literature makes it evident that these are not the only cells and interconnections postulated to be involved in these circuits. It is therefore useful to remember to consider these as representative of classes of neurons with similar physiological properties but perhaps belonging to different morphological categories. Although there is only a small number of reported studies in which cells recorded from *in vivo* have been identified by intracellular labeling, the prevailing hypotheses relating physiology and morphology found in the literature have been adhered to. To recapitulate: (1) MP-II cells display onset chopper PSTH patterns and have broad type I/III re-

TABLE I. The three types of AN fibers used in the simulations and basic characteristics for fibers with CF between 1 and 20 kHz (below 1 kHz, the threshold is raised because of outer/middle-ear filtering).

Type	Spont. rate	Threshold (dB)	Saturation (dB)
ANH-1	65	5	30
ANH-2	40	20	60
ANL	5	30	>90

sponse areas (RAs), (2) ST cells display chopper PSTHs and type I/III or III RAs, (3) TUB cells display chopper PSTHs and type II RAs. The neuron models and the connectivities between groups of neurons are described below.

## A. Stimuli

Simulation runs consisted of 100 ms of silence, followed by a signal during 100 ms, and finally 50 ms of silence again. The background noise had an onset at 0 ms, with a ramp of 40 ms. The broadband noise was low-pass filtered at 20 kHz. Noise *levels* refer to total energy except where otherwise indicated. Notch noise was generated from broadband noise, and filtered so as to obtain notch widths of 0.5, 1, 2, and 4 kHz centered arithmetically around 5 kHz. The floor of the notch was 30 dB below the sides.

Plots of neural discharge rate vs stimulus intensity (rate-intensity curves) were established for a pure tone of 5 kHz as well as for broadband noise. This was the chosen characteristic, or best, frequency for all the neurons shown in this paper.

## B. Auditory periphery

The auditory periphery model used for these simulations has been described in a previous paper (Robert and Eriksson, 1999). To summarize, the model consists of: (1) a passive filter which simulates the external and middle-ear transfer function, (2) a set of parallel sections each representing a location along the basilar membrane (BM), (3) and IHC-AN synapse model, and (4) an AN spike generation model.

For the simulations, the BM filterbank consisted of 120 filters between 200 and 20 000 Hz. The spacing of the filters along the frequency axis was determined by measurements from cat cochleas (Liberman, 1982; Greenwood, 1990). Parameters were selected to produce three sets of IHC/AN synapse with characteristics shown in Table I. Each IHC became the source of several AN fibers (four in our case) by using independent random generators for each synapse, producing nonstationary Poisson processes modified by a dead time simulating the absolute refractory period. The maximum number of fibers used was therefore  $120 \times 3 \times 4 = 1140$ . All AN action potentials transmitted to the CN model neurons were delayed by a fixed interval (1.7 ms).

In the tests that were made to simulate the long-term adaptation (of the order of a second or more) known to take place in the auditory nerve, the output of the IHC, which determines the probability of firing, is reduced by a factor calculated on the basis of mean firing rate prior to the signal onset. This factor decreases linearly from 1 for low rates

(less than 80 spikes/s) to 0.6 at the maximum rate (250 spikes/s). For example, a rate of 200 spikes/s is reduced to 152 spikes/s.

### C. Neuron model

A MacGregor model (MacGregor, 1987) is used as the basis for all cells modeled in this work. As in previous studies (Arle and Kim, 1991), modifications in the form of additional conductances were made to the basic model. The basic model is characterized by four variables: (i) a transmembrane potential, (ii) an after-spike hyperpolarization potassium ( $K^+$ ) conductance, (iii) a threshold, and (iv) an all-or-nothing spiking variable.

The actual neuron models used include a somatic part and a dendritic part. Such a configuration can be considered a standard for modeling CN neurons, particularly chopper units in which dendritic filtering is believed to be an important element. The detailed multicompartment model of Banks and Sachs (1991), for example, has been used to demonstrate the effect of the location of excitatory input on the spiking regularity of the neuron.

The membrane potential (m.p.) of the soma is governed by the following equation (interchanging subscripts  $s$  and  $d$  gives the m.p. for the dendrite compartment):

$$\tau_0 \dot{V}_s(t) = -V_s(t) + \sum_i G_i(t, V_s) [E_i - V_s(t)] + G_{s;d} [V_d(t) - V_s(t)], \quad (1)$$

where  $V_s(t)$  and  $V_d(t)$  are the membrane potentials of the soma and dendrite, respectively,  $\tau_0 = c_m/g_L$  is a time constant,  $G_{s;d}$  and  $G_{d;s}$  are the axial conductances between the two compartments,  $G_i(t, V)$ ,  $E_i$  are conductances (and corresponding reversal potential) reflecting synaptic input, delayed  $K^+$  rectifier and additional nonlinear voltage-dependent conductances. These are applied on a case-by-case basis and are described briefly below.

The above membrane equation is obtained after normalization by the leakage conductance  $g_L$  and replacing the actual membrane potential  $E_s(t)$  by  $E_s(t) = E_L + V_s(t)$ , where  $E_L$  is the resting potential. Because  $g_L$  depends on the surface area of the membrane, it will generally be different for the soma and dendrite compartments (hence the different axial conductances  $G_{s;d}$  and  $G_{d;s}$  for the soma and dendrite membrane equations). The parameter  $\rho (= g_d/g_s)$  is the ratio between the dendritic and soma areas. According to Manis (Manis, Marx, and White, 1996), values of  $\rho$  can show considerable variability even for a single cell type. An average value for stellate cells in the VCN is about 3. All other constants, including  $\tau_0$ , are equal in both compartments. The dendritic part of the neuron can be further subdivided into smaller compartments. In this case,  $G_{d;s}$  (and  $G_{d;d}$ , the conductance between two dendritic segments) needs to be adjusted to match the dendritic area represented.

An action potential is generated when the membrane potential  $V_s(t)$  crosses a fixed threshold  $\theta$ . Spiking activity is blocked for a fixed duration to simulate the spike duration and the absolute refractory period, which are not explicitly modeled.

TABLE II. List of parameter values for the four neuron models used in this study.

Parameter	MP-II	TUB	ST
$E_+, E_{Na}$ (mV)	60	60	60
$E_-, E_K$ (mV)	-10	-10	-10
$\tau_0$ (ms)	3	5	2
$\theta$ (mV)	12	12	10
$G_{s;d}$	2	5	5
$\rho$	3.3	2.5	2.5
abs refr (ms)	2	2	1.5
$G_{K_0}$	2.6	1.5	4.5
$\tau_{g_K}$ (ms)	1.5	1.5	1.2
$G_{K,A}$			3.25
$\tau_{G_{K,A}}$ (ms)			12
$d_{G_{K,A}}$ (mV)			-9
$c_{G_{K,A}}$ (mV)			-7
$m_{G_{K,A}}$			1
$\tau_{G_{K,A},B}$ (ms)			3
$d_{G_{K,A},B}$ (mV)			-2
$c_{G_{K,A},B}$ (mV)			2.5
$n_{G_{K,A}}$			4

Excitatory/inhibitory postsynaptic conductances and the after-spike hyperpolarization potassium conductance are modeled using similar linear differential equations. For the latter ( $G_K$  conductance), included in the soma compartment, the equation is

$$\dot{G}_K(t) = -G_K(t)/\tau_{g_K} + G_{K_0} \cdot S, \quad (2)$$

where  $\tau_{g_K}$  is the time constant,  $G_{K_0}$  the conductance strength, and  $S$  the spiking variable (0 or 1).

For the excitatory/inhibitory postsynaptic conductances  $G_+(t)$  and  $G_-(t)$ , the time constants were set to 1 and 6 milliseconds, respectively, and  $S$  represents a presynaptic spike.

In addition to these basic elements, the neuron model may include specific voltage-dependent conductances,  $G_i$ . Following Kim and colleagues (Kim *et al.*, 1994) the following equations were used to describe these voltage-dependent conductances:

$$G_i(V, t) = \overline{G}_i \cdot A_i^m(V, t) \cdot B_i^n(V, t), \quad (3)$$

$$A_i(V, \infty) = \frac{1}{1 + \exp\left[\frac{d_{A_i} - V}{c_{A_i}}\right]}, \quad (4)$$

$$\dot{A}_i(V, t) = [A_i(V, \infty) - A_i(V, t)]/\tau_{A_i}, \quad (5)$$

and similarly for  $B_i$ .

Index  $i$  denotes the conductance type.  $A$  corresponds to the activation and  $B$  to the inactivation components of the conductance. For noninactivating conductances,  $B$  is set to 1. All parameters for the different cell types are listed in Table II. All conductance values given in the text refer to normalized (by  $g_L$ ) conductances.

The equations are numerically solved by using the implicit (backward) Euler method. The integration step was 0.1 or 0.5 ms, depending on the temporal resolution required for the analysis of the responses (e.g., analysis of spike rates

does not require fine resolution). Comparisons between simulations involving different time steps produced near-identical responses, provided the excitatory conductance amplitudes (whose time constants are small) were adjusted to produce the same change in the membrane potential.

#### D. Interconnections

In the model, the connectivity between cell-type groups maintained a tonotopic arrangement, i.e., cells with similar best frequencies had the highest probability of being connected. A probabilistic connectivity was used, in which the probability of two cells of different groups being connected followed a Gaussian distribution whose width depended on the cell types connected. The probability of connection between two cells depended on their characteristic frequency (CFs) and is given by the equation

$$P = P_0 \exp(-\log(f_1/f_2)^2/\omega^2). \quad (6)$$

Values of  $P_0$  were adjusted to obtain different numbers of connections. In the subsequent sections,  $\omega$  is expressed in octaves.

## II. STANDARD RESPONSE PROPERTIES

### A. Inhibitory units: MP-II and TUB cells

Onset choppers, the physiological category most likely to correspond to MP-II cells, have been the subject of several recent investigations (Winter and Palmer, 1995; Jiang, Palmer, and Winter, 1996; Palmer, Jiang, and Marshall, 1996). These studies support the idea that this group of cells receives input from auditory-nerve fibers with a wide range of characteristic frequencies, although their tuning curves at low stimulus levels are as sharp as those of AN fibers. Furthermore, unlike AN fibers, the threshold to noise of these cells is about the same as that for pure tones. The time course of the response to pure tones consists of one or a few precisely timed spikes followed by a low rate of activity, at least for low/medium stimulus intensities. Mechanisms proposed for the lower activity after onset include depolarization block, coincidence detection, and delayed inhibition. Depolarization block may be a factor for certain types of onset cells (e.g., onset lockers) but not onset choppers which continue to fire after the initial spike, albeit at a reduced rate. The fact that cell bodies of type II multipolar cells appear to receive inhibitory input (Smith and Rhode, 1989) supports the inhibitory hypothesis. A rectifying ( $K^+$ ) conductance might also be proposed as a mechanism for the delayed rise in firing threshold, but intracellular recordings seem to indicate that these are not present in stellate cells (Oertel, Wu, and Hirsh, 1988).

The MP-II and TUB cell models were based on the two-compartment cell model with no additional voltage-dependent conductances. Both cell types received excitatory input on both dendrites and soma. Based on data from Liberman (Liberman, 1993), the MP-II cells received input on the soma from high-spontaneous fibers from a narrow range of CFs only. The synaptic strength  $G_{+0}$  was equal to  $2.0/n$ , where  $n=20$  (number of AN inputs). Dendrites received input from all three AN fibers types within a range of two

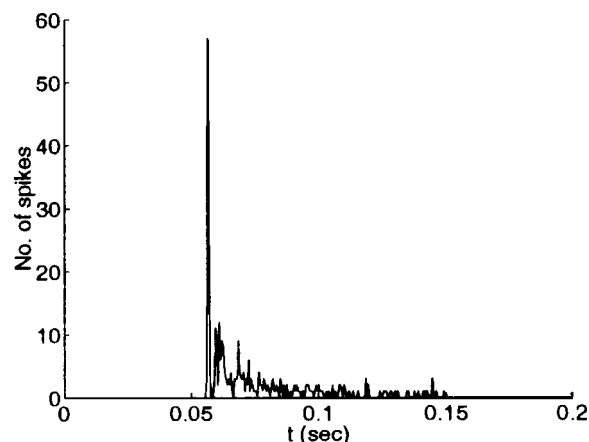


FIG. 2. Onset-chopper PSTH pattern in MP-II cell in response to a 5-kHz stimulus at 50 dB.

octaves above and below CF, and  $G_{+0}=2.0/n$ , where  $n = 60$ . One set of MP-II cells also received an inhibitory input from TUB cells ( $G_{-0}=2.3/n$ ,  $n=12$ ). The TUB cell model received excitatory input from AN fibers (range:  $\pm 0.1$  octaves,  $G_{+0}=2.3/n$ ,  $n=21$ ) and inhibitory input from the MP-II cells (range:  $\pm 0.4$  octaves,  $G_{-0}=1.5/n$ ,  $n=8$ ). In the model, the onset response of the MP-II cells is generated by the high-spontaneous rate fiber's large onset response, whose subsequent steady-state input does not carry the cell's membrane potential above threshold. Only when lateral excitation arrives at the soma does the firing rate rise. This occurs both for multiple-tone or wideband signals, as well as through spread of excitation in AN fibers as a pure-tone stimulus intensity increases.

Figure 2 shows a typical PSTH from an MP-II onset-chopper cell at medium stimulus intensity. It consists of a set of short-latency spikes followed by a low rate of activity. The onset is rapid and with little variability in first-spike latency. Rate-intensity curves and response areas for MP-II and TUB cells with CF of 5 kHz from each of the three groups are shown in Fig. 3. Onset rate was calculated during the first 10 ms of the response to a 5-kHz tone and the steady-state rate during the last 70 ms. Also shown is the steady-state response to the noise stimulus alone. The onset rate for MP-II (Fig. 3) cells rises steeply with stimulus intensity. The steady-state rate for both tone and noise stimuli increases linearly and does not saturate even at high intensities. The pure-tone response of the MP-II cell with inhibitory input has a firing rate which remains low for stimuli below 70 dB (Fig. 3, middle panels).

The response areas show the activity evoked by tones of varying frequency and intensity. In this case, the rate was calculated over the full stimulus duration (100 ms). The response area for the second MP-II cell illustrates the fact that the reduced activity due to inhibition is restricted to frequencies near the CF. Such response areas are sometimes observed in onset cells, including a category labeled onset lockers,  $O_L$  (see, e.g., Rhode and Smith, 1986; Rhode, 1991). Rate-intensity curves and response areas for TUB cells exhibit the easily recognized type II pattern of the EIRA scheme. These cells typically have nonmonotonic rate-intensity functions and respond little to noise.



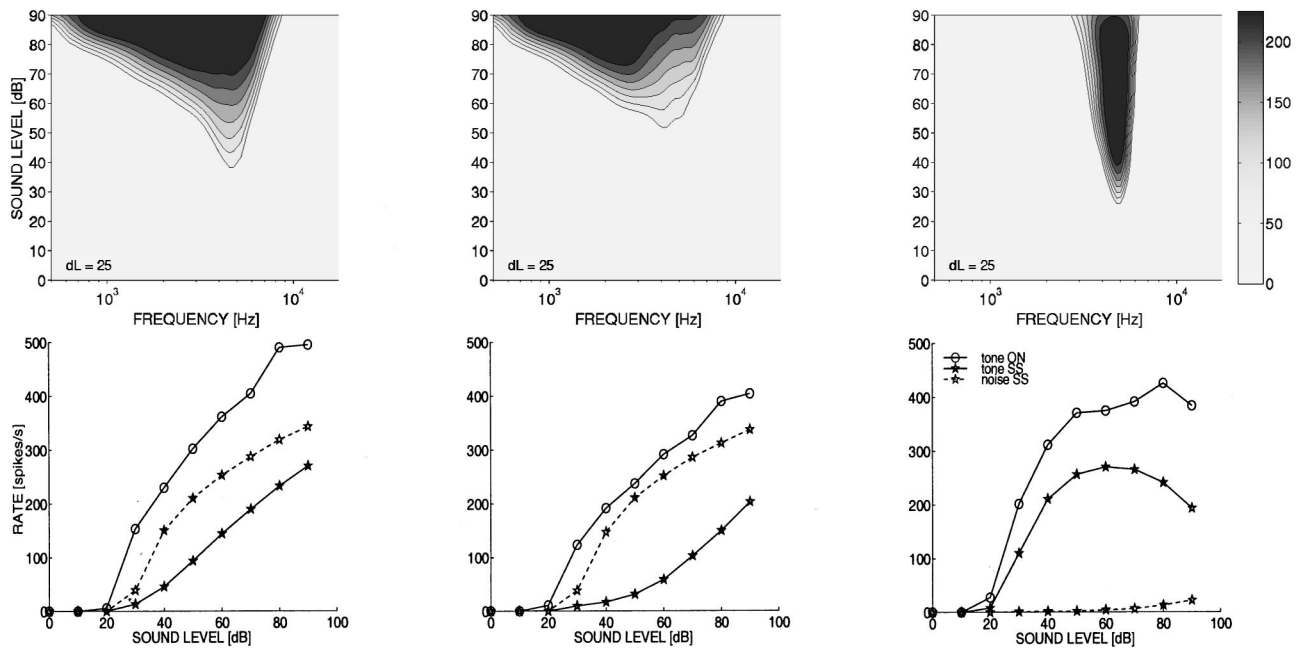


FIG. 3. Top: Response areas for the same cell types. Bottom: Rate-intensity curves for inhibitory units with characteristic frequency of 5 kHz. Onset (ON) and steady-state (SS) firing rates in response to pure tone and to noise (SS only). Left to right: MP-II cell, MP-II cell with inhibitory input, and TUB cell.

## B. Stellate cells

Responses of stellate cells are usually regarded as belonging in one of two basic categories: the regular sustained choppers ( $C_S$ ) and the more irregular transient choppers ( $C_T$ ) (see, e.g., Blackburn and Sachs, 1989). Most of these cells have thresholds and dynamic ranges similar to high-spontaneous AN fibers (Rhode and Smith, 1986; Rhode and Kettner, 1987). Their cell bodies tend to be sparsely innervated, although transient choppers, in particular, are believed to correspond to cells whose somata are innervated as well (Cant, 1981). Input to AVCN stellate cells from the auditory nerve appears to be biased towards medium/low-spontaneous rate fibers (Rouiller *et al.*, 1986; Liberman, 1991). This does make sense if one considers that most stellate cells are more sharply tuned than the majority of AN fibers. However, there must undoubtedly be some input from high-spontaneous rate fibers to account for their generally low thresholds. Stellate cells therefore appear to perform the selective listening function mentioned in the introduction. Inhibitory input, if present, is believed to occur at proximal dendrites and soma (Cant, 1981; Smith and Rhode, 1989), although it has recently been reported that distal dendrites may be contacted by terminals with pleomorphic vesicles as well (Josephson and Morest, 1998) which could suggest that these cells are capable of more complex signal processing.

Several authors have suggested that the highly regular “chopping” response pattern of stellate cells is due to a combination of the convergence of many nerve fibers and these cells’ intrinsic membrane properties, and a number of simulation experiments support this view (Banks and Sachs, 1991; Hewitt and Meddis, 1993). Banks and Sachs’s approach of using a compartmental model has been adopted, although MacGregor’s neuron model has been chosen instead of the computationally more demanding Hodgkin and Huxley model they use. The model neuron consisted of two

dendritic compartments and a soma. Comparisons of tuning sharpness between AN fibers and choppers ( $Q_{10}$  factors) indicate that these are very similar, in some cases slightly sharper, in others slightly less sharp (Rhode and Smith, 1986; Jiang, Palmer, and Winter, 1996). For the ST neurons, input from all three AN fiber groups spanned a range of  $\pm 0.1$  octaves relative to CF (5 kHz). Inhibitory input, when present, had a range of  $\pm 0.4$  octaves. The time constants chosen for the excitatory/inhibitory conductances, 1 and 6 milliseconds, respectively, resulted in excitatory postsynaptic potential/inhibitory postsynaptic potential (EPSP/IPSP) durations in line with data reported by Oertel (1983). Inputs from inhibitory synapses were delayed by 1.2 ms, with a jitter of  $\pm 0.5$  ms.

A voltage-dependent conductance was included in all compartments, following Manis’s (Manis, Marx, and White, 1996) observation that A-type currents are present in stellate cells. Manis’s data indicate that this conductance is not necessarily completely inactivated at the level of depolarization induced by the spontaneous activity of the AN fibers. One consequence is that there is an outward current which lowers the membrane potential and prevents or reduces the cell’s spontaneous activity. As the membrane potential is depolarized, this conductance is inactivated. The result is a steeper slope of the rate-intensity curve. Another consequence of including this conductance is that the depolarization caused by the strong AN input at stimulus onset is reduced, producing an output rate that shows little sign of adaptation through the stimulation.

The number of synapses, target location, and synaptic strength were systematically varied. Changing the first two parameters had little impact on rate measurements (provided synaptic strength was adjusted). For the results reported in

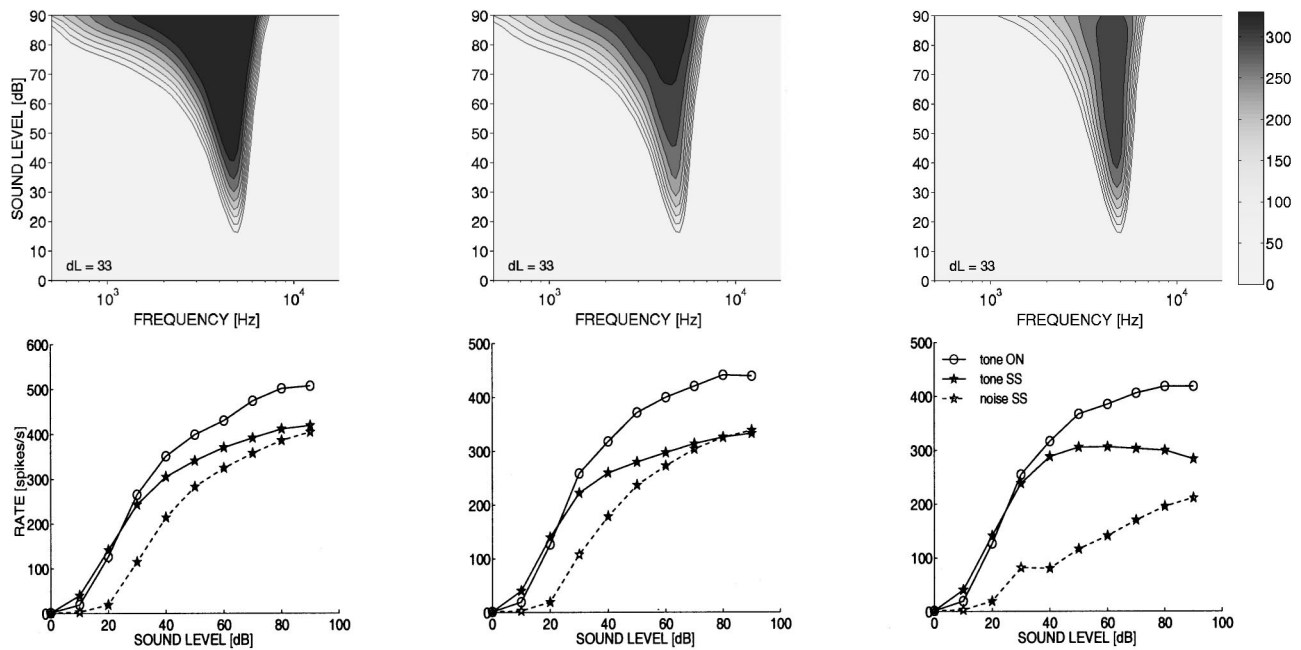


FIG. 4. Response areas for stellate cells. Left to right: No inhibition, TUB inhibitor input (weak), MP-II inhibitory input (strong). Bottom: Rate-intensity curves for the same cell types. Onset (ON) and steady-state (SS) firing rates in response to pure tone and to noise.

this paper, the ST cells fell into two categories: those cells which received excitatory input on dendrites only (type 1) and those with inputs on both dendrites and soma (type 2). Cells of type 1 received inputs from either 6 or 15 AN fibers (i.e., two, or five, respectively, from each of the three AN sets). The synaptic strength  $G_{+0}$  was equal to  $8.5/n$  (where  $n$ =number of AN inputs). With the 15-fiber configuration, a presynaptic spike produced a dendritic EPSP of approximately +7.5 mV, which was attenuated to +2.5 mV at the soma. Cells of type 2 received an equal number of AN inputs ( $n=15$ ) on both dendrite and soma compartments. For the type 2 ST cells, the synaptic strength for each compartment was equal to  $2.0/n$ . A presynaptic spike input to the soma produced an EPSP of approximately +2.4 mV, whereas input to the dendrites produced a +0.9 mV EPSP at the soma. Both of these ST cell types could receive inhibitory input. The number of inhibitory inputs from MP-II cells was 3 or 8 for the type 1 ST cells, and 3 for type 2. The synaptic strength  $G_{-0}$  was either weak ( $0.5/n$ ), moderate ( $1.0/n$ ), or strong ( $1.5/n$ ). For the 3 inhibitory inputs configuration, an inhibitory presynaptic spike at the soma produced a -0.6-mV IPSP. Type 1 and 2 ST cells had characteristics of regular and irregular choppers, respectively. The ST cells' PSTH patterns will be analyzed in detail in Sec. IV.

Rate-intensity curves and response areas for a stellate cell with different degrees of inhibitory input are shown in Fig. 4. In the left-most panel, no inhibition is present. Discharge rate increases monotonically and continues to increase due to input from low-spontaneous AN fibers. In the central panel, inhibitory input from TUB cells has been added. This has the effect of reducing the firing rate at medium-range pure-tone intensities, but has little/no effect on the noise rate-level functions. Increasing it further will reduce the response to the pure tone even more, and the rate-level functions will become nonmonotonic at medium-

range intensities. In the right panel, inhibition is from MP-II cells with the consequence that the response area is now more restricted to the area around the CF and saturation occurs at a lower level. For stronger inhibition, the rate-intensity curve can become nonmonotonic at high intensities and the response to noise is very limited.

### III. RESPONSES TO PURE TONES IN BACKGROUND NOISE

Having defined sets of parameters with which the different cells in the model reproduce patterns of activity in response to pure tones and noise similar to those observed in electrophysiological recordings, it is now possible to examine whether the model can generalize to stimuli involving pure-tone and background noise combinations.

The alteration in the dynamic range of a cell when a background noise is present is typically shown in a set of masked rate-intensity curves (MRC), a family of rate-intensity curves each with a different background noise level. A frequency-intensity response map with background noise, or masked response area (MRA), is similar to the excitatory-inhibitory response area, except that it is determined relative to a fixed noise background level (Rhode and Greenberg, 1994). Each point on these plots represents the change in mean firing rate when the pure-tone stimulus is added to the noise background.

Understanding the inhibitory cells' responses to complex sounds such as tone/noise combinations is obviously important since these condition the responses of most of the other cells in the CN. The MRA for an MP-II model neuron is shown in the top right of Fig. 5 next to the one of an onset-chopper unit recorded by Rhode and Greenberg (1994). Below, an  $O_C$  MRC from Rhode and Greenberg (1994) is shown next to that of the MP-II cell with inhibitory

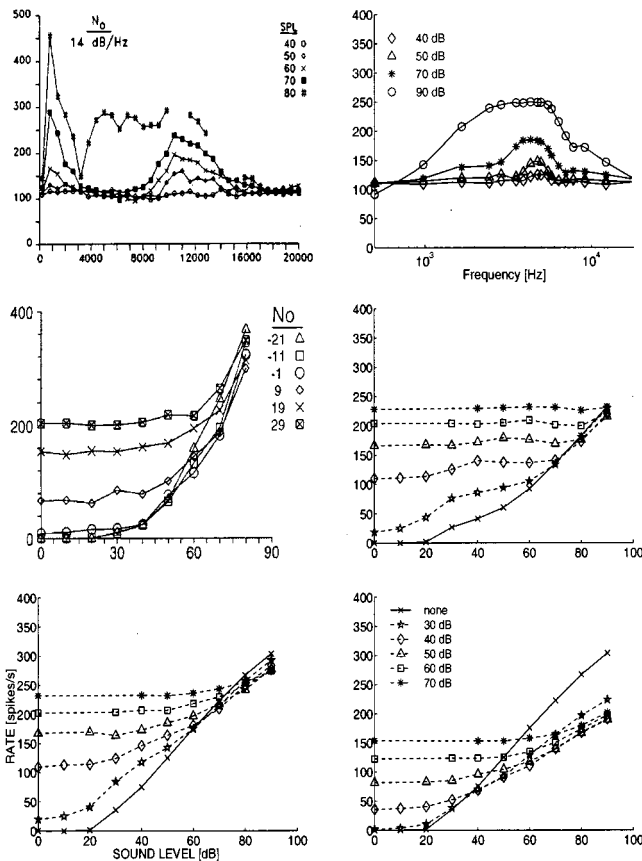


FIG. 5. Masked response areas and masked response curves for MP-II cells. Top row: MRA for  $O_C$  unit from Rhode and Greenberg (1994) and MP-II model neuron with fixed background noise level of 40 dB. Middle: MRC for  $O_C$  unit from Rhode and Greenberg (1994) and MP-II cell with inhibitory input. Bottom: MRCs for MP-II cell with synchronous and continuous noise.

input. At the bottom, the MRCs in conditions of synchronous and continuous background noise of an MP-II cell are displayed. The most distinctive feature here is that these cells show little sign of suppression. The strong response to noise that these cells exhibit often tends to completely mask their response to tonal stimuli over a large range of intensities. For the lower noise levels, the two inputs (tone and noise) are additive, as can be seen from the fact that the curves rise above the curve for the tone signal alone. In the continuous background noise condition, the response of the MP-II cell is reduced.

In contrast to MP-II cells, TUB cells' output is suppressed when noise is present, either in the form of a shift of the rate-intensity curves towards higher stimulus levels (Fig. 6) or a combination of a shift in dynamic range with a marked reduction in maximum firing rate. The continuous vs synchronized background noise cases did not differ much for TUB cells.

Responses of stellate cells to tone/noise combinations vary with the amount of inhibition they receive. In Fig. 7, masked response areas are displayed for ST cells with no inhibition, and two different inhibitory strengths, along with the MRA of a chopper unit recorded by Rhode and Greenberg (1994). The MRAs for ST cells exhibit sideband inhibi-

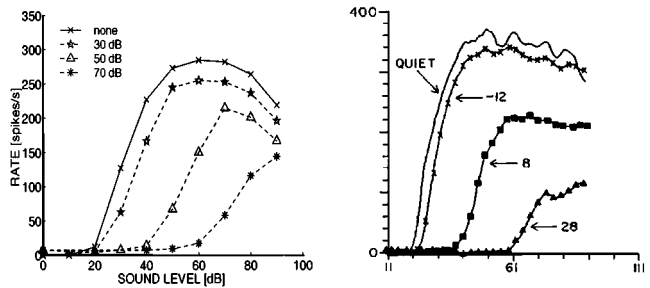


FIG. 6. Masked response curves for a TUB model cell and a DCN type II cell from Gibson *et al.* (1985), with tones in continuous background.

tory regions, but these are small unless there is additional inhibitory input.

In Fig. 8, masked rate-level curves are displayed along with units recorded by Rhode and Greenberg (1994) and Gibson *et al.* (1985). In all cells, there is a shift in the dynamic range towards higher intensities. The shift is already present in the AN fibers' responses (Robert and Eriksson, 1999), but is more evident when inhibitory input is added. The masked rate-intensity curves shown here differ in the baseline response to the noise stimulus (response at 0-dB tone intensity). When normalized rate-intensity curves (Gibson, Young, and Costaloupes, 1985; Robert and Eriksson, 1998) are plotted (not shown here), the shift is the same for all cells, approximately a 1-dB shift for every 1-dB increase in noise level. The figures on the right represent responses in continuous background noise; in the case of continuous noise, the reduced AN input was accompanied by a reduced inhibitory input from MP-II cells (Fig. 5, bottom right). For ST cells with moderate to strong inhibition, the curves are similar with the exception that the maximum rate decreases faster with noise level in the case of continuous noise.

#### IV. DISCRIMINATION CAPABILITY OF ST CELLS

It is instructive to consider responses to fixed-intensity pure tones in situations in which the background noise is

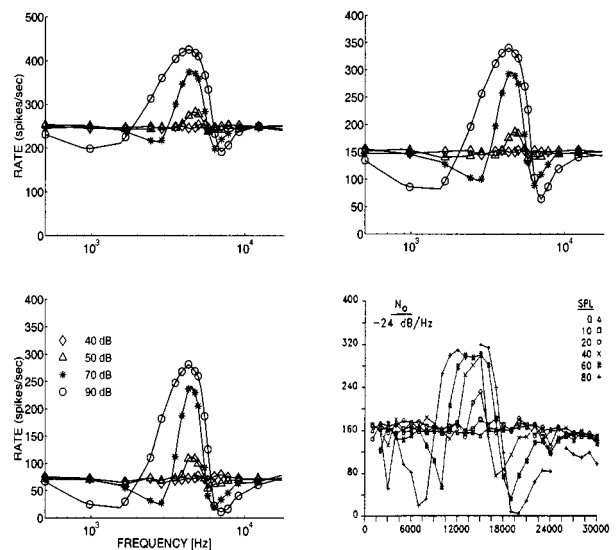


FIG. 7. Top: Masked response areas ST cells. Top row: ST with zero and moderate inhibition. Bottom row: ST with strong inhibition and MRA for a chopper unit from Rhode and Greenberg (1994).

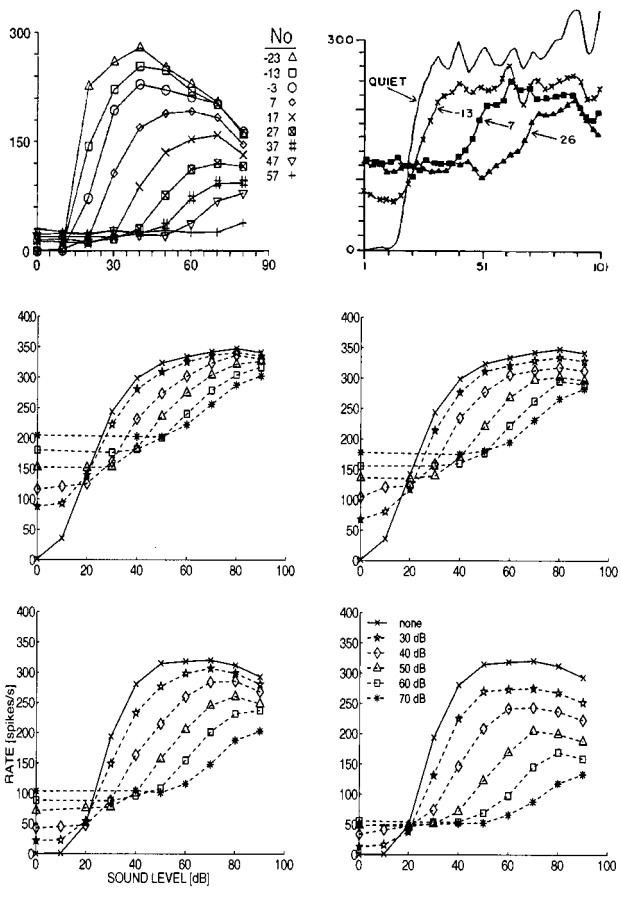


FIG. 8. Masked response curves for ST cells. Top row: MRCs for  $C_T$  unit from Rhode and Greenberg (1994) and PVCN unit from Gibson *et al.* (1985). Middle: MRCs for ST cells with moderate inhibition. Bottom: MRCs for ST cells with strong inhibition. Right column involves tones in continuous noise background.

varied. This gives an idea of how the representation of a stimulus is affected by decreasing the signal-to-noise ratio. A method for expressing a cell's discrimination capability based on its firing rate is to include the rate variability in the measurements. This relationship has been formalized by Green and Swets as a  $d'$  statistic (Green and Swets, 1966). Here,  $d' = (\mu(\text{tone+noise}) - \mu(\text{noise})) / \sqrt{\sigma(\text{tone+noise})^2 + \sigma(\text{noise})^2}$ , where  $\mu$  is the mean and  $\sigma$  the standard deviation in firing rates. To show how the representation of pure tones is modified by background noise, pure tones at a fixed intensity were combined with broadband noise, at signal-to-noise ratios (SNRs) ranging from +20 to -10 dB, and the responses were then recorded. Figure 9 shows the results obtained from two types of AN fibers and two stellate cells. The curves represent  $d'$  values for a 5-kHz tone at four tone intensities as a function of SNR. If  $d'$  values greater than +1 are regarded as significant, it can be seen that for SNRs less than +10 dB, significant encoding is limited to low tone intensities. The low-spontaneous fibers provide a reliable encoding of the stimulus over a wider range of SNRs for the tone intensities tested. In contrast to AN fibers, CN cells can exhibit significant  $d'$  values over the full range of SNRs:  $d'$  values for the ST cells are two to three times greater than any of the AN fibers. One observation that was made is that adding inhibitory input from TUB

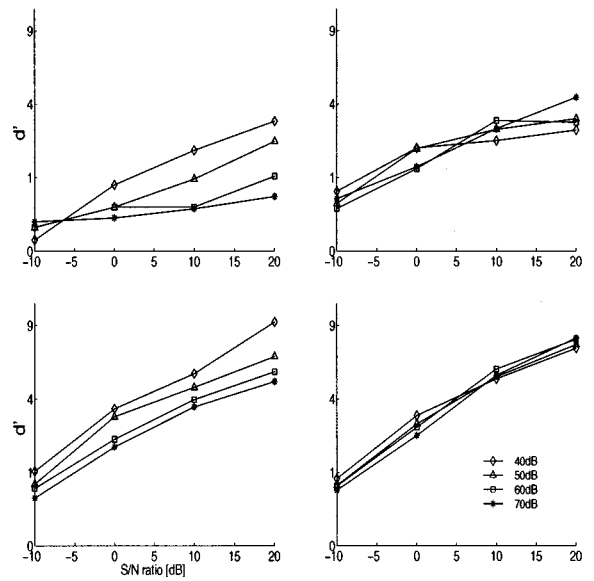


FIG. 9. Plots of  $d'$  values for a pure-tone stimulus of 5 kHz at four different intensities (40–70 dB). Top row, left to right: Auditory-nerve fibers with high (ANH-1) and low (ANL) spontaneous rates. Bottom row, left to right: a regular and irregular ST cell with moderate inhibitory input. For each tone intensity, the noise level was varied to yield signal-to-noise ratios between -10 and +20 dB.

cells onto ST cells made the  $d'$  values smaller. This should not come as a great surprise; TUB cells affect pure tones, not broadband noise stimuli.

## V. RESPONSES TO NOTCH NOISE

Figure 10 shows the effect of changing notch width on the responses of AN fibers, an MP-II cell, and ST cells with either weak or strong inhibition from MP-II cells. The plots represent the output rate as a function of stimulus intensity. The plots also show the responses (dashed line) that are seen when wideband noise of equal spectral levels to the notch-noise sidebands and to the notch-noise floor are used as inputs. The presence of suppression/inhibition is evidenced by the fact that larger notch widths can result in response rates lower than those of wideband noise at levels equal to the notch-noise floor. This is seen in all model cells except the MP-II cell. Excitation is seen for all notch widths in all cells at higher intensities. Among the ST cells in the bottom row of Fig. 10, it can be observed that even moderate inhibition results in a shift of the notch rate-level functions toward higher intensities. If the threshold also changes (Fig. 10, bottom right) the responses are like those of a low-spontaneous AN fiber.

## VI. PSTH PATTERNS OF ST CELLS

The response pattern commonly associated with stellate cells is the chopper response, which is characterized by a regularity in the interval between successive spikes. This pattern has been grouped in two categories based on the coefficient of variation [CV, the ratio of the standard deviation in interspike interval (ISI) over the mean; see, e.g., Blackburn and Sachs, 1989]. The mean and standard deviation in ISI are calculated for short time bins following the stimulus on-



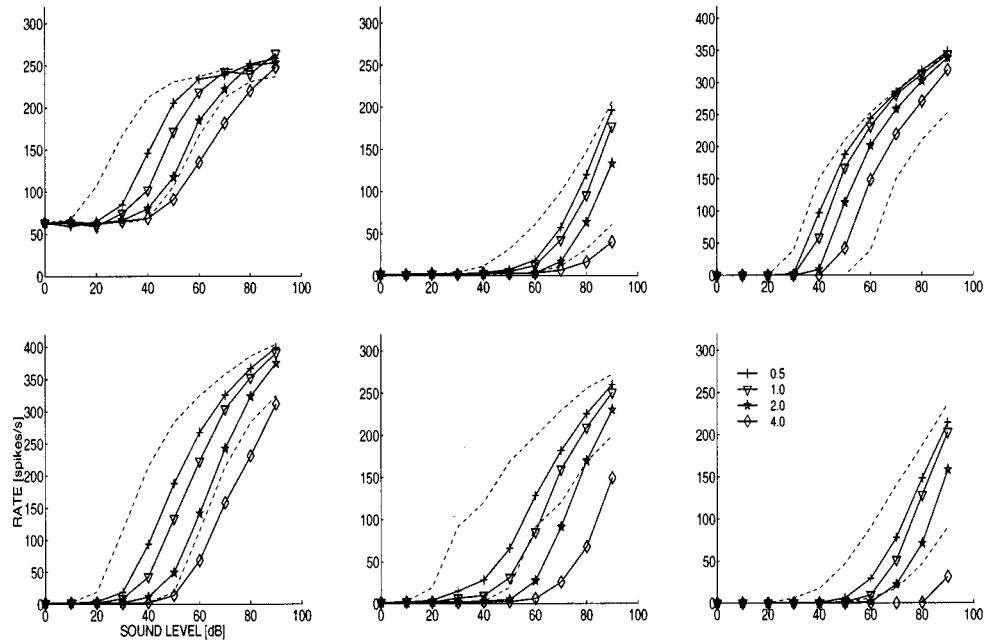


FIG. 10. Responses to notch noise centered at 5 kHz. Notch widths: 0.5, 1, 2, and 4 kHz. Dashed lines indicate response to broadband noise at the level of the sides and the floor of the notch stimuli. Top row: High- and low-spontaneous AN fibers, and MP-II cell. Bottom row: ST cell (regular) with zero and moderate inhibition, ST cell (irregular, threshold 10 dB higher) with moderate inhibition.

set (actual method given in Young, Robert, and Shofner, 1988). So-called sustained choppers fire at regular intervals with little sign of temporal adaptation, whereas transient choppers fire regularly for 10–15 ms and then become irregular. One explanation given for this difference is that delayed inhibition produces the less regular firing in transient choppers. In a modeling study, Banks and Sachs (1991) showed that this could be the case if the inhibition was delayed, with delays increasing with distance from the soma. However, in their model, Hewitt and Meddis (1993) showed that transient responses can be produced without inhibition by varying model parameters such as threshold, number of inputs, and current delivered by each input. The issue then is what role, if any, inhibition plays in generating these responses.

In a follow-up to their original study, Blackburn and Sachs (1992) examined the effects of off-best frequency (BF) tones, in which they observed that response patterns could be influenced by such stimuli. They noted that what they termed the temporal adaptation pattern (TAP, the change in ISI with time) would either remain the same or change in certain ways when off-BF tones were presented. The changes were from sustained to slowly adapting, slowly adapting to transiently adapting, and transiently adapting to slowly adapting. Four of these ISI patterns are shown in Fig. 11, to the right of the PSTH for the same cell and condition. The values in the plots represent the mean and standard deviations of the interspike interval as a function of time. On the right, the upper panel represents a regular chopper and below it the same unit when an off-BF tone is added to the tone at the cell's CF. The cell response changes from a steady, sustained pattern to a slowly adapting one. In the panels below these responses of an irregular chopper whose TAP changes from slowly (top) to transiently adapting (bottom) are shown. In several cases,

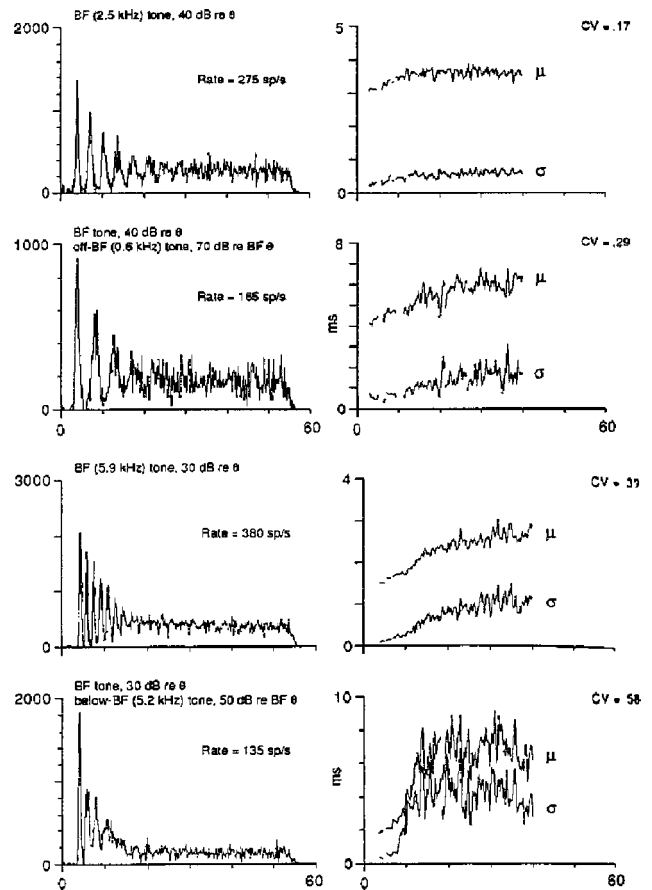


FIG. 11. Data presented in Blackburn and Sachs (1992). Top four panels: Regular chopper and its change from a sustained to a slowly adapting pattern. Bottom four panels: Irregular chopper and its change from a slowly adapting to transient pattern.

changes in TAPs due to off-BF stimulation was accompanied by a change in regularity. Since regularity changes with mean rate, Blackburn and Sachs compared the CV for off-BF stimulation with the CV for a pure tone which yields the same mean rate, and found that these could be different. This change appears to be more frequent in sustained choppers.

The ST cells in this model display many of the same temporal responses found in recordings from choppers cells. These patterns depend on how the input (excitatory and inhibitory) is distributed, on the number of synapses, and on the strength of synaptic input. As mentioned, these parameters have been used to explore chopper cell response properties in previous studies. The results obtained with this model complement these studies by examining how the type of inhibitory inputs modeled can generate changes in chopper responses similar to those observed in the laboratory. The principal difference with Banks and Sachs's model is that the duration of inhibitory synaptic events is in the upper range of those reported by Oertel (1983). This produces a more sustained and gradual buildup of effect, which is effective in generating the TAPs similar to those reported. Another difference is in the use of the MP-II cells' onset-chopper response pattern to effect the change from slow to transient adaptation.

Figure 12 shows some of the response patterns obtained with the model cells. Each panel shows the mean and standard deviation in ISI as well as the PSTH and the CV (insets). Values for the mean, standard deviation, and the coefficient of variation shown in the plots are average values obtained between 30 and 40 ms after stimulus onset. The top four panels are for the same cell (15 excitatory inputs to dendrites) with responses to a 5-kHz tone at 50 dB. In the top left there is no inhibition. The PSTH and ISIs are characteristic of a sustained regular chopper. To obtain a steady response, it was necessary to include an A-type current in the model (the fact that Banks and Sachs did not require this in their model may be due to a difference in the AN input model). In the top left, inhibition from a TUB cell changes the TAP to a slowly adapting pattern. Hence, one way of obtaining this type of pattern is if the inhibition is sustained (recall that the TUB model cell has a chopper-like response as well). In the second row, moderate inhibition from an MP-II cell is added (to dendrites). An off-BF tone reduced the ST output and increases its CV (right). However, the pattern of adaptation is too rapid for a regular chopper. The third row of panels is for a cell with excitatory input to the cell soma and dendrites (15 connections each) and with strong MP-II inhibitory input. In this cell, an increase in stimulus intensity resulted in a regular firing for about 10 ms, which then became irregular. Finally, the bottom panel shows a similar unit whose response changes from slowly adapting to transiently adapting when a 3-kHz tone is added.

To show how ISI mean and CV are related, one can plot the standard deviation vs the mean (Fig. 13). The points fall on a line whose intersection with the  $x$ -axis corresponds to the minimum ISI. Points above the line correspond to higher CVs for an equal mean ISI. This occurs when excitatory input is shifted to the soma, as well as when the number of inputs is reduced (top two panels). The bottom-left panel

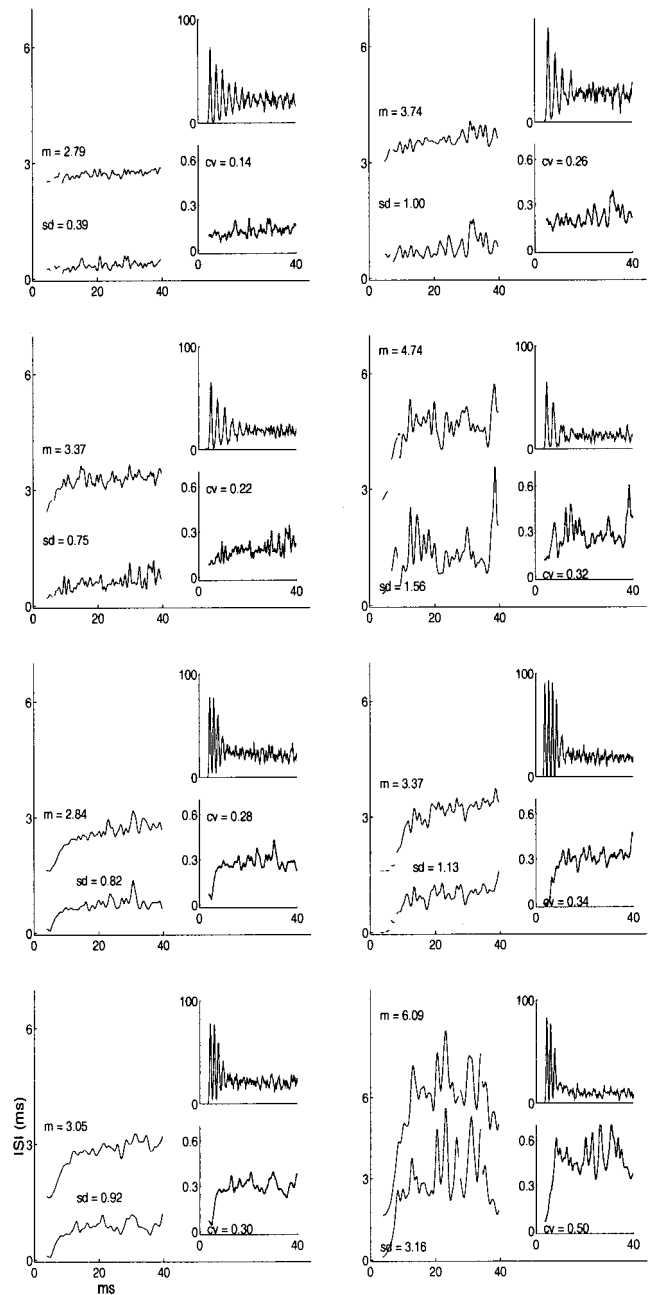


FIG. 12. Response patterns generated by ST cells for different model parameters and stimulus conditions. See the text for details.

compares the responses of a cell to single pure tones when inhibition is present or absent (excitatory input to dendrites). Inhibition on dendrites or soma produced the same results. The bottom right shows the responses of the same cell as that in the bottom of Fig. 13, with and without off-BF stimuli.

## VII. DISCUSSION

The first part of the present work has been to evaluate spectral properties of cochlear nucleus neurons in the presence of a wideband noise background. To evaluate the effect of inhibition in the CN, a set of hypothetical schemes of connectivity has been tested in a series of simulations. The main conclusion is that the model can qualitatively reproduce a large part of the data reported by Rhode and Green-

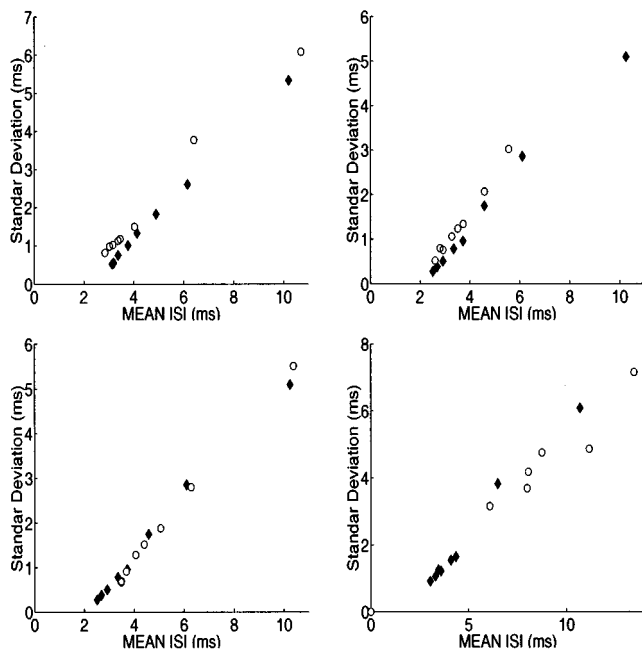


FIG. 13. Plots of ISI mean vs standard deviation. Top left: Responses to different levels of a 5-kHz tone for an ST cell with excitatory input to dendrites (diamonds) and one with excitatory input to soma and dendrites (circles). The first is a regular chopper, the second irregular. Top right: Same stimuli but for cells with 15 connections to dendrites (diamonds) versus 6 connections (circles). Bottom left: Comparison between cell with or without inhibitory input (excitatory input to dendrites). Bottom right: Responses of an irregular chopper (excitatory/inhibitory input to soma and dendrites for 5-kHz tones (diamonds) and to combined 5-kHz/3-kHz stimuli at different levels.

berg (1994). It also corresponds to data reported by others (Gibson, Young, and Costaloupes, 1985). Certain intermediate results obtained in a quiet background should also be of interest. These points will now be discussed individually.

The model of the auditory periphery used here includes a basilar-membrane model which can exhibit two-tone suppression and an IHC/AN synapse model which exhibits both rapid ( $<10$  ms) and short-term adaptation to a stimulus. For long-term adaptation an *ad hoc* method was employed to reduce the AN firing rate. These mechanisms can account for most of the effects produced by tone/noise combinations observed in the auditory nerve. Suppression of noise by pure tones can also be produced, and vice versa.

Cochlear nucleus connectivity schemes were motivated by findings in physiological and anatomical studies. Responses of VCN stellate cells with different degrees of input from inhibitory neurons have been compared. How variations in response patterns observed in putative inhibitory (type II multipolar and tuberculoventral) neurons may arise have also been examined. One of the configurations considered included a inhibitory link originating from TUB cells and terminating on MP-II cells. The resulting rate-intensity curves and response areas are similar to those observed in many onset-chopper cells in the PVCN. Inhibitory cells with wide response areas but attenuated response around its CF could be effective in reducing the transmission to higher-order auditory centers of the low-frequency “tail” response seen in AN fibers. A comparison between simulated responses of onset-chopper (MP-II) neurons with the masked

response curves shown in Rhode and Greenberg’s study shows that these match very well. Their MRC would correspond to the MP-II cell with inhibition. Although they do not show the response area for the same neuron in their paper, a unit with similar CF and threshold in Rhode (1991) does have a response area with a slightly asymmetrical increase in response rate, similar to Fig. 3 (top row, right panel). They also show a unit with a narrow response area and a non-monotonic rate-intensity curve characteristic of type II units, whose responses to tones in background noise are almost completely absent. The MRC for a type II unit in Gibson *et al.* (1985) is very similar to the TUB cell (Fig. 6).

Responses of stellate cells to tones in a noisy background were shown in Figs. 7 and 8. Masked rate-level curves for stellate cells with inhibitory input exhibit marked shifts toward higher stimulus levels in agreement with MRCs obtained for VCN chopper cells in Rhode and Greenberg (1994). As pointed out in the previous section, this shift is already present in AN fibers, and normalized MRCs for ST cells are not very sensitive to the degree of inhibition. The regular MRCs are influenced by the inhibition principally via the reduction in the response to the noise alone.

Stellate cells and type III fusiform cells show the greatest amount of lateral inhibition of all cells recorded in the CN (Rhode and Greenberg, 1994). If the function of these cells is to transmit spectral information to higher brain centers, the inhibition would serve to enhance the contrast between the noise-driven background activity and the spectral peaks in the signal. In Rhode and Greenberg’s study, both sustained and transient choppers showed prominent inhibitory sidebands. Because the magnitude and threshold of the suppression differed from that of AN fibers, they concluded that additional suppression must come from inhibitory neurons in the CN. The difference between sustained and transient choppers may be more a matter of how inhibition is distributed, rather than an all-or-none distinction. In this paper, it is suggested that stellate cells with regular chopper patterns receive inhibitory input on the dendrites. Moreover, with the parameters used for this model, inhibitory input is more regular than excitatory input. Due to these two factors, inhibition is not the main contributor in making a cell’s response irregular.

The wideband inhibitory input from the MP-II (or onset-chopper) cells, in combination with the adjustment occurring in the cochlea reflected in the AN recordings reported by Costaloupes and colleagues, probably constitutes the major mechanism by which CN neurons adjust their dynamic range in the presence of a background noise. Furthermore, the simulations involving notch noise indicated that responses of ST cells to these stimuli vary with the strength of inhibition. Naturally occurring notches are encoded by the discharge rate of AN fibers and are known to be important for sound localization (Poon and Brugge, 1993). Figure 10 showed responses of model AN fibers to notches of different widths. For the largest widths tested, the response rate fell below that of a flat broadband noise at the level of the notch floor. Although we are not aware of any experimental data to support these results, May and colleagues (May *et al.*, 1996) show that the AN rate response to background noise can be

reduced when the signal is a vowel and the fiber's CF is near the trough between two formants. They also show that both sustained and transient choppers maintain or enhance this spectral contrast (May, Prell, and Sachs, 1998). The model ST cell's responses to notch noise displayed increased contrast relative to AN responses when inhibition was turned on, but this was less evident when the ST cell's threshold was higher (Fig. 10, bottom-right panel). These results could also be interesting for models of dorsal cochlear nucleus neurons, since there is evidence that the DCN receives excitatory input from the cochlear nucleus itself in addition to the auditory nerve.

A comparison of responses of AN fibers with ST neurons using a  $d'$  metric has also been made. The fact that  $d'$  values for ST neurons were larger than those for AN fibers agrees with the observations made by Kim and colleagues in comparisons made between AN fibers and PVCN neurons (Kim *et al.*, 1991). They stated that the transformation from AN fibers to PVCN chopper neurons involves an amplification of the mean discharge rate and a increase in the ratio of mean to standard deviation. Although signal-detection methods have not been employed in electrophysiological studies for tones in noise, very similar conclusions were made for vowels in the cat VCN (May, Prell, and Sachs, 1998). One factor that may influence mean to standard deviation ratios in neural discharges is the firing rate itself. The postspike refractory period (absolute and/or relative) imposes a regularity on strongly driven cells. Another factor is the integration of synaptic events coming from several independent sources, which many neurons in the CN are known to perform. This increase in reliability is probably one of the functions of the cochlear nucleus.

Finally, the response patterns that the model is able to generate can be quite diverse. The temporal adaptation patterns (TAPs) noted by Blackburn and Sachs (1992) can be generated by changing the model neuron's synaptic input. The involvement of the MP-II type cell accounts well for generating the transient TAPs when off-BF tones are used. The use of a long-time constant for the inhibitory input conductances (6 ms) was sufficient for generating the TAPs observed in real cells. This provides an alternative to Banks and Sachs's proposal that inhibitory input is delayed by up to several milliseconds. The inhibition affects the membrane potential more gradually and smoothly, even when few (3) inhibitory connections are used. Therefore, although the coefficient of variation (CV) is increased, it corresponds to that of a rate-matched CV without inhibition. This appears to be the case for most irregular chopper units recorded by Blackburn and Sachs. However, the model does not account for two experimental observations. The first is the transition, in regular choppers, from a sustained to slowly adapting pattern with off-BF tones, and the second is the sometime associated change in regularity. The model could produce responses with *increased* regularity with off-BF inhibition (not shown). This occurred when the excitatory input resulted in a membrane potential with little variability and the inhibitory input was smooth enough to simply lower that membrane potential, resulting in a lower mean rate but equal variance. *Decreased* variability may have to do with how excitatory and

inhibitory input is distributed on the dendritic trees of regular choppers. If inhibition is unevenly distributed on different branches (some branches receiving strong inhibition and others none), EPSPs would be effectively reduced in numbers rather than size. Since it has been shown here and elsewhere that the number of excitatory inputs is related to the regularity of chopper models, and since there is evidence that excitatory and inhibitory inputs are found at the same spatial locations, this is a plausible hypothesis that should be further investigated.

## VIII. CONCLUSION

This work provides a summary of some existing hypotheses concerning the function of neural circuits of the cochlear nucleus. We have assembled a model with a large number of realistic features which can mimic many of the responses observed in physiological recordings. A model of the auditory periphery with several realistic properties was used for these simulations. Auditory-nerve tuning curves with typical shapes can be obtained, and the model also exhibits compressive nonlinearity at high intensities as well as two-tone suppression. Model multipolar onset-chopper and tuberculoventral neurons provide the inhibitory input onto stellate cells of the ventral cochlear nucleus. The aim was to compare the encoding of sounds in simulated auditory-nerve fibers and cells of the cochlear nucleus with published data. The main conclusion is that, of the cell types modeled here, the MP-II cell has a special role in adjusting the responses of the principal projection neurons of the cochlear nucleus to a broadband background noise. It appears to be a good candidate for generating transient chopper responses as well.

As an investigative tool, modeling is becoming increasingly important for studying the nervous system. In this regard, the model presented here can be considered to be fairly detailed, while at the same time ignoring other important aspects. For example, we have included both dendritic and somatic compartments and, in certain cases, nonlinear voltage-dependent conductances in the model neurons. These details contribute little to the results which involve mean discharge rates over a 100-ms period. They do matter for the fine temporal structure of the response, however, and the aim has been to test a single composite model under several different conditions. On the other hand, we have not included systems such as the olivocochlear reflex system. These appear to play an important role in adjusting the dynamic range of auditory-nerve fibers when a continuous background noise is present, as has been demonstrated both by electrical stimulation in the anesthetized cat (Winslow and Sachs, 1988). It would be useful to incorporate this type of feedback (rather than the *ad hoc* method used here) in future models when its mechanisms are further elucidated.

This work builds upon several previous modeling studies of the cochlear nucleus neurons mentioned in the paper, and bears resemblance to these in several respects. However, there has so far been no modeling work related to the representation of sounds in the presence of background noise. We think that these topics deserve to be studied and that insights into these should provide a firmer base from which to investigate more abstract problems involving complex sounds.



## ACKNOWLEDGMENTS

This work is supported by the Swiss National Scientific Foundation, Grant Number 2150-045689.95.

- Adams, J. C. (1979). "Ascending projections to the inferior colliculus," *J. Comp. Neurol.* **183**, 519–538.
- Arle, J., and Kim, D. (1991). "Neural modeling of intrinsic and spike-discharge properties of cochlear nucleus neurons," *Biol. Cybern.* **64**, 273–283.
- Banks, M. I., and Sachs, M. B. (1991). "Regularity analysis in a compartmental model of chopper units in the anteroventral cochlear nucleus," *J. Neurophysiol.* **65**, 606–629.
- Blackburn, C. C., and Sachs, M. B. (1989). "Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis," *J. Neurophysiol.* **62**, 1303–1329.
- Blackburn, C. C., and Sachs, M. B. (1992). "Effect of OFF-BF tones on responses of chopper units in ventral cochlear nucleus. I. regularity and temporal adaptation patterns," *J. Neurophysiol.* **68**, 124–143.
- Cant, N. B. (1981). "The fine structure of two types of stellate cells in the anterior division of the anteroventral cochlear nucleus of the cat," *Neuroscience (Oxford)* **6**, 2643–2655.
- Cant, N. B. (1992). "The cochlear nucleus: Neuronal types and their synaptic organization," in *The Mammalian Auditory Pathway: Neuroanatomy*, edited by A. N. Popper and R. R. Fay (Springer, Berlin), Chap. 3.
- Costaloupes, J. A. (1985). "Representation of tones in noise in the responses of auditory nerve fibers in cats," *J. Neurosci.* **5**, 3261–3269.
- Doucet, J. R., and Ryugo, D. K. (1997). "Projections from the ventral cochlear nucleus to the dorsal cochlear nucleus in rats," *J. Comp. Neurol.* **385**, 245–264.
- Geisler, C. D., and Sinex, D. G. (1980). "Responses of primary auditory fibers to combined noise and tonal stimuli," *Hearing Res.* **3**, 317–334.
- Gibson, D. J., Young, E. D., and Costaloupes, J. A. (1985). "Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei," *J. Neurophysiol.* **53**, 940–958.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Krieger, Huntington, NY).
- Greenwood, D. D. (1990). "A cochlea frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Hewitt, M., and Meddis, R. (1993). "Regularity of cochlear nucleus stellate cells: A computational modeling study," *J. Acoust. Soc. Am.* **93**, 3390–3399.
- Jiang, D., Palmer, A. R., and Winter, I. M. (1996). "Frequency extent of two-tone facilitation in onset units in the ventral cochlear nucleus," *J. Neurophysiol.* **75**, 380–395.
- Josephson, E. M., and Mostert, D. K. (1998). "Synaptic diversity of the distal dendrites of type I stellate cells in the cochlear nucleus," *Assoc. Res. Otolaryngol.* (abstract).
- Kim, D., Ghoshal, S., Khant, S. L., and Parham, K. (1994). "A computational model with ionic conductances for the fusiform cell of the dorsal cochlear nucleus," *J. Acoust. Soc. Am.* **96**, 1501–1514.
- Kim, D. O., Parham, K., Sirianni, J. G., and Chang, S. O. (1991). "Spatial response profiles of posteroventral cochlear nucleus neurons and auditory nerve fibers in unanesthetized decerebrate cats: Responses to pure tones," *J. Acoust. Soc. Am.* **89**, 2804–2817.
- Lieberman, M. C. (1982). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," *J. Acoust. Soc. Am.* **72**, 1441–1449.
- Lieberman, M. C. (1991). "Central projections of auditory nerve fibers of differing spontaneous rate. I. Anteroventral cochlear nucleus," *J. Comp. Neurol.* **313**, 240–258.
- Lieberman, M. C. (1993). "Central projections of auditory nerve fibers of differing spontaneous rate. II. Posteroventral and dorsal cochlear nuclei," *J. Comp. Neurol.* **327**, 17–36.
- MacGregor, R. (1987). *Neural and Brain Modeling* (Academic, New York).
- Manis, P. B., Marx, S., and White, J. (1996). "Integrative mechanisms of ventral cochlear nucleus stellate cells: Membrane conductances and electrotonic structure," in *Speech, Hearing and Language Processing*, edited by W. Ainsworth, E. Evans, and C. Hackney (JAI).
- May, B. J., Huang, A., Prell, G. S., and Heinz, R. (1996). "Vowel formant frequency discrimination in cats: comparison of auditory nerve representations and psychophysical thresholds," *Aud. Neurosci.* **3**, 135–162.
- May, B. J., Prell, G. S. L., and Sachs, M. B. (1998). "Vowel representations in the ventral cochlear nucleus of the cat: Effects of level, background noise, and behavioral state," *J. Neurophysiol.* **79**, 1755–1769.
- Nelken, I., and Young, E. (1994). "Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type iv units to wideband stimuli," *J. Neurophysiol.* **71**, 2446–2462.
- Oertel, D., Wu, S., Garb, M. W., and Dizack, C. (1990). "Morphology and physiology of cells in slice preparations of the posteroventral cochlear nucleus of mice," *J. Comp. Neurol.* **295**, 136–154.
- Oertel, D., Wu, S., and Hirsh, J. (1988). "Electrical characteristics of cells and neuronal circuitry in the cochlear nuclei studied with intracellular recordings from brain slices," in *Auditory Function*, edited by G. Edelman, G. Wall, and W. Freeman (Wiley, New York).
- Oertel, D. (1983). "Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus," *J. Neuroscience* **3**, 2043–2053.
- Palmer, A. R., and Evans, E. F. (1982). "Intensity coding in the auditory periphery of the cat: Responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise," *Hearing Res.* **7**, 305–323.
- Palmer, A. R., Jiang, D., and Marshall, D. (1996). "Responses of ventral cochlear nucleus onset and chopper units as a function of signal bandwidth," *J. Neurophysiol.* **75**, 780–794.
- Phillips, D. P., and Cynader, M. S. (1985). "Some neural mechanisms in the cat's auditory cortex underlying sensitivity to combined tone and wide-spectrum noise stimuli," *Hearing Res.* **18**, 87–102.
- Poon, P. W. F., and Brugge, J. F. (1993). "Sensitivity of auditory nerve fibers to spectral notches," *J. Neurophysiol.* **70**, 655–666.
- Rees, A., and Palmer, A. R. (1988). "Rate-intensity functions and their modification by broadband noise for neurones in the guinea-pig inferior colliculus," *J. Acoust. Soc. Am.* **83**, 1488–1499.
- Rhode, W. S. (1991). "Physiological-morphological properties of the cochlear nucleus," in *Neurobiology of Hearing: The Central Auditory System*, edited by R. A. Altschuler, R. P. Bobbin, B. M. Clopton, and D. W. Hoffman (Raven, New York), Chap. 3.
- Rhode, W. S., and Greenberg, S. (1992). "Physiology of the cochlear nuclei," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, Berlin), Chap. 3.
- Rhode, W. S., and Greenberg, S. (1994). "Lateral suppression and inhibition in the cochlear nucleus of the cat," *J. Neurophysiol.* **71**, 493–514.
- Rhode, W. S., and Kettner, R. E. (1987). "Physiological study of neurons in the dorsal and posteroventral cochlear nucleus of the unanesthetized cat," *J. Neurophysiol.* **57**, 414–442.
- Rhode, W. S., and Smith, P. H. (1986). "Encoding timing and intensity in the ventral cochlear nucleus of cat," *J. Neurophysiol.* **56**, 261–286.
- Robert, A., and Eriksson, J. L. (1999). "A composite model of the auditory periphery for simulating responses to complex sounds," *J. Acoust. Soc. Am.* **106**, 1852–1864.
- Rouiller, E. M., Cronin-Schreiber, R., Fekete, D. M., and Ryugo, D. K. (1986). "The central projections of intracellularly labeled auditory nerve fibers in cats: an analysis of terminal morphology," *J. Comp. Neurol.* **249**, 261–278.
- Rouiller, E. M., and Ryugo, D. K. (1984). "Intracellular marking of physiologically characterized cells in the ventral cochlear nucleus of the cat," *J. Comp. Neurol.* **225**, 167–186.
- Sachs, M. B., and Young, E. D. (1979). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," *J. Acoust. Soc. Am.* **66**, 471–479.
- Schreiner, C. E., and Langer, G. (1997). "Laminar fine structure of frequency organization in auditory midbrain," *Nature (London)* **388**, 383–386.
- Smith, P. H., and Rhode, W. S. (1989). "Structural and functional properties distinguish two types of multipolar cells in the ventral cochlear nucleus," *J. Comp. Neurol.* **282**, 595–616.
- Voigt, H. F., and Young, E. D. (1990). "Cross-correlation analysis of inhibitory interactions in dorsal cochlear nucleus," *J. Neurophysiol.* **64**, 1590–1610.
- Wickesberg, R., and Oertel, D. (1988). "Tonotopic projection from the dorsal to the anteroventral cochlear nucleus of mice," *J. Comp. Neurol.* **268**, 389–399.
- Winslow, R. L., Barta, P., and Sachs, M. B. (1987). "Rate coding in the auditory nerve," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).
- Winslow, R. L., and Sachs, M. B. (1988). "Single-tone intensity discrimi-

- nation based on auditory-nerve responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle," *Hearing Res.* **35**, 165–190.
- Winter, I. M., and Palmer, A. R. (1995). "Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise," *J. Neurophysiol.* **73**, 141–159.
- Young, E. D., Robert, J.-M., and Shofner, W. P. (1988). "Regularity and latency of units in ventral cochlear nucleus: Implications for unit classification and generation of response properties," *J. Neurophysiol.* **60**, 1–29.
- Zhang, S., and Oertel, D. (1993). "Tuberculoventral cells of the dorsal cochlear nucleus of mice: Intracellular recordings in slices," *J. Neurophysiol.* **69**, 1409–1421.
- Zhang, S., and Oertel, D. (1994). "Neuronal circuits associated with the output of the dorsal cochlear nucleus through fusiform cells," *J. Neurophysiol.* **71**, 914–930.
- Zhang, Y., Suga, N., and Yan, J. (1997). "Corticofugal modulation of frequency processing in bat auditory system," *Nature (London)* **387**, 900–903.

# Solution of the inverse problem for a linear cochlear model: A tonotopic cochlear amplifier

Emilios K. Dimitriadis<sup>a)</sup>

*Bioengineering and Physical Sciences Program/ORS/OD, National Institutes of Health, Bethesda, Maryland 20892*

Richard S. Chadwick

*Laboratory of Cellular Biology, Section on Auditory Mechanics/NIDCD, National Institutes of Health, Bethesda, Maryland 20892*

(Received 3 June 1997; revised 11 February 1999; accepted 18 March 1999)

The extraordinary fine-tuning characteristic of normal mammalian hearing is attributed to physiological mechanisms collectively known as the cochlear amplifier (CA), which amplifies and sharpens the basilar membrane (BM) vibration response to incoming acoustic pressure oscillations. Electromechanical properties of outer hair cells (OHCs) are believed to be the critical component of the CA, but its “circuitry” as yet remains unknown. Here, the required frequency-space response characteristics of the CA are computationally determined when typical *in vivo* tuning data are introduced as input to a linear hydroelastic cochlear model whose cross-sectional dynamics are represented by two coupled vibrational degrees of freedom. It is assumed that the CA senses motion at the tectorial membrane (TM) reticular lamina (RL) and applies proportional, equal, and opposite forces to the BM and the RL. The results show the CA to be tonotopically tuned, meaning it conforms to a space-frequency similarity principle like other cochlear dynamical responses. This requires that the active mechanism use information distributed along the cochlear partition. The physiological mechanism responsible for such behavior remains unknown, but here the computed CA characteristics can be qualitatively reproduced by a circuit spanning the length of the cochlea. This does not preclude other mechanisms, but is intended to motivate closer experimental investigation of extracellular and intercellular ionic flow pathways. [S0001-4966(99)00109-5]

PACS numbers: 43.64.Bt, 43.64.Kc, 43.64.Ld, 43.64.Nf [RDF]

## INTRODUCTION

Many questions concerning the physiological mechanisms responsible for extraordinary auditory fine tuning, i.e., the putative cochlear amplifier (CA) (Gold, 1948; Davis, 1983), as yet remain unanswered. The difficulty lies in measuring and even conceptualizing the complex dynamics of the structurally complex cochlear partition at the micromechanical level. It has long been known that the cochlea itself performs not only tonotopic frequency analysis, but it also possesses a control mechanism that strongly sharpens and amplifies the organ of Corti (OC) response to the incoming sound pressure fluctuations (Rhode, 1971; Davis, 1983). This mechanism is inactivated by trauma or postmortem (Lieberman and Dodds, 1984), and its existence became evident only through the use of powerful measurement techniques such as Mössbauer (Rhode, 1971; Sellick *et al.*, 1986; Johnstone *et al.*, 1986; Robles *et al.*, 1986), laser interferometry (Khanna and Leonard, 1982; Russell and Nilsen, 1997), and laser velocimetry techniques (Nuttal *et al.*, 1991; Ruggero and Rich, 1991a; Nuttal and Dolan, 1996). The issue whether the control mechanism generates energy has not been fully resolved (Allen and Fahey, 1992) as it is rather difficult to measure its characteristics without an *a priori* model. Drugs or trauma that affect either the stereocilia or the hair cell cytoskeleton seem to strongly influence the per-

formance of the controller and appear to support the notion of an active mechanism (Lieberman and Dodds, 1984; Ruggero and Rich, 1991b). Both sensors (Hudspeth and Corey, 1977) and actuators (Brownell *et al.*, 1985; Kachar *et al.*, 1986; Ashmore, 1990) for this controller have been associated with the motile action of OHCs. The central problem with a CA based on independently acting OHCs is the  $1/\omega$  dependence of the receptor potential on stereocilia displacement above the cell's corner frequency (Palmer and Russell, 1986; Kidd and Weiss, 1990; Housley and Ashmore, 1992; Santos-Sacchi, 1992; Dallos and Evans, 1995a, b; Kolston, 1995), which results in a strong attenuation of the active force at high frequencies. It has also been shown that, with one degree-of-freedom models, the force applied by independently acting OHCs could never produce amplification due to the phase of such force (Markin and Hudspeth, 1995). The problem has usually been overcome by a shift in the phase of the force by a mechanism which could not be accounted for within the framework of the model unless a second degree-of-freedom was added. Proposed compensation for this “deficiency” of independently acting OHCs may be classified under two different broad mechanisms. The first such mechanism postulates the existence of a second mechanical filter in the form of a radial resonance of the TM (Zwislocki and Kletsky, 1979; Allen, 1980), whose role is to amplify the stereocilia deflections. In most of the previous models the

<sup>a)</sup>Electronic mail: dimitria@helix.nih.gov

TM radial motion does not add another degree of freedom, but rather acts as an external controller. Models with one (Zwislocki and Kletschy, 1979; Allen, 1980; Mammano and Nobili, 1993; Geisler and Sang, 1995) or two (Hubbard, 1993; Neely, 1993) transverse degrees of freedom have incorporated this mechanism with some success although, with most models, the required OHC active force tends to be excessive (cf. Sec. III). Recent measurements of relative motions within the cochlear partition (CP) (Ulfendahl *et al.*, 1989; Mammano and Ashmore, 1995; Xue *et al.*, 1995; Gummer *et al.*, 1996) suggest that the CP cross section be modeled with at least two fully coupled oscillators. A fully coupled system is presented by Neely (Neely, 1993) where realistic responses were obtained only with model parameters that effectively eliminated one of the two transverse degrees of freedom while maintaining the radial tectorial motion. The realistic response, however, was achieved at the expense of the OHCs which were required to generate unrealistic forces. This point is further discussed at the end of this paper, where it is proposed that the outer hair cell force generation required in the model could serve as the basis of comparison for the different modeling approaches. The second proposed mechanism to overcome high frequency attenuation of OHC motile response (Dallos and Evans, 1995a, b; Kolston, 1995) postulates that the transmembrane potential modulation is accomplished at high frequencies by extracellular potential oscillations caused by ionic flows through known and perhaps as yet unknown electrical pathways. This would effectively short-circuit the slow stereocilia/cell membrane circuit. One of these two competing mechanisms or some combination of the two could, in the future, elucidate the macroscopic nature of the CA.

Our previous cochlear model with two fully coupled transverse degrees of freedom maintained realistic structural parameter values and active force generation for independently acting OHCs (Chadwick *et al.*, 1996). However, frequency responses with physiological  $Q$ -factors were never achieved even with the implementation of a radial TM resonance. Large amplification of the passive wave was achievable only for frequencies below about 3–4 kHz. At higher frequencies performance deteriorated and amplification could not be achieved. In addition, frequency response  $Q$ -factors fell far short of *in vivo* measurements, even though active power was shown to be supplied to the CP. Evidently, the CA axial distribution seems to be very important, but the relatively small variations in the axial distribution of a CA resulting from the axial graduation in OHC lengths is inadequate. Those results point to an important difference between single- and multi-partition models. Energy paths in single-partition models can only exist between the OHCs and the partition. In multi-partition models, CA control energy can be fed or absorbed at both the BM and the RL, and depends on the tonotopic axial variation of the mode shape.

Here, the same model is treated in a different manner that can indicate possible ways to overcome the high-frequency motility attenuation problem. The initial objective is to mathematically describe the CA without prior assumptions about the underlying physiological mechanism by computing the electromechanical space-frequency characteristics

of the CA consistent with *in vivo* tuned responses which are used here as an input to the model. This is the so-called inverse problem. This technique was previously used to advantage for single degree of freedom models of the cochlea by Zweig (Zweig, 1991) and by deBoer (deBoer, 1995a, b). Here, the only assumption in formulating the problem is that, irrespective of the actual active mechanism, the CA senses motion at the stereocilia and applies equal but opposite forces to the BM and the RL. No *a priori* knowledge of electro-mechanical connections or parameter values is required. Under these conditions the following questions are posed: What should the CA space-frequency response map be if the model were to predict frequency responses similar to those obtained from *in vivo* experiments? Can possible mechanisms be proposed in the form of electromechanical circuitry for such a CA? The mathematical answer to these questions was sought via a tedious but conceptually simple numerical procedure. Clearly, if these active filter characteristics were available, the efficacy of a variety of proposed physiological mechanisms could be evaluated. The inverse problem is solved here using several input data sets and the results consistently indicate that the CA should be tonotopically tuned. This implies that the CA must sense more than the local motion, which is what independently acting OHCs would do. At this point we can only speculate on the mechanism of the tonotopic tuning of the CA. As one such speculative example, we investigated a distributed electrical cable model allowing for the modulation of the extracellular potential. The frequency response characteristics of the circuit are qualitatively similar to those computed from the inverse problem, provided a low-impedance, axial ionic pathway exists between neighboring OHCs. The proposed circuit is an invitation for further experimental work to elucidate the ionic pathways in the OC. The computed active filter characteristics are open to other possible interpretations and suggestions for new proposals for the possible physiological mechanism underlying the tonotopic character of the CA.

## I. MODEL DESCRIPTION

The linear model previously proposed and analyzed by the authors (Chadwick *et al.*, 1996) with independently acting OHCs was developed with the criterion that controller elements be placed in a structurally realistic setting. In that model, the cochlea is represented by a straightened duct with rectangular cross section bounded by rigid walls. The dynamic elements representing the organ of Corti (OC) separate the two fluid-filled spaces. Different parts of the OC cross section are allowed to oscillate with different amplitudes and phases. Such a representation is consistent with experiments demonstrating significant relative motions within the OC (Ulfendahl *et al.*, 1989; Gummer *et al.*, 1996). If OHCs are the active elements of the CA, they will apply forces at their two ends, their regions of attachment to the Deiter's cells and the RL. Since the RL is attached to the relatively rigid pillar cells, it is expected to undergo rocking motion about the top of the tunnel of Corti. The transverse component of this motion must be followed by the TM since it has been shown that there is no transverse relative motion between TM and RL (Chadwick *et al.*, 1996). Thus the CP



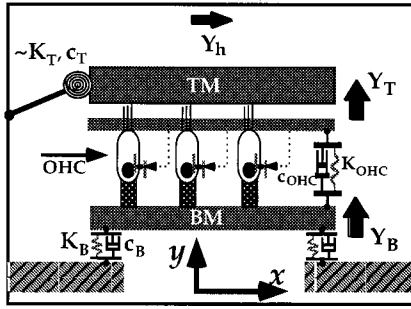


FIG. 1. Lumped parameter representation of cochlear partition cross-section structural elements. The TM/RL complex motion has both vertical ( $Y_T$ ) and horizontal ( $Y_h$ ) components which are assumed to be kinematically coupled. Explanation of the parameters in nondimensional form can be found in the Appendix. Stiffness and damping associated with the RL are not explicitly shown for clarity of the figure.

model has to explicitly allow for independent motions of the BM and the RL–TM complex. The minimal required complexity for such a model is a lumped parameter model with two degrees of freedom. The experimental observation of both transverse and radial displacements in the OC (Ulfendahl *et al.*, 1989; Gummer *et al.*, 1996) need not complicate matters unless significant deformations perpendicular to the OHC axis were to take place. If not, radial and transverse motions can be treated as kinematically coupled and need not be modeled as distinct degrees of freedom. Based on the above arguments, the minimal required system on which proposed CA models may be tested is the two subpartition system shown in Fig. 1. The coordinates ( $X, Y, Z$ ) denote the ‘radial,’ ‘transverse’ (perpendicular to the partitions), and ‘axial’ (along the duct length) directions, respectively. Our only assumption concerning the CA is that OHC stereocilia sense motion at the TM–RL subpartition, while somatic motility of the OHC acts between the TM–RL complex and the BM as occurs *in vivo*. Mechanical axial coupling is included through the fluid–structure interaction. Two fluid channels representing the scalae vestibuli and tympani, each filled with an incompressible liquid provide the fluid loading to the subpartitions. All properties such as mass, stiffness, and damping are allowed to slowly vary along the axial coordinate  $z$ . Locally, the masses perform pistonlike oscillations displacing liquid in the adjacent channels. Elastic coupling of adjacent sections is neglected. Thus the model is a continuous axially arranged array of elastically uncoupled two-degree-of-freedom oscillators with dynamic coupling affected by the fluid above and below this array.

### A. Computation of the traveling wave

The mathematical problem is written in terms of time harmonic waves traveling along the CP, using the Wentzel–Kramers–Brillouin (WKB), or two-scale, approximation. The method, which has been shown to be especially accurate for at least the passive problem, is based on determining the local wave number  $k(\omega, z)$  of the wave traveling along the cochlear axis (Lighthill, 1981). For the active cochlea, the argument for using the WKB method is stronger since the system is slowly varying on the scale of the wavelength, which in the presence of activity should be even shorter. This argument is supported by the comparison of the computed

wave numbers and waveforms for passive and active systems presented by Chadwick *et al.* (1996). It is seen there that the increase in the damping component of the wave number causes the cutoff of the wave before the wave number rate of change becomes very high. Therefore, possible WKB approximation errors will tend to accumulate in regions apical to the wave cutoff location, which would make those errors unimportant. In terms of  $k(\omega, z)$ , the vector of the partition displacement functions for pure tone stimulation may be described by

$$\vec{Y}(t, z; \omega) = \vec{A}(\omega, z) e^{i(\omega t - (1/\varepsilon) \int_0^z k(\omega, z') dz')}, \quad (1)$$

where  $z \in [0, 1]$ , the axial coordinate, is normalized by the cochlea length  $L_c$ . The wave number is normalized by the BM width at the base,  $W_0$ , and is of order 1, and  $\varepsilon = W_0/L_c \ll 1$ . Thus, the wave is comprised of a slowly varying envelope,  $\vec{A}(\omega, z)$ , and a fast oscillating complex exponential. The two disparate length scales,  $L_c$  and  $W_0$ , form the basis of the approximation method. The same wave, but with different amplitudes and phases, travels on both subpartitions. The dynamic equations of motion for the cochlear cross section per unit axial length at  $z$  can be written as

$$[\mathbf{M}(z) \partial_{tt} + \mathbf{C}(z) \partial_t + \mathbf{K}(z)] \vec{Y}(t, z) = \vec{F}_f(t, z) + \vec{F}_c(t, z), \quad (2)$$

where  $\vec{Y}(t, z) = \{\tilde{Y}_T(t, z), \tilde{Y}_B(t, z)\}'$  with subscripts  $T$  and  $B$  referring to TM and BM, respectively and where  $\vec{F}_f$  represents the fluid forcing vector  $W(z)(-\tilde{P}_u, \tilde{P}_l)$  while the prime ( $'$ ) symbol on matrices and vectors indicates transposition.  $W(z)$  is the local BM width and  $\tilde{P}_u, \tilde{P}_l$  are the fluid pressures, averaged over the partition width and computed at the upper TM and the lower BM surfaces, respectively. If  $\tau = RC$  is the time constant of the OHC membrane,  $R$  and  $C$  being its resistance and capacitance, respectively, the OHC control forcing vector can be written in the frequency domain as

$$\vec{F}_c(\omega, z) = \beta(\omega, z) \{1, -1\}' \tilde{Y}_T(\omega, z). \quad (3)$$

If  $\beta(\omega, z) = \text{const}(1 + i\omega\tau)$ , this reduces to the well-known low-pass filter of the isolated OHC. The mechanism that presumably complements the CA characteristics beyond those of independently acting OHCs is mathematically embedded in the function  $\beta(\omega, z)$  which, therefore, represents the additional filter needed to obtain realistic model responses in the low-amplitude linear regime. To clarify the terminology used in this paper, CA refers exactly to this function  $\beta(\omega, z)$  and not the function  $\beta(\omega, z)Y_T(\omega, z)$ . The factor in brackets represents the fact that equal and opposite forces are generated on the BM and the TM. Details for the mass, damping, and stiffness parameter matrices used in the numerical computations are given in the Appendix, item (1). The fluid dynamics in the two scalae are well described by the unsteady Stokes' flow equations. For the upper fluid chamber the fluid velocity and pressure satisfy the equations

$$\rho \partial_t \tilde{\mathbf{v}}_u = -\nabla \tilde{P}_u + \mu \nabla^2 \tilde{\mathbf{v}}_u, \quad \nabla \cdot \tilde{\mathbf{v}}_u = 0, \quad (4)$$

with the boundary condition at the TM–fluid interface:

$$\tilde{\mathbf{v}}_u(x, 0, z, t) = (0, \partial_t \tilde{Y}_T, 0). \quad (5)$$

It is assumed that the top and side walls of the channel are far enough into the ambient fluid compared to the local wavelengths so that there is no interaction with the oscillating OC resulting in effectively a two-dimensional ( $y, z$ ) flow field. Also, the unsteady boundary layer thickness is much smaller than the local wavelength. The solution of these equations, and of the corresponding set of equations for the lower channel, in a slowly varying medium is well known (Lighthill, 1992). The main assumptions here correspond to what is known as the “deep water” and “short wave” approximation which is known to be more accurate for shorter wavelengths than for longer ones. It is, therefore, expected that the amplitude approximation errors will be greater basally but quite small near the peak. It is interesting to note that the first-order WKB solution for the surface pressures is quasi-local in the  $(k, \omega)$ -space in the sense that, apart from the exponential term from Eq. (1), the fluid pressure at any location depends only on the local wave number  $k = k(\omega, z)$ , i.e.,

$$\tilde{P}_u(\omega, z; k) = -\frac{\rho\omega^2\tilde{Y}_TW_0}{k}[1 + k\sqrt{-i\nu/(\omega W_0^2)}], \quad (6)$$

where  $\rho$  and  $\nu$  are the density and kinematic viscosity of the fluid, respectively. This simplification reduces the zeroth-order dynamic equations of motion to a homogeneous algebraic form that can be written in matrix notation as

$$[Z_{ij}(\omega, z; \beta, k)]\{Y_{0,i}\} = 0, \quad (7)$$

where  $[Z_{ij}]$  is the  $(2 \times 2)$  local CP impedance matrix,<sup>1</sup>  $\beta = \beta(\omega, z)$  represents the CA transfer function as discussed in connection with Eq. (3),  $k = k(\omega, z)$  as defined in Eq. (1), and the subscript “0” refers to the zeroth-order displacement amplitudes. The impedance matrix elements are explicitly given in the Appendix, item (1), together with the parameter values used. Equation (7) is a nonstandard eigenvalue problem and its solution determines the local complex wave number  $k(\omega, z; \beta)$  and the complex mode or displacement ratio between the two subpartitions  $y_{BT} = Y_B/Y_T$ . The dispersion relation which equates the determinant of the impedance matrix to zero,  $|Z_{ij}| = 0$ , is a quadratic polynomial in  $k$ , but only one of its solutions represents a right traveling wave. The second root, for this model and for the parameter ranges tested, is backward traveling with  $\text{Im}(k) \gg 0$ , which makes it an evanescent wave of no interest except for a region very near the helicotrema. The wave amplitude has to be determined from a forced problem, which in this case is the solvability condition for the first-order WKB problem:

$$[Z_{ij}(z, \omega)]\{Y_{1,i}\} = F_{f_{1,0}}(z, \omega), \quad (8)$$

where the right-hand term is the part of the first-order hydrodynamic load that depends on the zeroth-order amplitude, i.e., the inhomogeneous part of the first-order term of the hydrodynamic load expansion; the homogeneous part will always result in convergent solutions. Since the determinant of the impedance matrix is equal to zero,  $|Z_{ij}| = 0$ , Eq. (8) will only have a solution if (Horn and Johnson, 1985)

$$(\mathbf{F}_{f_{1,0}}, \boldsymbol{\eta}_0) = \sum_{i=1}^2 F_{f_{1,0i}} \bar{\eta}_{0i} = 0, \quad (9)$$

where  $\tilde{\boldsymbol{\eta}}_0$  is the eigenvector of the adjoint matrix  $[Z_{ij}]^* = [\bar{Z}_{ij}]'$ , i.e., the complex conjugate of the transpose of  $[Z_{ij}]$ . This condition ensures that the solution of the first- and higher-order problems do not diverge by requiring orthogonality between the zeroth-order solution vector or eigenvector and the first-order forcing vector. The first-order hydrodynamic load vector was determined by applying Green's theorem to the zeroth- and first-order pressure fields over the cochlear duct cross section (Chadwick *et al.*, 1996) where it was shown that the first-order pressure term integrated over the partition width can be written as

$$\tilde{P}_{u,1} = \frac{i}{\rho\omega^2 Y_{T,0}} \frac{\partial}{\partial z} \left( kw \int_0^\infty \tilde{P}_{u,0}^2 dy \right) \quad (10)$$

with a corresponding relation for the lower fluid channel. Here,  $\tilde{P}$  is the pressure complex amplitude where the first subscript distinguishes between upper ( $u$ ) and lower ( $l$ ) chambers and the second subscript refers to the order in the WKB expansion. These relationships can be substituted into the solvability condition, Eq. (5), which can finally be expressed in terms of the zeroth-order pressures in the two scalae as

$$y_{BT} \frac{d}{dz} \left( kw \int_0^\infty P_{u,0}^2 dy \right) + \eta_{BT} \frac{d}{dz} \left( kw \int_{-\infty}^0 P_{l,0}^2 dy \right) = 0, \quad (11)$$

where  $w = W(z)/W_0$ ,  $y$  is the coordinate perpendicular to the subpartitions, and  $\eta_{BT}$  is the eigenfunction of the adjoint problem,  $[Z_{ij}]^* \{\eta\} = 0$ . Substituting the pressures derived from the oscillatory Stokes flow problem into Eq. (11) the latter can be integrated to establish the actual waveforms in terms of  $k(\omega, z)$  and  $\beta(\omega, z)$ , which can be written in the convenient form

$$Y_p(\omega, z) = A_p(\omega, z; \beta, k) \exp \left[ i \left( \omega t + \int_0^z \left( \Phi(\omega, z'; \beta, k) - \frac{k(\omega, z')}{\varepsilon} \right) dz' \right) \right], \quad (12)$$

where  $p = B$  or  $T$  and  $A_p$  and  $\Phi$  are functions of the mode ratios obtained from Eq. (1), which for the BM can be written as

$$A_{BM}(\omega, z; k, \beta) = \frac{c_0 k [1 - \sqrt{-i\nu/(\omega W_0^2)}] y_{BT}}{\rho\omega^2 \sqrt{w}}, \quad (13a)$$

$$\Phi(\omega, z; k, \beta) = i \frac{\eta_{BT} \partial_z y_{BT}}{1 + y_{BT} \eta_{BT}}, \quad (13b)$$

where  $c_0$  is an integration constant that was computed by matching the volume velocity at the stapes to the integrated volume velocity along the subpartition.

## B. Inverse problem for the CA transfer function

### 1. Use of *in vivo* data

The goal is to determine the characteristics of the CA which, when implemented in the present linear model, will result in a system whose response is similar to *in vivo* measured responses. These *in vivo* responses must correspond to

sound pressure level (SPL) inputs which are low enough so that the system may be assumed to behave linearly.

Two sets of *in vivo* data which were collected independently at different times using different animals and by different investigators were used for this study. The first set of data was taken from the early work of Rhode (Rhode, 1971) who used the Mössbauer technique to record from squirrel monkeys and the second set was obtained by courtesy of Nuttall and deBoer (unpublished) who recorded from guinea pigs using laser Doppler velocimetry (Nuttall and Dolan, 1996). The frequency response amplitude and phase data were used to fit the model predictions. It was thus necessary to obtain *in vivo* frequency response functions in the linear system regime which typically corresponds to input SPL of 30 dB or less. Unfortunately, such data were not directly available from the work of Rhode. It was therefore decided to adopt the method devised by Zweig (Zweig, 1991), who first extrapolated the collected amplitude versus SPL data to lower SPL where the cochlea is presumed to behave linearly. He thus obtained linear tuning curves from which corresponding causal propagating waves were reconstructed. To achieve this he had to assume a simple self-similarity relation among responses across the frequency spectrum. Although extrapolation to low SPL can always be questioned, it can be safely assumed that most qualitative features of the response were not seriously distorted and that the tuning curves obtained by that treatment are valid for our purposes here.

A critical element for the ultimate credibility of the inverse problem solution results is the choice of the parameters for the passive elements. The problem of lacking geometric and material property data for a complete cochlea is compounded here by the two degrees of freedom for the system, since it is ambiguous which geometric regions of the OC the two degrees-of-freedom correspond to. A reasonable resolution for this problem could be reached if the passive parameters of the model were chosen so that the theoretical predictions matched the linear *post-mortem* passive responses. A drawback of using Rhode's results was that no response data for the passive cochlea were available. This meant that comparison of our passive model response with *post-mortem* data to adjust model parameters was not possible. The second set of data was more extensive (Nuttall and deBoer, unpublished) and we used their passive data to adjust our model parameters. The passive model parameters were adjusted to obtain a reasonable fit of the passive frequency response amplitude and phase at one location. The quality of such a fit as shown in Fig. 2 is excellent for the amplitude except at the lower frequencies. A fairly good approximation for the phase was also achieved as seen in the same figure. These fits were achieved by trial-and-error and probably better fits could be achieved by more rigorous methods in the future. The model response was most sensitive to the cochlea length, to the damping parameters, and to the BM width and stiffness variations. The final passive model parameters used for the guinea-pig model are given at the end of the Appendix, item (1), along with those used for the squirrel monkey.

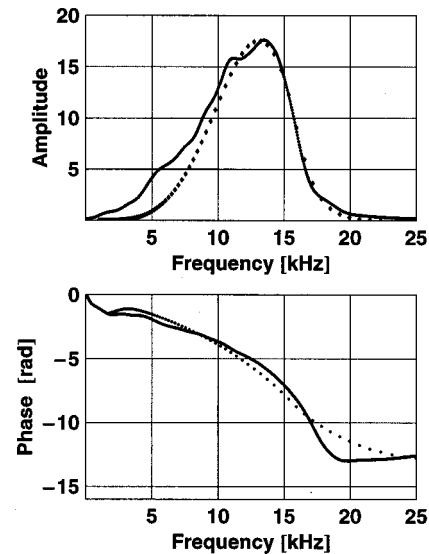


FIG. 2. Matching amplitude and phase frequency response for the passive guinea pig cochlea. The solid line represents the *in vivo* data and the dashed line the theoretical predictions of the present linear, passive model. Amplitudes are normalized by the stapes velocity.

## 2. Solution of the inverse problem

To obtain the frequency response curves which are needed as input for the inverse problem solution, the *in vivo* data collected by Rhode (Rhode, 1971) were treated entirely according to the method developed by Zweig (Zweig, 1991), who extrapolated nonlinear, high-SPL data to the lower, linear domain. We used frequency response curves at 25 dB SPL which were collected by Nuttall and deBoer using laser Doppler velocimetry similar to (Nuttall and Dolan, 1996). These frequency response data were used to reconstruct causal propagating waves by computing the corresponding wavenumber,  $k(\omega, z)$ . Because frequency response functions were measured only at one axial location, the complete 2-D  $k(\omega, z)$  function was computed using the simple self-similarity relation proposed by Zweig (Zweig, 1991). That relation postulates the existence of a universal, nondimensional variable,  $s$ ,

$$S = \frac{f}{f_c e^{-\alpha z}} \quad (14)$$

to which both  $\omega$  and  $z$  scale. The denominator is the cochlear frequency map for the specific animal (Greenwood, 1990). All response fields  $\Psi(\omega, z)$  can be expressed as functions of the single variable  $s$ , as  $\Psi(s)$ . This allows for the full two-dimensional field to be reconstructed from data collected from a single axial location.

With the frequency response curves available the formulation and solution of the inverse problem is relatively straightforward. First, the frequency response functions were substituted on the left-hand side of Eq. (12). The activity term on the right-hand side of the same equation can be eliminated by expressing the function  $\beta$ , defined in Eq. (3), in terms of  $k$  as  $\beta(\omega, z; k)$ . This is accomplished by solving the characteristic equation of the eigenvalue problem of Eq. (5) for  $\beta(\omega, z; k)$  after the active force from Eq. (3) is sub-



stituted into the equations of motion, Eqs. (2). Equation (12) thus becomes a rather complicated equation for the space-frequency map for the wave number that corresponds to an experimentally measured tuning curve and includes the modification of the passive wave by the CA. The inversion of this equation can be accomplished by differentiating the natural logarithm of both sides of Eq. (12) with respect to  $z$ . This leads to a first-order, nonlinear, ordinary differential equation for the CA-modified (or active) wave number,  $k = k(\omega, z)$ , in the form of a generalized Riccati equation. Once a solution for the wave number is obtained from this equation, the space-frequency CA map can be computed by back-substitution into the expression for  $\beta(\omega, z; k)$  and into Eq. (3). The equation is complicated by the fact that no simple closed form mathematical expression exists for the frequency response functions and by the lengthy expressions obtained for the differential equation coefficients. The task was accomplished by using the symbolic manipulation features of Mathematica (Wolfram Research). Simple algebra, however, indicates that an approximate solution is easily obtained by exploiting the presence of the small parameter  $\varepsilon$  in Eq. (12). After differentiation of the natural logarithm of Eq. (12) and rearrangement of its terms, the resulting equation can be written as

$$k = \varepsilon \Phi + \varepsilon \frac{\partial_z A}{A} - i\varepsilon \frac{\partial_z Y}{Y} \approx -i\varepsilon \frac{\partial_z Y}{Y}, \quad (15)$$

where  $\Phi$  and  $A$  are defined in Eqs. (13a) and (13b). The approximation in Eq. (15) is valid since  $\Phi$  is of order 1 and  $\partial_z A/A$  is of order  $\partial_z k/k$  which, according to the basic assumption of the WKB approximation, is of order  $\varepsilon$  away from the best place and may be order 1 in its vicinity, while  $\partial_z Y$ , by virtue of Eq. (1), is of order  $1/\varepsilon$ , at least near the best place. This first-order approximation of the active wave number is very simple to calculate since it involves no differential equation and no concomitant complications involving solution nonuniqueness for nonlinear differential equations. The validity of the approximation was tested by comparing the approximate wave number to the exact solution of the nonlinear Riccati differential equation. The differences were not qualitative in nature and the error generally remained less than 10% to 15% at all locations along the cochlear partition. Since the uncertainty in the values of many of the physical parameters in the model are at least equal or larger than this error in the estimated wave number, it was decided to try using the first-order approximation and avoid the solution of the Riccati equation. This was done after the resulting CA frequency response was computed using both the exact and approximate wave numbers and it was ascertained that the differences were again less or equal to about 10%.

## II. RESULTS

Figure 3 shows the magnitude and phase of the computed CA frequency and spatial response for the squirrel monkey. At low frequencies the CA looks very much like the low-pass filter of isolated OHCs. However, unlike the low-pass filter, the CA filter reaches a minimum near the best

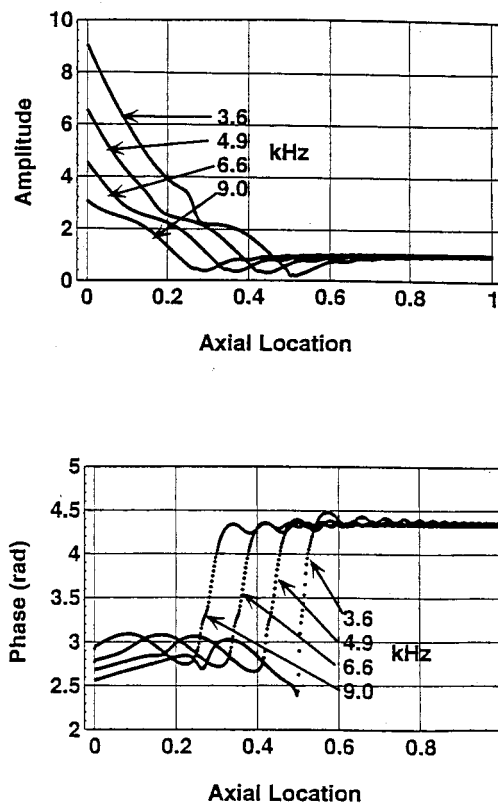


FIG. 3. Amplitude and phase of the cochlear amplifier transfer function as a function of axial location for several frequencies (kHz) for the squirrel monkey. Amplitude is normalized by the inertia of the basilar membrane per unit length.

place and remains practically constant thereafter. Clearly, the low-pass isolated OHC characteristic has been overridden by a different mechanism. Also note that the OHC activity minimum near the best location does not necessarily indicate the location of minimum power generation as that depends on the impedances on either side of the OHC against which the cell is doing work. Finally, the phase of the CA is clearly not  $-\pi/2$ , as one would expect for independently acting OHCs, but ranges from somewhat less than  $\pi$  to  $3\pi/2$  with respect to a TM upward displacement. These characteristics occur at frequencies where characteristic features also exist in the frequency response of other dynamical variables such as mode shapes. This means that the CA frequency response exhibits characteristic features which probably obey a similarity rule across space and frequency as do other dynamic responses. Such features cannot be achieved by an array of independently acting, low-pass OHCs. It is clear that the force applied by the CA is tonotopically tuned because it follows the system response itself. What is somewhat surprising here is that the filter that modulates that CA response [the function  $\beta(\omega, z)$ ] is itself tonotopic.

Figure 4 shows the corresponding result for the guinea pig. Due to the different treatment of the raw, guinea pig *in vivo* data these were much noisier and hence the fluctuations observed in the CA characteristics. It is again seen, however, that the low-pass character of the CA is overridden by a different mechanism as the best frequency is approached and



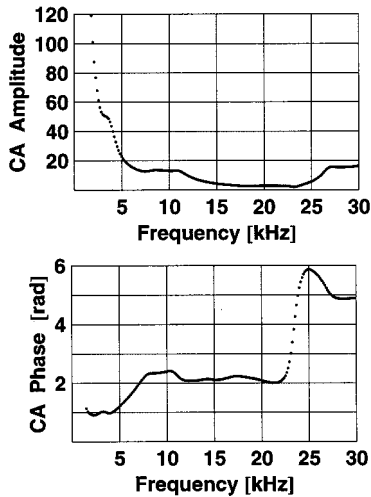


FIG. 4. Amplitude and phase of the cochlear amplifier transfer function as a function of axial location for several frequencies (kHz) for the guinea pig. The plotted curves represent the impedance part of the expression in Eq. (3).

this mechanism persists beyond that frequency.

Figures 3 and 4 strongly suggest that the CA looks like a tonotopic mechanism as opposed to a CA consisting of independently acting OHCs whose low-pass characteristics vary insignificantly along the cochlear axis. Here it is seen that the CA frequency response possesses features that parallel the tonotopic variation of response fields. For both sets of data, the CA filter amplitude reaches a minimum around the best place and its phase has a definite transition somewhere beyond the best place. Of course the tonotopicity shown in Fig. 3 across the different frequencies is the result of the self-similarity relation adopted in Eq. (14). Even if that relationship is not very accurate, it is generally accepted that self-similarity does exist in some form and that the use of Eq. (14) does not lead to qualitatively wrong conclusions.

The validity of the CA properties that were computed from the inverse problem was tested by reconstructing the model response given the computed CA from the inverse problem. Again, for computational efficiency the CA frequency response used in these computations was the first-order approximation which is valid for frequencies down to about 2 kHz. Figure 5 shows the resulting tuning curves for the squirrel monkey which agree closely, with respect to most features, to the observed *in vivo* responses. Figure 6 shows a tuning curve for the guinea pig. Again the match is remarkable especially around the best frequency. Away from that, the accuracy of the WKB approximation deteriorates and that is probably in part the reason for the discrepancies on either side of the best frequency. These results show that the approximation used to compute the active wave number is sufficiently accurate.

The robustness of the above results was examined extensively during the study by computing wave numbers and resulting CA frequency responses for wide ranges of model parameters. The results were remarkably insensitive to most model parameters. The parameters to which the sensitivity was considerable were the cochlea length and the stiffness

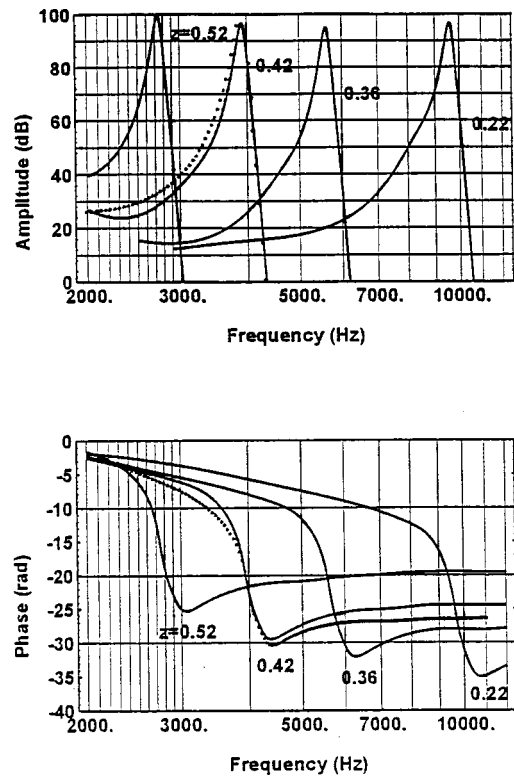


FIG. 5. Reconstructed amplitude and phase of basilar membrane response as a function of frequency for several axial locations (solid lines). For one axial location, the experimentally measured frequency response curve (Rhode, 1971) is shown (dotted lines). These represent typical input data for the inverse problem.

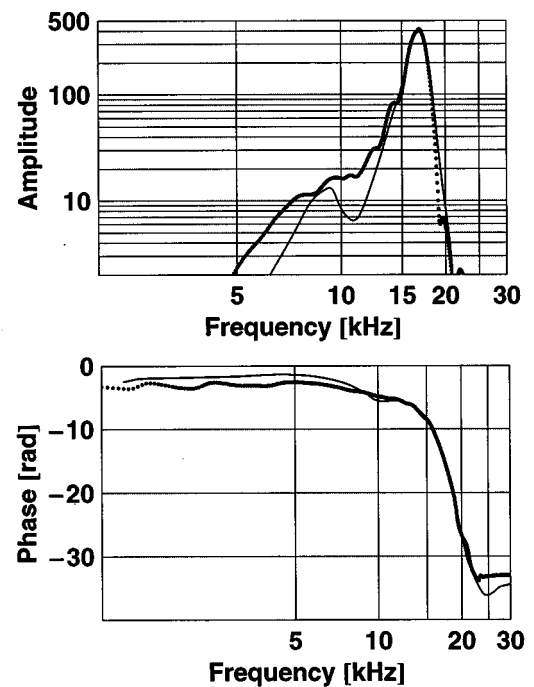


FIG. 6. Reconstructed amplitude and phase of basilar membrane response as a function of frequency for  $z=0.2$  (solid lines). The experimentally measured frequency response curve by Nuttall and deBoer (unpublished data) which was obtained by methods described in (Nuttall and Dolan, 1996) is also shown (dotted lines). Amplitudes are normalized by the stapes velocity.

variation parameters of the BM. Fortunately, the cochlea length is accurately known for both animals referred to here, as is the axial variation of the BM width. The stiffness variation parameters were adjusted so that the model resulted in an active cochlear map as close as possible to the map that has been measured and presented in the literature (Greenwood, 1990), while at the same time, those same system parameters resulted in passive frequency response closely matching the *in situ* experimental data used here.

### III. DISCUSSION

An important issue relating to the validity of any theoretical prediction concerning the function of the CA is the magnitude of the force that each OHC is required to apply. For characteristic frequencies, using Eq. (3) it was estimated that the active pressure is of the order of 5 mN/cm<sup>2</sup> which can be converted to force per cell by taking into account the BM width, the OHC spacing, and the three rows of OHCs. The resulting force is of the order of 20 pN/cell. If it is assumed that the transmembrane potential oscillates with about 0.1 mV amplitude and that the cell can apply about 100 pN of force for each mV transmembrane potential variation (Iwasa and Adachi, 1997), the force is about 10 pN cell when measured against a very high impedance (immeasurably small length change). This value is reasonably close to the force predicted here. By way of comparison, another two-degree-of-freedom cochlear model (Neely, 1993) predicted active pressure of the order of 150 mN/cm<sup>2</sup> and the ratio of the two masses was  $M_T/M_B=62$  as contrasted with the present study where that ratio is equal to a more reasonable 0.75. A very recent two-degrees-of-freedom linear model (Fukazawa, 1997) predicts 4 mN/cm<sup>2</sup>, a prediction very close to that of the present model. The active pressure predictions from most other published models were even higher. For example, the models by Mammano and Nobili (1993), Kolston (1995), and Kanis and deBoer (1996) predict active pressures between 200 and 600 mN/cm<sup>2</sup>. deBoer's model (1995b), based on a one-degree-of-freedom model, predicts a force of about 3500 mN/cm<sup>2</sup>. Still other models (Kolston *et al.*, 1989; Geisler and Sang, 1995) predict much higher forces. These numbers need to be evaluated based on experimental measurements of the force generation capacity of OHCs.

Previous difficulties encountered by single- and multi-partition models because of the low-pass, independently acting OHCs were partially overcome either by adopting unrealistic lumped parameter distributions (Neely, 1993) or by postulating mechanisms such as the radial TM resonance (Zwislocki and Kletsy, 1979; Allen, 1980) or the OHC feed-forward effect due to their axial tilt (Geisler and Sang, 1995). It is difficult, however, to see how the latter two mechanisms could modify the low-pass OHC filter to the one presented in Figs. 3 and 4. In the case of single-partition models where the energy pathways are clearly defined, these mechanisms could be made to work. In the multi-component cross-section model, however, energy pathways are more complex, and the CA characteristics required to achieve correct tuning are not at all intuitively obvious. It should be noted that in the approximation given by Eq. (15), the CA-

modified wave number is independent of system parameters as it can be described purely on the basis of the propagating wave, irrespective of the slow axial variations along the cochlea. This is probably an indication that the required CA characteristics would primarily be the result of the system response itself. Hence, the suggested CA would be a distributed electromechanical dynamical system and not a static structure imprinted on the CP. The natural next step is to elucidate the physiological mechanisms underlying such a CA transfer function.

A mechanism has recently been proposed by which large enough OHC motile responses can be achieved *in vivo*. Instead of the attenuated intracellular potential oscillation caused by stereociliary ion channel modulations, the transmembrane potential might be modulated by fast extracellular potential oscillations generated by ion pathways within the OC and without encountering the cell membrane barrier (Dallos and Evans, 1995a, b). Such a feature of the CA essentially implies some type of, probably electrical and/or mechanical, coupling between OHCs. It has recently been demonstrated that fluorescent dyes injected into an OHC find their way into nearby OHCs (Zwislocki *et al.*, 1992). Die pathways through Deiter and Hensen cell assemblies suggest physical pathways among the various cell groups. Although other observations are not consistent with physical pathways such as gap junctions, between hair cells and supporting cells (Santos-Sacchi and Dallos, 1983; Santos-Sacchi, 1986), the above findings (Zwislocki *et al.*, 1992; Dallos and Evans, 1995a), including those in the present paper, call for renewed investigations of electrical pathways in the cochlea.

#### A. Proposed CA electrical pathways

The evidence from previous research and our finding from the inverse problem solution suggest that the CA is a nonlocally sensing transducer. Its tonotopic tuning suggests that it senses features of the wave and not merely a local displacement. The reason behind this suggestion lies in the surprising computational observation that the CA filter of Eq. (3) could be qualitatively represented as proportional to the local wavelength

$$\beta(\omega, z) \propto k(\omega, z)^{-1}. \quad (16)$$

According to this observation, OHCs are required to receive information about remote events since the wavelength cannot be sensed purely locally. In fact, the approximation in Eq. (16) was best for a power of the wavelength slightly higher than one. This holds true for both the magnitude and phase of the CA transfer function. An effort was made to reconcile these observations with possible underlying mechanisms. Having failed to think of a possible structural feature that could account for the computed CA transfer function (see discussion at the end of this section), we decided to revisit the issue of electrical pathways within the CP and seek possible electrical circuits that might manifest the desired features. After some computational experimentation with different electrical circuits, we propose the circuit for the CA shown in Fig. 7. It consists of three axial cables transversely connected by the OHC electrical impedances. The cables represent, from top to bottom, two extracellular spaces, and

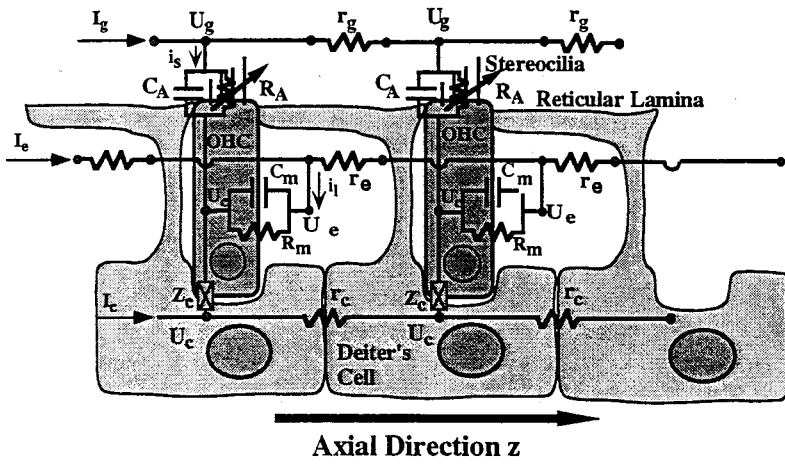


FIG. 7. Proposed cable circuit of the cochlear amplifier with axial coupling. The impedance element  $Z_c$  is included in the figure but not in the equations. Computations indicated that its presence, whether capacitive or resistive, had little qualitative effect on the frequency response of the OHC transmembrane potential.

connections between OHCs through the underlying Deiter's cells. A constant dc potential,  $U_c - U_e$ , is maintained between the two extracellular spaces by the stria vascularis. The activation of the CA is caused by the wave traveling on the CP, which modulates the stereocilia resistance [see Eq. (A10)] in a pattern mirroring the shape of the wave as given by Eq. (1). To endow the CA with tonotopic spatial features not achievable by arrays of independent cells, the circuit space-frequency response now depends on the traveling wave. To show how this is accomplished, one needs to formulate and solve the set of three coupled cable differential equations that describe the voltage wave propagation within the CA. Such a voltage wave needs to travel along with the mechanical wave. The related mathematical formulation is given in the Appendix, item (2).

The main obstacle to testing the performance of this circuit comes from the lack of experimental data on the magnitudes of the various electrical parameters in the proposed CA circuit model. In particular, values for the axial cable resistances ( $r_g, r_e, r_c$ ) can only be crudely estimated. Although the resistance and capacitance of the OHC membrane have been extensively measured and discussed, the nature and magnitude of electrical impedances presented to the extracellular space and to their points of attachment to other structures within the OC remain unclear. Reasonable, order-of-magnitude guesses can be made for the top two axial cables based on geometry of the extracellular spaces and on estimates of the fluid specific conductances. The lower cable representing an electrical pathway from OHC to OHC through the supporting cells is being postulated here. Neither its nature nor its conductance magnitude are known with any certainty. Although there is some supporting evidence of existence of a conductive pathway between OHCs and Deiter's cells (Zwislocki *et al.*, 1992), other experiments and morphological studies have, so far, failed to reproduce the finding (Santos-Sacchi and Dallos, 1983; Santos-Sacchi, 1986). When a capacitive transmembrane connection between OHCs and Dieter cells was assumed, the changes in the circuit characteristics were quite insignificant. If the presence of such pathways were to be confirmed, their quantification would provide one more reliable parameter for the CA circuit model.

Computations were made to establish the transfer function of the proposed circuit and subsequently the transfer function of the transmembrane potential. Parameter values were sought that would result in transmembrane transfer function, qualitatively similar to an inverse power of the wave number as indicated by Eq. (16). It was found that for reasonable parameter values the transfer function could be made similar to a power between one and two of the inverse of the wave number. As can be easily verified by observation of the determinant of Eq. (A15), the CA frequency response is a ratio of a second-order to a fourth-order polynomial of the wave number. It is of interest to note that when the lower cable is omitted, the order of the polynomials changes, and the resulting CA loses its tonotopic features. Hence it is felt at this point that all three cables are vital to the proposed model function. The final magnitude and phase of the transmembrane potential transfer function for the three-cable model are shown in Fig. 8. The qualitative similarity of the magnitude of this function to the magnitude of the CA filter in Fig. 4 is remarkable. Modulation amplitudes for the transmembrane potential of the order of 0.1 to 10 mV was achievable when the circuit parameters were constrained according to

$$\left\{ \frac{\kappa^2 R_m}{r_g}, \frac{\kappa^2 R_m}{r_e}, \frac{\kappa^2 R_m}{r_c} \right\} \gg \omega R_m C_m, \quad (17)$$

where  $\kappa$  is the local dimensional wave number and it is introduced in the Appendix, item (2), in connection with the stereocilia resistance modulation. This inequality states that the ratio of the extracellular to the OHC lateral membrane conductance must be much larger than the phase lag introduced by the lateral OHC membrane. Notice that the electrotonic space constant of the circuit cables is the square root of the ratio of the transverse to the axial resistance, namely,

$$\lambda_s^2 = \frac{R}{r}. \quad (18)$$

Inequality (17) can now be rewritten in the generic form

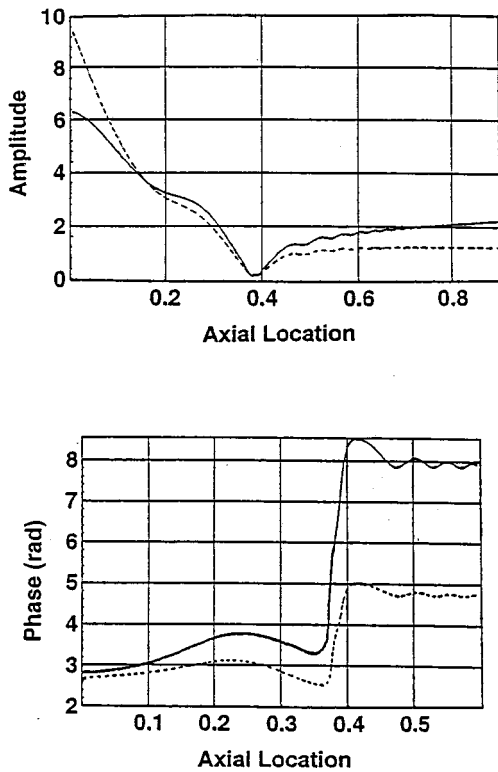


FIG. 8. Comparison of amplitude and phase of the transmembrane potential transfer function for the proposed cochlear amplifier circuit shown in Fig. 7 (solid lines) and of the corresponding functions computed from the inverse problem (dashed lines), both at 5 kHz. The parameter values for the electrical impedance elements were as follows:  $R_a = 10^9 \Omega$ ,  $R_m = 5 \times 10^6 \Omega$ ,  $C_a = 5 \times 10^{-12} \text{ F}$ ,  $C_m = 2 \times 10^{-12} \text{ F}$ ,  $r_s = 15 \times 10^3 \Omega/\text{m}^2$ ,  $r_e = 10^3 \Omega/\text{m}^2$ , and  $r_c = 1 \Omega/\text{m}^2$ .

$$(\kappa \lambda_s)^2 \gg 2\pi f \tau, \quad (19)$$

which allows a lower bound estimate of  $\lambda_s$  for a given frequency  $f$ . If it is assumed that the smallest wavelength of the traveling wave on the CP is  $\lambda_w = (2\pi/\kappa) \approx 250 \mu\text{m}$  and that the time constant of the OHC lateral membrane is of the order of  $\tau \approx 1 \text{ ms}$ , then at 1 kHz  $\lambda_s \gg 100 \mu\text{m}$ , which corresponds to a minimum axial extent of four to five OHC rows. The actual number may potentially be much larger, indicating the importance of OHC synergy for proper CA function. The phase of the transmembrane transfer function, at low frequencies or basal locations, was also similar to the one computed for the CA from the inverse problem. The phase jump occurs at the same place but it is almost twice as large as that of the inverse problem. This may be of no consequence as it is expected that the higher phase values are reached after the wave cutoff. Still, a correction for the phase discrepancy is possible if one speculates that the stereocilia W-shaped bundle orientation at a slight angle to the radial direction would cause a component of resistance modulations which are proportional to the wave slope rather than the wave amplitude itself. Then the transmembrane potential varies nearly as  $\kappa^{-1}(\omega, z)$ , and the magnitude stays qualitatively the same as before but the phase is much closer to that of the inverse problem. This is a possible form of OHC me-

chanical coupling in addition to the electrical coupling within the proposed CA electrical circuit. Another possible form of such synergy may result from the mechanical coupling of neighboring OHCs due to the anchoring of the stereocilia to the underside of the TM. Such coupling, although it could mirror the waveform, should not be effective because of the great compliance magnitude difference between the stereocilia and the TM local structure.

Clearly, these theoretical findings concerning the proposed nature and circuitry of the CA cannot be considered either as conclusive or complete, but are intended to stimulate and challenge experimenters to search for new electro-mechanical connections in the cochlea.<sup>2</sup>

## APPENDIX

(1) Mechanical impedance relations: The cochlear model impedance matrix can be written as

$$[Z_{ij}] = \begin{bmatrix} Z_{11}(\omega, z) - \alpha(\omega, z) - \frac{W(z)}{H_b k_w} & -Z_{12}(\omega, z) \\ \alpha(\omega, z) - Z_{12}(\omega, z) & Z_{22}(\omega, z) - \frac{W(z)}{H_b k_w} \end{bmatrix}, \quad (A1)$$

where  $W(z)$  is the BM half-width,  $H_b$  is the effective thickness of the BM,  $k_w$  is the nondimensional wave number normalized to  $W(z)$ , and  $\alpha(\omega, z)$  is the CA activity term defined by [cf. Eq. (3)]

$$\alpha(\omega, z) = \frac{W(z)}{H_b} \frac{\beta(\omega, z)}{1 + i\omega\tau}. \quad (A2)$$

The remaining impedance elements were defined by dividing the equations of motion by the BM mass per unit length as follows:

$$Z_{11} = \left( \frac{\omega_{\text{RL}}}{\omega} \right)^2 + \left( \frac{\omega_{\text{OHC}}}{\omega} \right)^2 + \left( \frac{\omega_{\text{TM}}}{\omega} \right)^2 - m_{\text{RB}} - m_{\text{TB}} + i(\zeta_{\text{TM}}(\omega) + \zeta_{\text{RL}}(\omega)) - \nu_c(\omega), \quad (A3)$$

$$Z_{12} = \left( \frac{\omega_{\text{RL}}}{\omega} \right)^2 + \left( \frac{\omega_{\text{OHC}}}{\omega} \right)^2 + i\zeta_{\text{RL}}(\omega), \quad (A4)$$

$$Z_{22} = \left( \frac{\omega_{\text{BM}}}{\omega} \right)^2 + \left( \frac{\omega_{\text{RL}}}{\omega} \right)^2 + \left( \frac{\omega_{\text{OHC}}}{\omega} \right)^2 - 1 + i(\zeta_{\text{RL}}(\omega) + \zeta_{\text{BM}}(\omega)) - \nu_c(\omega), \quad (A5)$$

where the subscripted  $\omega$ 's are the square roots of the ratios of the corresponding stiffnesses and the BM mass per unit length,  $m_{\text{RB}}$  and  $m_{\text{TB}}$  are mass ratios of RL and TM, respectively, to the BM mass,  $\zeta_i$  are the viscous damping terms for the BM, RL, and TM defined as  $\zeta_i = c_i / (2\omega \rho_{\text{BM}} W(z) H_b)$ , and  $\nu_c(\omega)$  is the boundary layer viscous correction term defined as  $\nu_c(\omega) = (1 - i)\sqrt{\nu/(2\omega)}$  (Batchelor, 1966).



The various parameters used throughout for the squirrel monkey and for the guinea pig are shown in the table below:

	Squirrel monkey	Guinea pig
$W_0$	84 $\mu\text{m}$	60 $\mu\text{m}$
$H_b$	140 $\mu\text{m}$	140 $\mu\text{m}$
$\rho$ (fluid)	1 $\text{kg/m}^3$	1 $\text{kg/m}^3$
$\rho_b$	1.05 $\text{kg/m}^3$	1.05 $\text{kg/m}^3$
$\mu = \nu\rho$	0.0015 $\text{kg/m s}$	0.0015 $\text{kg/m s}$
$W(z)$	$W_0(1+5z)$	$W_0(1+2.6z)$
$L_c$	30 mm	18 mm
$f_{\text{BM}} = \omega_{\text{BM}}/2\pi$	26000 $\exp(-4.5z)$	45000 $\exp(-6z)$
$f_{\text{TM}} = \omega_{\text{TM}}/2\pi$	4000 $\exp(-4.5z)$	12000 $\exp(-z)$
$f_{\text{RL}} = \omega_{\text{RL}}/2\pi$	1100 $\exp(-z)$	5000 $\exp(-z)$
$f_{\text{OHC}} = \omega_{\text{OHC}}/2\pi$	2000 $\exp(-z)$	3000 $\exp(-0.25z)$
$m_{\text{RB}} = m_{\text{RL}}/m_{\text{BM}}$	0.25	0.15
$m_{\text{TB}} = m_{\text{TM}}/m_{\text{BM}}$	0.5	0.25
$c_i$	0.05 N s/m <sup>2</sup>	0.75 N s/m <sup>2</sup>
Cochlear map	45 000 $\exp(-2z)$	45 000 $\exp(-5.16z)$

(2) Electrical cable equations: The axial currents and voltages in the circuit shown in Fig. 7 are related through

$$r_g I_g = -\frac{\partial U_g}{L_c \partial z}, \quad r_e I_e = -\frac{\partial U_e}{L_c \partial z}, \quad r_c I_c = -\frac{\partial U_c}{L_c \partial z}, \quad (\text{A6})$$

where  $L_c$  is the cochlea length and  $z \in [0,1]$  is the nondimensional axial location. The axial current rates of change are caused by the cell (transverse) currents which, according to Kirchoff's law, may be written as

$$\begin{aligned} i_s &= \frac{U_g - U_c}{R_A} + C_A \frac{\partial}{\partial t} (U_g - U_c) - \frac{V_g - V_c}{R_a}, \\ i_l &= \frac{U_e - U_c}{R_m} + C_m \frac{\partial}{\partial t} (U_e - U_c) - \frac{V_e - V_c}{R_m}, \end{aligned} \quad (\text{A7})$$

where the  $V$ 's signify corresponding voltage dc components. Current balance requires that transverse currents be related to axial ones by

$$i_s = -\frac{\partial I_g}{L_c \partial z}, \quad i_l = -\frac{\partial I_e}{L_c \partial z}, \quad i_s + i_l = \frac{\partial I_c}{L_c \partial z}. \quad (\text{A8})$$

Differentiation of Eqs. (A6), substitution into Eqs. (A8) and combination with Eqs. (A7) results in the final cable equations:

$$C_A \frac{\partial}{\partial t} (U_g - U_c) - \frac{\partial}{L_c^2 \partial z} \left[ \frac{1}{r_g} \frac{\partial U_g}{\partial z} \right] + \frac{U_g - U_c}{R_A} = \frac{V_g - V_c}{R_a}, \quad (\text{A9a})$$

$$C_M \frac{\partial}{\partial t} (U_e - U_c) - \frac{\partial}{L_c^2 \partial z} \left[ \frac{1}{r_c} \frac{\partial U_c}{\partial z} \right] + \frac{U_e - U_c}{R_M} = \frac{V_e - V_c}{R_M}, \quad (\text{A9b})$$

$$\begin{aligned} C_m \frac{\partial}{\partial t} (U_e - U_c) + C_A \frac{\partial}{\partial t} (U_g - U_c) + \frac{\partial}{L_c^2 \partial z} \left[ \frac{1}{r_c} \frac{\partial U_c}{\partial z} \right] \\ + \frac{U_e - U_c}{R_m} + \frac{U_g - U_c}{R_A} = \frac{V_e - V_c}{R_m} + \frac{V_g - V_c}{R_a} = 0. \end{aligned} \quad (\text{A9c})$$

The final cable equations are second-order partial differential equations (Rall, 1977) where Eq. (A9a) describes the current flow in the extracellular space between the stereocilia, Eq.

(A9b) describes the current flow through the OHC lateral membranes into the extracellular spaces within the OC, and Eq. (A9c) describes a transcellular current flow at the OHC bases which is here postulated to represent the communication link between OHCs through the other cells (Deiter and Hensen cells) in the OC. The symbols used are explained in connection with Fig. 7. A dc voltage is maintained across the OHCs and the various other potentials are perturbed and modulated because of the stereocilia resistance modulation which can be written as

$$R_A(t, z; \omega) = R_a \left( 1 + \delta \exp \left[ i \left( \omega t - L_c \int_0^z \kappa dz \right) \right] \right), \quad (\text{A10})$$

where  $\kappa$  is the dimensional wave number and  $\delta$  is a small parameter. This will induce voltage modulations across all circuit segments. The voltages then can be written in terms of the small parameter  $\delta$  as follows:

$$U_e(t, z; \omega) = V_e + \delta v_e \exp \left[ i \left( \omega t - L_c \int_0^z \kappa dz \right) \right],$$

$$U_c(t, z; \omega) = V_c + \delta v_c \exp \left[ i \left( \omega t - L_c \int_0^z \kappa dz \right) \right], \quad (\text{A11})$$

$$U_g(t, z; \omega) = V_g + \delta v_g \exp \left[ i \left( \omega t - L_c \int_0^z \kappa dz \right) \right].$$

When Eqs. (A10) and (A11) are substituted into Eqs. (A9), these can be expanded in terms of the small parameter  $\delta$ . The zeroth-order terms for the first two equations are identically satisfied but the third relates the dc voltages at the three OHC electrical nodes of Fig. 7. The latter may be solved for the dc voltage at the OHC base:

$$V_c = \frac{R_a V_e + R_m V_g}{R_a + R_m}. \quad (\text{A12})$$

The potential across the whole cell,  $V_g - V_c$ , or between the extracellular space in the scala vestibuli and the intercellular spaces among the various cell groups within the OC is maintained by the high impedance voltage source at the stria vascularis which keeps it constant at about 80 mV. One of the two voltages can be set to zero as the reference voltage and, therefore, all dc voltages are known. The first-order terms of the expansion give a set of three algebraic equations for the small voltage modulation amplitudes which, in matrix form, can be written as

$$[\mathbf{A}_v] \{v\} = \{b\}, \quad (\text{A13})$$

where

$$\{v\} = \begin{Bmatrix} v_g \\ v_e \\ v_c \end{Bmatrix}, \quad \{b\} = \begin{Bmatrix} 0 \\ r_g(V_e - V_g)/R_a \\ R_a(V_g - V_c)/R_m \end{Bmatrix}, \quad (\text{A14})$$

and

$$[A_v] = \begin{bmatrix} 0 & i\omega r_c C_M + \frac{r_c}{R_M} & -\kappa^2 - i\omega r_c C_M - \frac{r_c}{R_M} \\ -\kappa^2 - i\omega r_g C_A - \frac{r_g}{R_a} & i\omega r_g C_A + \frac{r_g}{R_a} & 0 \\ 1 + i\omega R_a C_A & -1 - i\omega R_a C_A - i\omega R_a C_M - \frac{\kappa^2 R_a}{r_e} & i\omega R_a C_M + \frac{R_a}{R_M} \end{bmatrix}. \quad (\text{A15})$$

This set of equations has a unique solution for the voltage modulation amplitude vector. The difference  $v_c - v_0$  is, within a constant  $\delta$ , the OHC transmembrane potential modulation which controls the motile action of these cells. This is, therefore, the main quantity of interest since, as already discussed, the active force generated by the CA is assumed to be proportional to the transmembrane potential.

<sup>1</sup>Function arguments following the semi-colon are included for clarity only since they themselves are functions of the arguments preceding the semi-colon.

<sup>2</sup>Some axial interactions between neighboring OHCs have been recently experimentally demonstrated by Zhao and Santos-Sacchi (1999).

Allen, J. B. (1980). "Cochlear micromechanics—a physical model of transduction," *J. Acoust. Soc. Am.* **68**, 1660–1670.

Allen, J. B., and Fahey, P. F. (1992). "Using acoustic distortion products to measure cochlear amplifier gain on the basilar membrane," *J. Acoust. Soc. Am.* **92**, 178–188.

Ashmore, J. F. (1990). "Forward and reverse transduction in the mammalian cochlea," *Neurosci. Res. Suppl.* **12**, S39–S50.

Batchelor, G. K. (1966). *An Introduction to Fluid Mechanics* (Cambridge U.P., Cambridge, England).

Brownell, W. E., Bader, C. R., Bertrand, B., and de Ribaupierre, Y. (1985). "Evoked mechanical responses of isolated cochlear outer hair cells," *Science* **227**, 194–196.

Chadwick, R. S., Dimitriadis, E. K., and Iwasa, C. H. (1996). "Active control of waves in a cochlear model with sub-partitions," *Proc. Natl. Acad. Sci. USA* **93**, 2564–2569.

Dallos, P., and Evans, B. N. (1995a). "High frequency motility of outer hair cells in the cochlear amplifier," *Science* **267**, 2006–2009.

Dallos, P., and Evans, B. N. (1995b). "High frequency motility: Corrections and addendum," *Science* **268**, 1420–1421.

Davies, H. (1983). "An active process in cochlear mechanics," *Hearing Res.* **9**, 79–90.

deBoer, E. (1995a). "The inverse problem solved for a three dimensional model of the cochlea. I. Analysis," *J. Acoust. Soc. Am.* **98**, 896–903.

deBoer, E. (1995b). "The inverse problem solved for a three dimensional model of the cochlea. II. Application to experimental data," *J. Acoust. Soc. Am.* **98**, 904–910.

Fukazawa, T. (1997). "A model for cochlear micromechanics," *Hearing Res.* **113**, 182–190.

Geisler, C. D., and Sang, C. (1995). "A cochlear model using feed-forward outer hair cell forces," *Hearing Res.* **86**(1), 132–146.

Greenwood, D. G. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.

Gold, T. (1948). "The physical basis of the action of the cochlea," *Proc. R. Soc. London, Ser. B* **135**, 492–498.

Gummer, A. W., Hemmert, W., and Zenner, H. P. (1996). "Resonant tectorial motion in the inner ear: Its crucial role in frequency tuning," *Proc. Natl. Acad. Sci. USA* **93**, 8727–8732.

Horn, R. A., and Johnson, C. R. (1985). *Matrix Analysis* (Cambridge U.P., Cambridge, England).

Housley, G. D., and Ashmore, J. F. (1992). "Ionic currents of outer hair cells and cochlear amplifier," *J. Physiol. (London)* **448**, 73–98.

Hubbard, A. (1993). "A traveling wave amplifier model of the cochlea," *Science* **259**, 68–71.

Hudspeth, A. J., and Corey, D. P. (1977). "Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled me-

chanical stimuli," *Proc. Natl. Acad. Sci. USA* **74**, 2407–2411.

Iwasa, K. H., and Adachi, M. (1997). "Force generation in the outer hair cell of the cochlea," *Biophys. J.* **73**(1), 546–555.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (1986). "Basilar membrane measurements and the traveling wave," *Hearing Res.* **22**, 147–153.

Kachar, B., Brownell, W. E., Altschuler, R. A., and Fex, J. (1986). "Electrokinetic changes in cochlear outer hair cells," *Nature (London)* **322**, 365–368.

Kanis, I. J., and deBoer, E. (1996). "Comparing frequency-domain and time-domain solutions for a locally active non-linear model of the cochlea," *J. Acoust. Soc. Am.* **100**, 2543–2546.

Khanna, S. M., and Leonard, D. G. B. (1982). "Basilar Membrane Tuning in the Cat Cochlea," *Science* **215**, 305–306.

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

Kolston, P. J. (1995). "A faster transduction mechanism for the cochlear amplifier," *Trends Neurosci.* **18**, 427–429.

Kolston, P. J., Viergever, M. A., and deBoer, E. (1989). "Realistic mechanical tuning in a micromechanical cochlear model," *J. Acoust. Soc. Am.* **86**, 133–140.

Lieberman, M. C., and Dodds, L. W. (1984). "Single neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," *Hearing Res.* **16**, 55–74.

Lighthill, M. J. (1981). "Energy flow in the cochlea," *J. Fluid Mech.* **106**, 149–213.

Lighthill, M. J. (1992). "Acoustic streaming in the ear itself," *J. Fluid Mech.* **239**, 551–606.

Mammano, F., and Ashmore, J. F. (1995). "A laser interferometer for sub-nanometer measurements in the cochlea," *J. Neurosci. Methods* **60**, 89–94.

Mammano, F., and Nobili, R. (1993). "Biophysics of the cochlea: Linear approximation," *J. Acoust. Soc. Am.* **93**, 3320–3332.

Markin, V. S., and Hudspeth, A. J. (1995). "Modeling the active process of the cochlea: Phase relations, amplification, and spontaneous oscillation," *Biophys. J.* **69**, 138–147.

Neely, S. T. (1993). "A model of cochlear mechanics with outer hair cell motility," *J. Acoust. Soc. Am.* **94**, 137–146.

Nuttall, A. L., and Dolan, D. F. (1996). "Steady-state sinusoidal velocity responses of the basilar membrane in the guinea pig," *J. Acoust. Soc. Am.* **99**, 1556–1565.

Nuttall, A. L., Dolan, D. F., and Avinash, G. (1991). "Laser Doppler Velocimetry of basilar membrane vibration," *Hearing Res.* **51**, 203–214.

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

Rall, W. (1977). "Core conductor theory and cable properties of neurons," in *Handbook of Physiology, Section I: The Nervous System, Part I, Cellular Biology of Neurons*, edited by J. M. Brookhart and V. B. Mountcastle (American Physiological Society, Bethesda, MD), pp. 39–97.

Rhode, W. J. (1971). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," *J. Acoust. Soc. Am.* **49**, 1218–1231.

Robles, L., Ruggero, M. A., and Rich, N. C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlea I. Input-output functions, tuning curves and response phases," *J. Acoust. Soc. Am.* **80**, 1364–1374.

Ruggero, M. A., and Rich, N. C. (1991a). "Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar membrane vibration," *Hearing Res.* **51**, 215–230.

Ruggero, M. A., and Rich, N. C. (1991b). "Furosemide alters organ of Corti mechanics: Evidence for feedback and of outer hair cells upon the basilar membrane," *Neuroscience* **11**, 1057–1067.

- Russell, I. J., and Nilsen, K. E. (1997). "The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane," *Proc. Natl. Acad. Sci. USA* **94**(6), 2660–2664.
- Santos-Sacchi, J. (1986). "Dye coupling in the organ of Corti," *Cell Tissue Res.* **245**, 525–529.
- Santos-Sacchi, J. (1992). "On the frequency limit and phase of outer hair cell motility: Effects of the membrane filter," *J. Neurosci.* **12**, 1906–1916.
- Santos-Sacchi, J., and Dallos, P. (1983). "Intercellular communication in the supporting cells of the organ of Corti," *Hearing Res.* **9**, 317–326.
- Sellick, P. M., Patuzzi, R., and Yates, G. K. (1986). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.* **72**, 131–141.
- Ulfendahl, M., Flock, A., and Khanna, S. M. (1989). "A temporal bone preparation for the study of cochlear micromechanics at the cellular level," *Hearing Res.* **40**, 55–64.
- Xue, S., Mountain, D. C., and Hubbard, A. E. (1995). "Electrically-evoked basilar membrane motions," *J. Acoust. Soc. Am.* **97**, 3032–3041.
- Zhao, H.-B., and Santos-Sacchi, J. (1999). "Auditory collusion and a coupled couple of outer hair cells," *Nature* **399**, 359–362.
- Zweig, G. (1991). "Finding the impedance of the organ of Corti," *J. Acoust. Soc. Am.* **89**, 1229–1254.
- Zwislocki, J. J., and Kletschy, E. J. (1979). "Tectorial membrane: A possible effect on frequency analysis in the cochlea," *Science* **204**, 639–641.
- Zwislocki, J. J., Slepceky, N. B., Cefaratti, L. K., and Smith, R. L. (1992). "Ionic coupling among cells in the organ of Corti," *Hearing Res.* **57**, 175–194.

# Modeling otoacoustic emissions by active nonlinear oscillators

R. Sisto<sup>a)</sup>

Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1,  
00040 Monte Porzio Catone (Roma), Italy

A. Moleti<sup>b)</sup>

Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1,  
00133 Roma, Italy

(Received 11 March 1999; accepted for publication 17 June 1999)

The phenomenology of spontaneous otoacoustic emissions (OAEs) is compared to theoretical predictions given by models in which they are produced by active nonlinear oscillators. Along with the well-known Van der Pol oscillator, a new active oscillator model is proposed and analyzed here. Numerical simulations and multi-scale analytical computation results are compared to experimental data of neonatal spontaneous and evoked OAEs. A simple analysis technique is proposed, in which the time evolution after a click stimulus of the amplitude of each spectral line corresponding to a spontaneous OAE is studied. Apart from a few stationary lines, an approximately exponential decay law, with characteristic damping coefficients in the 20–200 Hz range, was found to fit the data. These results are not compatible with a Van der Pol oscillator model, and show that some important aspects of the OAE phenomenology can be better explained by the proposed oscillator. Other interesting features of the spontaneous end evoked OAE phenomenology, such as spontaneous OAE suppression by external tones and the following recovery, as well as stimulus/response curves in the linear and nonlinear mode of acquisition, are also well reproduced by the proposed model. © 1999 Acoustical Society of America. [S0001-4966(99)01310-7]

PACS numbers: 43.64.Jb [BLM]

## INTRODUCTION

The extraordinary performances shown by cochlear receptors, in terms of sensitivity around the auditory threshold, dynamic range, and frequency tuning, could not be explained if they were a simple system of mechanical-electrical passive transducers. Already by 1948 the intuitions by Gold (1948) led to the prediction of spontaneous otoacoustic emissions (SOAEs) produced by instabilities in the delicate active feedback mechanism, which allows the cochlea to perform quantum-limited measures in the proximity of the auditory threshold (Bialek and Wit, 1984). The presence of an active feedback mechanism, connected to the outer hair cells' motility, has found extraordinary confirmation by biomedical research (Brownell, 1983, 1990; Brownell *et al.*, 1985; Holton and Hudspeth, 1986; Zenner, 1986; Ashmore, 1987; Evans *et al.*, 1991). On the other hand, after the experimental discovery by Kemp (1978), there have been a large number of experimental observations of acoustic signals generated in the cochlea both in the presence and absence of an external stimulus, evoked otoacoustic emissions (OAEs) and SOAEs, respectively (reviewed in Probst *et al.*, 1991). These observations provide a deeper insight into cochlear micromechanics.

In this work, starting from a number of observations, which are partly found in the literature and partly original, a method is proposed to determine if the complex observed phenomenology can be satisfactorily explained by a model based on given kind of nonlinear oscillator. Along with the

well-known Van der Pol oscillator (Van der Pol, 1927; Hanggi and Riseborough, 1983), a new nonlinear oscillator model, which could better explain some aspects of the OAEs phenomenology, is proposed and analyzed here. In particular, starting from the modeling of SOAEs, some constraints on the functional form of the nonlinear term, which is responsible for the existence of SOAEs, have been found in the present work by studying the relaxation phenomena following the presentation of an external click stimulus. In other words, it is possible to discriminate between different models by following the decay of the tones corresponding to SOAEs after an externally applied wide band stimulus. Neonatal data have been chosen because of the larger prevalence and the typically larger amplitude of SOAEs with respect to adults' data (Burns *et al.*, 1992, 1994), which makes it possible to record high signal-to-noise ratio (SNR) data with commercially available recording systems.

In Sec. I, the basic data acquisition and analysis techniques used in this work are reviewed. In Sec. II, the nonlinear differential equations of the considered physical models are solved by means of the multi-scale analysis technique (Bender and Orszag, 1978), which is discussed in some detail in the Appendix. In this section, SOAE suppression and recovery dynamics are analytically computed by using the proposed model. In Sec. III, some important aspects of the OAE phenomenology, emerging from analysis of our data, are briefly discussed, and the models are compared to the experimental data. The results show that Van der Pol oscillator behavior is not consistent with the observed decay of the OAE spectral lines, while the oscillator model proposed in this work is consistent with the data. A qualitative repro-

<sup>a)</sup>Electronic mail: Renata.Sisto@romal.infn.it

<sup>b)</sup>Electronic mail: moleti@roma2.infn.it



TABLE I. Physical parameters of the 18 examined SOAEs and best-fit values of parameter  $b$  of Eq. (10).

Line #	subject	SOAE frequency (Hz)	SSOAE amplitude (dB SPL)	SOAE amplitude (dB SPL)	Damping coefficient $b$ (Hz)
1	A-R	2368	-9.9	-10	19
2	A-R	2502	-5.5	-2	stationary
3	A-R	2747	-16.0	-14	50
4	A-R	2893	-6.5	-6	50
5	A-R	3052	-8.1	-6	24
6	A-R	4175	-4.3	-5	50
7	A-R	4456	-13.5	-17	102
8	A-L	1440	0.1	2	stationary
9	A-L	3198	-7.5	-8	112
10	A-L	3760	-10.2	-10	170
11	A-L	3992	-0.6	-10	81
12	A-L	4236	13.7	10	stationary
13	B-R	2942	-0.4	5	stationary
14	B-R	3235	-15.2	-7.5	121
15	B-R	3845	2.1	-2	30
16	B-R	4419	3.5	4	stationary
17	B-R	4626	-3.2	-7.5	59
18	B-R	4871	-14.9	-13	98

duction of the experimental stimulus/response curves for evoked OAEs obtained with our model is also shown.

## I. METHOD

In this work neonate click evoked OAEs (CEOAEs), synchronous SOAEs (SSOAEs), and SOAEs, recorded as a part of an extensive neonatal hearing screening project, have been analyzed.

The data have been recorded using an ILO92 system (Otodynamics Ltd.). A total of 18 SOAE spectral lines of 3 ears of 2 female neonates were analyzed, whose parameters are listed in Table I.

The recordings were performed during a quiet sleep period, in a room dedicated to clinical testing. During the test, the neonates were placed in an incubator to diminish the environmental noise. Special probe fittings, which are part of the standard ILO neonatal hearing screening kit, were used.

SOAEs were measured: (a) by recording the signal in the ear canal for  $T=80$  ms with no external stimulus, and then by averaging the spectra of a large number of data streams (these are the properly named SOAEs); (b) by synchronously averaging a large number of data streams ( $T=80$  ms), each recorded after an identical click stimulus, and taking the FFT of the average waveform (these are called synchronous SOAEs, or SSOAEs). The stimulus peak amplitude was in the range 73–79 dB peSPL, and the total acquisition time was of the order of 100 s, for each SSOAE recording. This technique is analogous to that used for recording the CEOAEs in the linear mode of acquisition, the only difference being the elementary data stream length (for CEOAE recordings, usually  $T=10$ –20 ms). It is quite obvious that techniques (a) and (b) are not equivalent, because the SSOAE signal is the response to an external stimulus, even if it is observed for a relatively long time, while the SOAEs are really the spectral lines of spontaneously emitted tones. In this work only the SSOAEs that were clearly visible also in the SOAE spectrum were considered.

CEOAEs are often recorded in the so-called derived nonlinear mode, which differs from the linear mode in that the average is performed over the responses to a train of four stimuli in which three identical stimuli of a given polarity are followed by one of the opposite polarity and triple amplitude. In this acquisition mode, the averaged output signal is not sensitive to the linear component of the response, such as the ringing phenomena affecting the response in the first few ms.

In the present work the 80-ms SSOAE recordings have been FFT analyzed in four separate time intervals of  $T\sim 20$  ms duration by using the standard ILO92 method. In the first interval the first 2.5 ms have been windowed and eliminated to suppress the initial artifact due to linear ringing. The resulting low resolution ( $\Delta f\sim 50$  Hz) spectra provide a simple and effective tool for roughly estimating the decay of the spectral line amplitudes in the 80 ms following the stimulus, as will be shown in Sec. III. The SOAE recordings of the same neonates have also been examined, and these spectra, whose resolution was in our case  $\Delta f\sim 12$  Hz, have been compared to the spectra of the SSOAEs.

## II. PHYSICAL MODELS OF SOAE GENERATION

In this section the analytical properties of the Van der Pol oscillator and of the proposed model are discussed, with respect to both the response to a click stimulus and the suppression and recovery dynamics of the oscillator associated with an SOAE. It should be stressed that the model proposed here does not arise from a schematization of the real ear anatomy. This model is built starting from elementary physics concepts, giving a representation of the cochlear micro-mechanics that could be defined “model-independent.” In this representation, starting from the experimental evidence of high-level functional structures whose behavior is that of active nonlinear oscillators, the resonances are simply analyzed, thus condensing in a few mechanical parameters all the complex physiology of the OAE generation mechanisms.

The modeling scheme of the active element generating the SOAE assumed in the present work is that suggested by Tubis and Talmadge (1998), in which SOAEs are produced by noninteracting effective active oscillators, directly interfaced to the ear canal. So, both the transmission properties of the middle ear and of the cochlea itself and the localization of the SOAE oscillators are not included in the model. In the present work it is assumed that the incremental pressure in the ear canal, in the absence of an external stimulus, is proportional to the amplitude of excitation of the effective oscillator. The proportionality coefficient may depend on frequency, and also on the presence of other SOAEs in the subject, but it is considered here as a constant value, characteristic of any single SOAE. This means that all the relevant nonlinearity of the system is incorporated in the oscillator equation. In this work the excitation amplitudes of the oscillators corresponding to SOAEs are studied in their time evolution after the stimulus, and different spectral line amplitudes are never compared, so the absolute value of this constant is not important. Thus problems such as the spectral periodicity of SOAEs (Zweig and Shera, 1995), which have been studied by means of more complex full cochlear models (Talmadge *et al.*, 1998) by coupling the nonlinear active oscillator equation to propagation equations along the middle ear and the basilar membrane, are not addressed here. This class of models offers a unitary description of many different aspects of the OAE phenomenology, including the frequency spacing characteristics of the different kinds of emissions. These models are considered only with respect to the oscillator equation for the local basilar membrane displacement. In these models the oscillator equation also includes fast and slow feedback time-delayed stiffness terms, as suggested by Zweig (1991). Thus the equation is formally different from the Van der Pol equation, with the fine tuning of the feedback delay times [see the coefficients listed in Table I of Talmadge *et al.* (1998)] implying, however, that the fast delayed stiffness term is, in practice, a negative damping term, analogous to the Van der Pol anti-damping term, while stabilization is given by the Van der Pol nonlinear damping term. As shown below, this feature of the Van der Pol equation produces a compression of the oscillator dynamic range. In this work many OAEs, which are strongly excited by a click stimulus and decay exponentially over a large dynamic range, have been observed. Such OAEs are often referred to as resonant transiently evoked OAEs (e.g., Talmadge *et al.*, 1998), assuming that they represent something qualitatively different from SOAEs. The observations reviewed here, which will be amply discussed in the next section, of exponentially decaying tones with nonzero equilibrium amplitude (SOAEs), is not compatible with oscillator models in which the two behaviors are mutually exclusive. The above observations focused this work on the active oscillator equation, before including it in a more complex full cochlear model.

It has been shown (Bialek and Wit, 1984; Talmadge *et al.*, 1990, 1991) that spectral and statistical properties of SOAEs are compatible with limit-cycle self-sustained active oscillator models, perturbed by stochastic noise. The most studied model of this class is based on the Van der Pol oscillator. The simple Van der Pol oscillator can be modified

by including nonlinear stiffness terms or by assuming that the nonlinear damping term is proportional to the square of the time derivative of the displacement (e.g., Furst *et al.*, 1992), without significant changes in its analytical properties on a slow time scale. A different class of noise driven passive models seems to be ruled out (Talmadge *et al.*, 1991), mainly due to the narrow band filtering properties of SOAEs. A recent discussion concerning the ear canal reflectance structure near SOAE frequencies (Allen *et al.*, 1995; Burns *et al.*, 1998; Tubis and Talmadge, 1998) seems to confirm that an active mechanism producing a power gain is present at low excitation levels at SOAE frequencies. Thus the present work is restricted to the limit-cycle active oscillator class of models.

If the simple approximation in which the active amplification system produces feedback forces, whose main effect is that of reducing the mechanical damping (e.g., Bialek and Wit, 1984) is done, the equation that rules the motion of the effective oscillator associated with a given SOAE, in absence of an external stimulus, is simply:

$$\ddot{X} + (\Gamma_0 - \gamma(X))\dot{X} + \omega_0^2 X = 0, \quad (1)$$

where a stochastic term due to thermal motion has been neglected.

The damping coefficient  $\Gamma(X)$  is defined as the whole coefficient of the term proportional to  $\dot{X}$  in Eq. (1). The oscillator models discussed in this paper differ by the functional form of  $\Gamma(X)$ .

In Eq. (1), the coefficient  $\Gamma_0$  represents the mechanical damping, while  $\gamma(X)$  is a term produced by the active amplification system. Assuming the hypothesis that  $\gamma(X)$  is constituted by a constant term plus a term growing as the amplitude squared, which keeps in mind saturation phenomena of the feedback system (Bialek and Wit, 1984), the Van der Pol oscillator equation is obtained:

$$\ddot{X} + (\eta_1 X^2 - \eta_2)\dot{X} + \omega_0^2 X = 0, \quad (2)$$

where  $\eta_1$  and  $\eta_2$  are positive constants.

The oscillator described by Eq. (2) is an active, limit-cycle oscillator, with equilibrium amplitude:

$$X_{\text{SOAE}} = \sqrt{\frac{2\eta_2}{\eta_1}}. \quad (3)$$

$\eta_1$  is the coefficient of a damping term quadratically growing with the amplitude  $X$ , while  $\eta_2$  represents a constant anti-damping term.

Equation (2) admits an approximate analytical solution, which is perturbatively obtained with the multiple-scale analysis technique (Bender and Orszag, 1978). This method is more extensively discussed in the Appendix. The perturbative solution is:

$$X(t) = 2R(t)\sin(\omega_0 t + \theta_0), \quad (4)$$

with:

$$R^2(t) = \frac{R_\infty^2}{(1 - \alpha e^{-\eta_2 t})}, \quad (5)$$

where

$$R_\infty^2 = \frac{\eta_2}{2\eta_1} \quad \text{and} \quad \alpha = 1 - \frac{R_\infty^2}{R_0^2}. \quad (6)$$

The perturbative solution of Eq. (2) is true in the limit in which the oscillation amplitude does not become too large. In the case of a click stimulus, in the limit  $\alpha \approx 1$  and  $\eta_2 t \rightarrow 0$ , Eq. (5) describes a hyperbolic amplitude decay of the form:  $R^2(t) = 1/(2\eta_1 t)$ . This decaying law has a precise signature: the difference between the rms amplitudes, expressed in dB, at times  $t_1$  and  $t_2$  is simply given by  $10 \log(t_2/t_1)$ , independent of the value of the coefficients of the damping and anti-damping terms. This behavior is not compatible with the exponential decay, with decay coefficients variable in a wide range and decay amplitude range up to 30 dB that were observed in the neonatal data.

Analytical and numerical solutions of Eq. (2), in the case of a click stimulus of amplitude  $h_s$  and duration  $\tau_s$ , have been computed. The agreement between the numerical and analytical solutions is improved as the click is better approximated by a Dirac delta function. This condition is well matched with the stimulus duration  $\tau_s \ll 2\pi/\omega_0$  and its amplitude  $h_s \gg \omega_0^2 X$ . In this case the stimulus gives ‘‘initial conditions’’ on the velocity and on the amplitude:

$$\dot{X}(\tau_s) = h_s \tau_s, \quad X(\tau_s) = \frac{1}{2} h_s \tau_s^2, \quad (7)$$

which, with the time axis shift corresponding to the choice  $t=0$  at the end of the stimulus, imply in Eq. (4):

$$R_0 = \frac{h_s \tau_s}{2\omega_0}, \quad \theta_0 \cong \frac{\omega_0 \tau_s}{2}. \quad (8)$$

Here, and always in the following, all the amplitudes are given in dimensionless units as multiples of a reference amplitude  $X_0$ , which corresponds, for the given oscillator, to an ear canal incremental pressure level  $p_0 = 2 \times 10^{-5}$  Pa, the reference pressure. This choice of units allows the direct comparison of the results of the simulations to the experimental OAE data expressed in dB SPL, and the variable  $X$  will be used indifferently for the displacement of the oscillator and for the incremental pressure in the ear canal. In these units the amplitude  $h_s$  of the stimulus is given by the dimensionless driver pressure (always in units of the reference pressure  $p_0$ ) multiplied by the factor  $\omega_{\text{cav}}^2 = \rho_0 c^2 S_0^2 / m_0 V_e$ . Here  $\rho_0$  is the density of air in the ear canal,  $c$  is the wave velocity of sound,  $V_e$  is the ear canal volume, and  $S_0$  and  $m_0$  are, respectively, the effective area and the effective mass of the ear canal termination, according to the notation given by Tubis and Talmadge (1998).

The analytical solution of Eq. (2) is plotted in Fig. 1(a), for  $R_0 = 12.5$ ,  $\eta_1 = 100$  Hz, and  $\eta_2 = 1$  Hz. For a frequency  $f \approx 3$  kHz, and a click of effective duration  $\tau_s = 0.2$  ms, this value of  $R_0$  corresponds to a stimulus amplitude  $h_s = 2500$  (krad/s)<sup>2</sup>, in dimensionless units. For  $\omega_{\text{cav}} = 500$  rad/s, this amplitude corresponds to a driver pressure (using the notation of Tubis and Talmadge, 1998)  $P_{\text{dr}}(\text{mPa}) = 0.08 h_s = 200$  mPa (80 dB SPL). This value of  $\omega_{\text{cav}}$  is of the order of the highest value quoted by Tubis and Talmadge (1998), which seems to be a reasonable choice for neonatal data. In Fig. 1(b) the moving average of the square of the amplitude, which provides a rough estimate of the evolution

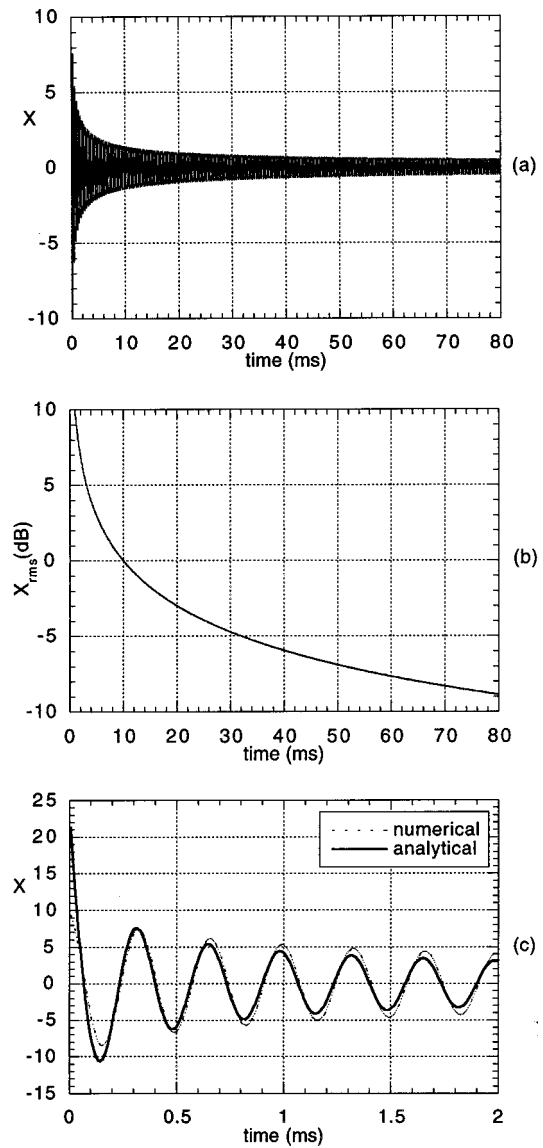


FIG. 1. Analytical solution (a) of the Van der Pol equation, moving average (b) of the amplitude squared, expressed in dB, which shows that the Van der Pol decay does not follow an exponential law, and (c) comparison with the numerical solution.

of the power spectral density, is plotted on a log scale. In Fig. 1(c) the analytical and numerical solutions are compared in the first 2 ms after the stimulus, showing good agreement, with respect to both the amplitude and the phase of the response waveform.

The Van der Pol oscillator has been shown to account for important features of OAEs, such as their already mentioned statistical properties, and the SOAE static and dynamic properties under the effect of external tones. In particular, the Van der Pol oscillator has been used to model SOAE suppression by external tones (Long *et al.*, 1991; Uppenkamp and Kollmeier, 1994; Murphy *et al.*, 1995a, b, 1996), synchronization by external tones (van Dijk and Wit, 1988, 1990a; Long *et al.*, 1991; Uppenkamp and Kollmeier, 1994; Tubis and Talmadge, 1998), and SOAE amplitude and frequency fluctuations (van Dijk and Wit, 1990b). However, as will be discussed in some detail in this paper, the limit-cycle Van der Pol oscillator described by Eq. (2) cannot

model with quantitative accuracy the active feedback mechanism which contributes to the cochlear extraordinary sensitivity in the small signal amplitude limit, and which is responsible for the existence of SOAEs. Studying the relaxation phenomena which follow the presentation of an impulsive stimulus, phenomenological evidence was uncovered that the limit-cycle oscillator associated with a SOAE has an exponentially decreasing amplitude regime, i.e., an amplitude regime in which the damping coefficient is approximately constant [see, e.g., Figs. 5(b), 7, and 8]. In Sec. III below, this experimental evidence will be discussed in detail.

Such an exponential decay phase is not expected for all the limit-cycle oscillator models in which, as in the Van der Pol oscillator, the SOAE equilibrium amplitude is the result of a competition between a constant anti-damping term and a nonlinear damping, growing with the excitation amplitude. This nonlinearity of the damping term is necessary to explain the observed saturation of the response/stimulus curve, but implies that the damping coefficient decreases continuously during the decay of the amplitude subsequent to the excitation by an external stimulus.

A new oscillator model is introduced here, which is capable of explaining both the existence of nonzero equilibrium amplitudes (i.e., SOAEs) and the observed constant  $\Gamma$  exponential decay regime. For simplicity, the saturation mechanism, of which there is clear evidence in the phenomenology of evoked OAEs, is not initially discussed. In other words, an equation is initially presented that describes only the “small amplitude” phenomenology, i.e., the slow relaxation phenomena that follow the stimulus presentation, and explains the presence of SOAEs. Saturation phenomena are described later, by introducing an appropriate saturation term.

The hypothesis is proposed that the coefficient  $\gamma(X)$  of Eq. (1) be constituted by a constant term  $\gamma_0$  and a term that tends to grow significantly in the small amplitude limit:

$$\gamma(X) \equiv \gamma_0 + \frac{c}{\langle X^2 \rangle^m}, \quad (9)$$

with  $\gamma_0 > 0$  and  $m > 0$ .

In Eq. (9)  $\langle X^2 \rangle$  represents the square amplitude averaged over a time scale which is “slow” with respect to the “fast” scale, characteristic of the proper oscillations of the system. The functional form of the nonlinear term of Eq. (9) is arbitrary. It is not determined by the experimental data analyzed here. The choice is based only on the need that the limit-cycle amplitude be due to a nonlinear anti-damping term, which is effective only in the small amplitude limit, so that a linear damping regime holds for a large amplitude range. Any functional form with the same properties would be equally acceptable. With this assumption, Eq. (1) becomes:

$$\ddot{X} + \left[ b - \frac{c}{\langle X^2 \rangle^m} \right] \dot{X} + \omega_0^2 X = 0, \quad (10)$$

where  $b = \Gamma_0 - \gamma_0 > 0$  is the new coefficient of the damping term, reduced by the  $\gamma_0$  term, due to the active system effect, while  $c$  is the coefficient of the anti-damping term. If  $c > 0$ , the system is unstable, i.e., it admits an equilibrium ampli-

tude which is not zero, corresponding to the condition  $\Gamma(X) = 0$ :

$$X_{\text{SOAE}} = \sqrt{2 \left( \frac{c}{b} \right)^{1/m}}. \quad (11)$$

SOAEs can be identified with these equilibrium amplitudes when they are detectable, clearly emerging from the background noise. Thus the definition of SOAE is strictly connected to the capability of improving the detector’s SNR.

Equation (10) admits an approximate analytical solution, which is perturbatively obtained with the multiple-scale analysis technique. In the Appendix the calculations leading to the perturbative solution given by Eq. (4) are reported, with:

$$R(t)^{2m} = R_\infty^{2m} + (R_0^{2m} - R_\infty^{2m}) e^{-bmt} \quad (12)$$

and

$$R_\infty^{2m} = \frac{c}{2^m b}. \quad (13)$$

Equation (12) is valid in the limit  $b \ll \omega_0$ . In the limit in which the amplitude  $R(t)$  is much larger than its equilibrium value:  $R_\infty^{2m} \ll R^{2m}(t) \ll R_0^{2m}$ , Eq. (10) describes oscillations with frequency  $\omega_0$  and an exponentially decreasing amplitude.

Analytical and numerical solutions of Eq. (10), in the case of a click stimulus of amplitude  $h_s$  and duration  $\tau_s$ , have also been computed. As in the Van der Pol case, the agreement between the numerical and analytical solutions is improved as the click is better approximated by a Dirac delta function. From now on, the model is restricted to the case  $m = 1$ , which has been chosen in analogy to the Van der Pol oscillator. The analytical solution of Eq. (10) is plotted in Fig. 2(a), for  $R_0 = 1.5$ ,  $b = 102$  Hz, and  $c = 0.55$  Hz. The exponential behavior is maintained over a wide amplitude range, until the equilibrium level  $X_{\text{SOAE}}^2$  is approached. This is more clearly visible in Fig. 2(b), where a moving average of the amplitude squared is plotted on a log scale, to show the difference with respect to the behavior of the solution of Eq. (2) [see Fig. 1(b) for comparison].

Summarizing, Eq. (10) represents an oscillator whose logarithmic rms amplitude, far from equilibrium, decreases linearly. Such an oscillator has a nonlinear amplification mechanism that becomes relevant only at oscillation amplitudes near the equilibrium, while  $\Gamma$  becomes constant when the oscillation amplitude grows far from the equilibrium. In the system modeled by Eq. (10), the existence of SOAEs is due to a nonlinear anti-damping term which represents an active amplification mechanism of the amplitudes in proximity of equilibrium. On the other hand, Eq. (2) represents an oscillator whose  $\Gamma(X)$ , far from equilibrium, is always strongly nonlinear. The external stimulus response strongly saturates and, after the stimulus, the oscillator relaxes with a hyperbolic envelope waveform.

The oscillator of Eq. (10) is not different from the Van der Pol oscillator with respect to the static properties of the nonzero equilibrium amplitude (SOAE) suppression by an



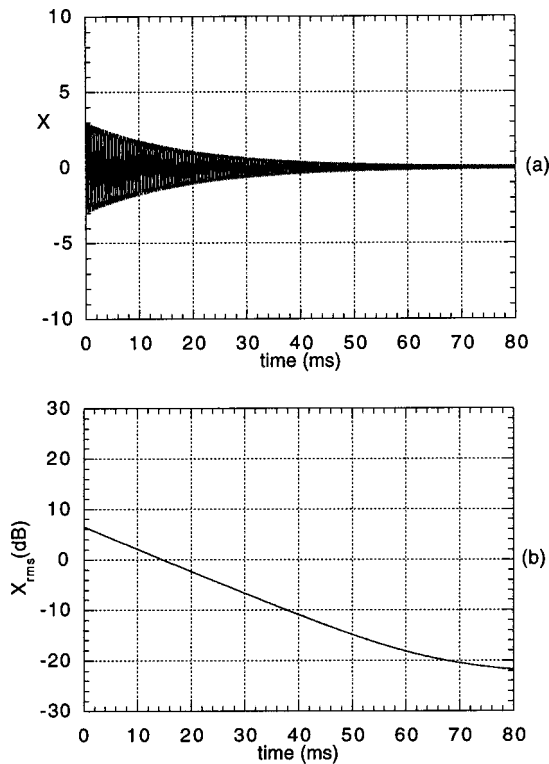


FIG. 2. Analytical solution (a) of Eq. (10) for  $m=1$ , and moving average (b) of the amplitude squared, expressed in dB, which shows that the exponential behavior is maintained over a wide amplitude range, until the equilibrium level  $X_{SOAE}^2$  is approached.

external tone (Murphy *et al.*, 1995a). In presence of an external tone stimulus of frequency  $\omega_1$  and amplitude  $E_1$ , the solution of Eq. (10) becomes:

$$X(t) = 2R(t)\sin(\omega_0 t + \theta_0) + 2\Lambda_1 \sin(\omega_1 t), \quad (14)$$

where

$$2\Lambda_1 = \frac{E_1}{(\omega_0^2 - \omega_1^2)}. \quad (15)$$

Using again the multiple-scale analysis technique, the equation that determines the new equilibrium amplitude, in presence of the suppressor tone is:

$$R_{sup}^2 = k^2 R_{SOAE}^2 \quad \text{with} \quad k^2 = 1 - \frac{\Lambda_1^2}{R_{SOAE}^2}, \quad (16)$$

which has the same functional dependence on external tone frequency and amplitude as in the Van der Pol oscillator, so the predicted suppression tuning curves are the same as those of the Van der Pol oscillator.

A difference arises in the SOAE suppression dynamics. If the same method described in the Appendix to solve Eq. (10) is followed, in the case in which, due to the presence of an external tone, the solution has the form given by Eq. (14). The time evolution of the amplitude of the tone corresponding to the SOAE is given implicitly by:

$$R^{2k^2}(t) - k^2 R_{SOAE}^2 R^{2(k^2-1)}(t) = (1 - k^2) R_{SOAE}^{2k^2} e^{-bk^2 t}, \quad (17)$$

where the initial condition  $R_0 = R_{SOAE}$  has been used.

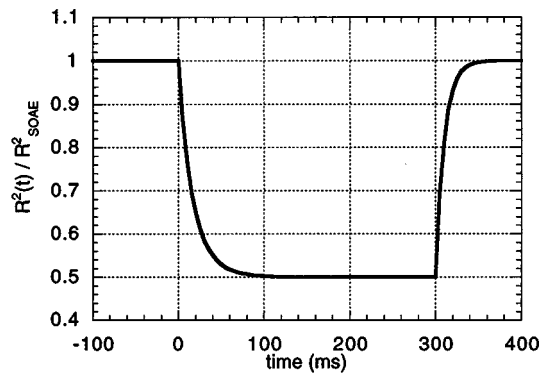


FIG. 3. Suppression and recovery dynamics, as predicted by Eq. (17) and Eq. (12), for  $k^2=0.5$  and  $b=100$  Hz. The dynamics is similar to that predicted by the Van der Pol oscillator model.

It is clear that, in the limit  $t \rightarrow \infty$ , the result of Eq. (16) is obtained. For the recovery from suppression, the time evolution of the amplitude is given by Eq. (12) with the initial condition  $R_0 = R_{sup}$ . In Fig. 3 the suppression and recovery dynamics, as predicted by Eqs. (17) and (12), respectively, is shown for  $k^2=0.5$  and  $b=100$  Hz. These dynamics have a different functional form, but are qualitatively very similar to that predicted by the Van der Pol model (Murphy *et al.*, 1995a), so only very accurate measurements of these phenomena could effectively discriminate between the two models.

As previously pointed out, the limit-cycle oscillator described by Eq. (10) can explain the existence of SOAEs as well as the constant  $\Gamma$  regime observed after presentation of an impulsive stimulus, but cannot explain saturation phenomena. These phenomena are responsible for the strong nonlinearity observed in the stimulus/response curve and for the observed coincidence between CEOAE waveforms recorded with the linear and the nonlinear acquisition mode.

Saturation phenomena could be reasonably modeled by a nonlinear damping term, of the form:  $aX^{2m}\dot{X}$ . In analogy to the Van der Pol equation, as noted above, the case  $m=1$  will be discussed. The limit-cycle oscillator equation that is proposed is of the form:

$$\ddot{X} + \left( aX^2 + b - \frac{c}{\langle X^2 \rangle} \right) \dot{X} + \omega_0^2 X = 0. \quad (18)$$

It should be stressed here that this oscillator is different from the Van der Pol oscillator, not only for the presence of the anti-damping term  $c/\langle X^2 \rangle$ , but mainly because the linear term  $b$  is in this case positive. This linear damping term gives to the solution an exponentially decaying phase, after an initial sharp decay. This oscillator admits an equilibrium amplitude which, in the limit  $4ac/b^2 \ll 1$ , is:

$$\langle X_{SOAE}^2 \rangle \approx \frac{c}{b}. \quad (19)$$

Following the method of multiple-scale analysis (see the Appendix), it is possible to solve perturbatively Eq. (18). The solution is always of the form given by Eq. (4) with:

$$R^2(t) = \frac{R_\infty^2 + R_L^2(1+r_3)r_4 e^{-b't}}{1 - r_4 e^{-b't}}, \quad (20)$$

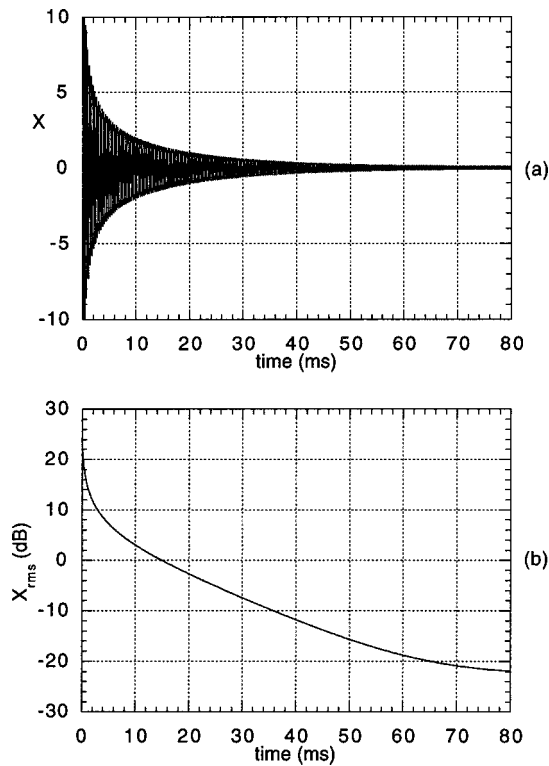


FIG. 4. Analytical solution (a) of Eq. (18), and moving average (b) of the amplitude squared, expressed in dB, showing the three distinct phases: the “saturation,” “linear,” and “equilibrium” phases.

where

$$b' = b + \frac{2ac}{b}; \quad R_\infty^2 \approx \frac{c}{2b}; \quad R_L^2 = \frac{b}{2a}; \quad (21)$$

$$r_1 = \frac{R_\infty^2}{R_0^2}; \quad r_2 = \frac{R_L^2}{R_0^2}; \quad r_3 = \frac{R_\infty^2}{R_L^2}; \quad r_4 = \frac{1-r_1}{1+r_1+r_2}.$$

The behavior of the oscillator described by Eq. (18), after presenting the click, consists of a short, strongly damped phase, which is called here the “saturation” phase, followed by a phase with constant  $\Gamma$ , which is referred to as the “linear” phase, and, finally, the system reaches an equilibrium amplitude that is identified with the SOAE level (“equilibrium” phase). Considering the limit in which:  $R_0^2 > R_L^2 \gg R_\infty^2$ , immediately after the stimulus presentation, the solution of Eq. (20) is similar to that given by Eq. (5), i.e.:

$$R^2(t) \approx \frac{R_L^2}{1 - (1 - r_2)e^{-bt}}. \quad (22)$$

Equation (22), in the limit  $t \rightarrow 0$ , represents a hyperbolically decreasing amplitude.

The “saturation” phase ends when  $R(t)^2 \approx R_L^2$ . As soon as  $R(t)^2 \leq R_L^2$  the “linear” phase begins. During this phase the amplitude decreases with a  $\Gamma$  that is approximately given by the constant  $b$ , until the equilibrium value  $R_\infty$  is approached.

The analytical solution of Eq. (18) is plotted in Fig. 4(a) for the same values of the parameters  $b$ ,  $c$  of Fig. 2, for  $a = 28$  Hz, and for  $R_0 = 12.5$ , corresponding to the same stimulus of Fig. 1. In Fig. 4(b), the moving average of the ampli-

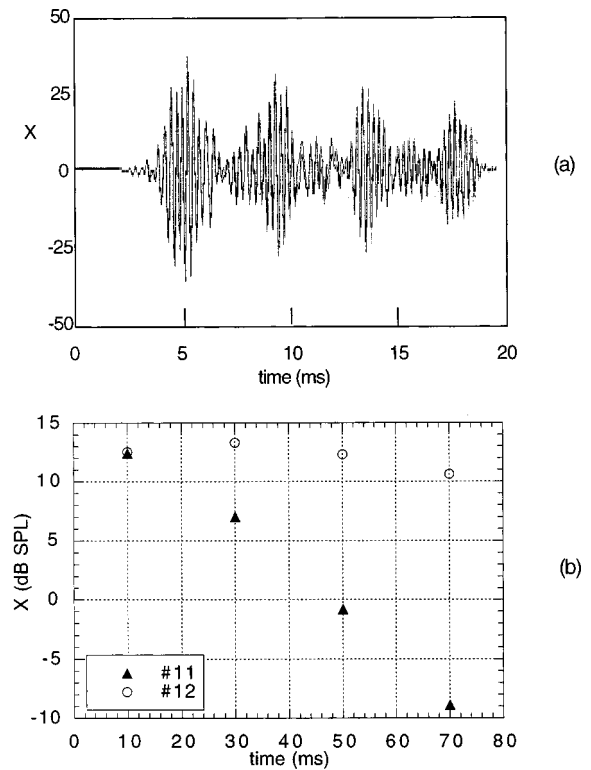


FIG. 5. Evoked waveform (a) in the first 20 ms after the click for subject A-L. The beating of two main tones of frequencies differing by  $\sim 250$  Hz, as shown by the  $\sim 4$ -ms period, is clearly visible in the waveform. The two tones can be identified with the two SOAE and SSOAE lines #11 and #12 of Table I. The response amplitude is expressed in units of  $p_0 = 2 \times 10^{-5}$  Pa, the reference pressure. The time evolution (b) of the two lines is quite different.

tude squared is plotted on a log scale, showing the three distinct phases. As predicted, after the sharp and short “saturation” phase, the exponential behavior is maintained over a wide amplitude range, until the equilibrium level  $X_{\text{SOAE}}^2$  is approached.

The functional form of  $\Gamma$  proposed in this work is qualitatively similar to that proposed by van Hengel *et al.* (1996). They proposed a functional form for the damping coefficient  $\Gamma$ , dependent on  $\dot{X}$  rather than on  $X$ , which, for appropriate values of their four parameters, is approximately constant in a large intermediate velocity interval. In the large velocity limit the damping grows exponentially, giving saturation for high levels of stimulation, while their anti-damping term becomes dominant in the small velocity limit, providing limit-cycle oscillations.

### III. RESULTS AND DISCUSSION

The neonatal data confirm the clear evidence that the CEOAE waveform is dominated by the contribution of tones corresponding to measurable SOAEs (Probst *et al.*, 1986; Kulawiec and Orlando, 1995). This property can be verified by the data of subject A-L shown in Fig. 5(a). That is, the response waveform in the 20 ms after the stimulus clearly shows the beating of two main tones of frequencies differing by  $\Delta f \sim 250$  Hz, as shown by the beating period  $T_B \sim 4$  ms. The two tones can be identified with two strong SOAEs, lines #11 and #12 of Table I. This is a general fact, and it is

an indication that the tones, which have a large equilibrium amplitude, that is observed as an SOAE, also have typically a large response to an external stimulus. If the time evolution of the amplitude of the two lines during the 80 ms of the SSOAE record [Fig. 5(b)] is examined, it is clear that the two behaviors are quite different. That is, the amplitude of line #12 is approximately constant, while that of line #11 decreases exponentially with a characteristic damping coefficient  $b=80$  Hz. The behavior of line #11 is particularly interesting, because, in the literature, such lines are often referred to as resonant transiently evoked OAEs (e.g., Talmadge *et al.*, 1998). Many lines with such a behavior have been observed in this work, with a damping coefficient  $b$  variable in a range 20–200 Hz, and the corresponding SOAE level is also variable over a large range. Some lines of the same subjects, which are not reported here, show a similar exponential decay until the amplitude reaches the noise level. For such lines, it is conceivable (but it cannot be verified) that the corresponding SOAE is not measurable, because its equilibrium level is lower than the noise level. In a cochlear oscillator model, the fact that an SOAE is associated with each of these lines suggests that the different behavior of the lines of Fig. 5(b) is due only to their very different SOAE levels compared to the level of the response to the external stimulus, and that the same oscillator equation can be used for both decaying and stationary lines, with different values of the physical parameters. One more interesting observation is that, in spite of the fact that the SOAE level is very different for the two lines of Fig. 5(b), the response to the external stimulus is of the same order of magnitude. This observation suggests that the physical parameters that determine the amplitude of the response immediately after the stimulus are partly independent from those that determine the equilibrium SOAE amplitude, so a reasonable oscillator model should parameterize independently the large amplitude and the small amplitude behavior of the cochlear oscillators.

It has also been observed (e.g., Burns *et al.*, 1998) that there is a good (but not complete) correspondence between the SOAE and SSOAE spectra, which is indeed the experimental basis for identifying the peaks in the SSOAE spectra with the properly named SOAEs. The good correlation between the levels of SOAEs and SSOAEs has been verified for the 18 lines analyzed in the present work, as shown in Fig. 6, finding a correlation coefficient  $r=0.86$ .

This experimental correlation between SOAEs, SSOAEs, and CEOAEs is expected in an oscillator model framework, and implies that the oscillators with the highest equilibrium amplitudes (i.e., SOAEs) are typically those having the highest quality factor, and so the highest response in the presence of an external stimulus. This is also in agreement with psychoacoustic results (Schloth, 1983), which show that SOAE frequencies coincide with minima in the hearing threshold fine structure, and with the observed correlation (Long and Tubis, 1988a, b; Furst *et al.*, 1992) between modifications of the hearing threshold fine structure and of the SOAE levels induced by ototoxic agents, such as aspirin and noise. These results suggest that the study of a model, which accurately describes both the SOAE level and

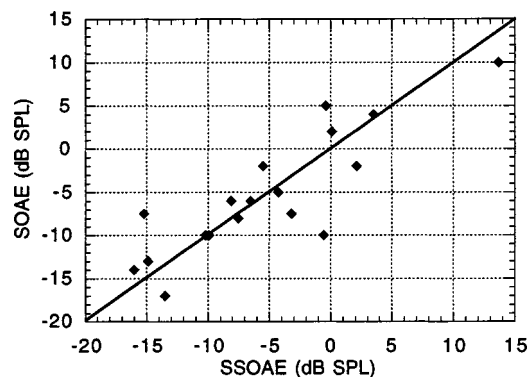


FIG. 6. SOAE vs SSOAE amplitudes for the 18 considered spectral lines. These amplitudes are strongly correlated.

the response to external tones would be useful for effectively parameterizing aspects of cochlear function. Therefore, the present study is important for better understanding the physiology of hearing and monitoring the effects of ototoxic agents in populations exposed to noise. For example, the idea of studying the characteristic decay time of the spectral peaks in the framework of a model in which they are important indicators of the hearing functionality, could prove to be particularly useful in the follow-up of populations exposed to noise. Indeed, such decay parameters could be effectively compared with the same parameters measured on the same subjects, regardless that the recordings are taken at different times, by different operators and even with different instrumentation, while criteria based on the follow-up of absolute amplitudes would be affected, in those cases, by large systematic errors.

The method described in Sec. I has been used to evaluate experimentally the decay of the SSOAE spectral line amplitudes. This is shown in Fig. 7 for the lines #3, 5, 6, and 7 of Table I of subject A-R. The amplitudes of the other lines of the same subject, #1, 2 and 4, which are slowly decaying or approximately constant in the 80-ms time interval, are not plotted here. The lines of Fig. 7 show an approximately exponential decay of the amplitude, with a large variety of characteristic decay times. This observation may provide a powerful clue for the investigation of the OAE phenomenology, as discussed in the previous section. On the other hand,

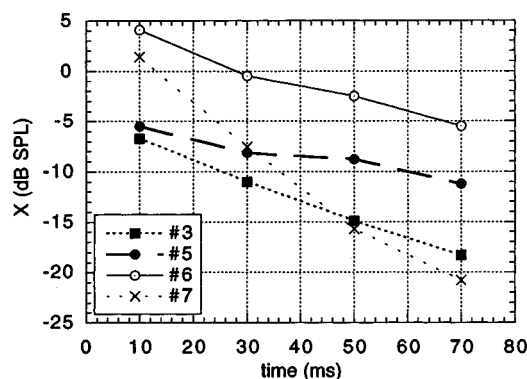


FIG. 7. Time decay of the spectral lines amplitude after the subadministration of a click stimulus for the subject A-R. The decay is approximately exponential. Lines # 1, 2 and 4, of the same subject, whose amplitudes are almost constant, are not plotted here for clarity.

there is clear evidence that the response to a click stimulus is strongly saturated by some nonlinear damping mechanism. Quantitative information about this mechanism is provided, for example, by the well-known correspondence between the CEOAE waveforms recorded with the linear and the derived nonlinear technique and by measurements of the stimulus/response curve (Grandori *et al.*, 1994). Thus any realistic oscillator model should explain also this saturation phenomenology.

In order to test quantitatively the models of Eqs. (2), (10), and (18) by a direct comparison to the data, the 80-ms SSOAE recordings were analyzed by measuring the Fourier amplitudes in the four 20-ms intervals for all the peaks that were clearly visible both in the SSOAE and SOAE spectra. These experimental data have been compared to the output of the models in the following way for each of the 18 lines: (a) the 80-ms response to a click stimulus has been analytically computed for a Van der Pol oscillator and for the oscillator of Eq. (10), in the case  $m = 1$ , as a function of the free parameters  $\eta_2$ ,  $R_0$ , and  $R_\infty$  (for the Van der Pol oscillator) and  $b$ ,  $R_0$ , and  $R_\infty$  [for Eq. (10)]; (b) the computed waveform has been cut into 20-ms strips, a Hanning window has been applied to the strips, and an FFT has been performed on each strip, obtaining four periodograms giving an estimate of the response spectrum in the four time intervals; (c) the spectral peak amplitude has been fitted to the experimental data by varying the oscillator free parameters, using the CERN library MINUIT software with the SIMPLEX and MIGRAD minimization methods. This software performs a least square fit method, with a minimization scheme that uses the simplex method, followed by a Newton minimization procedure, based on the analysis of the derivatives of the fit function.

The fit is performed using Eq. (10) instead of the complete Eq. (18) for simplicity, because the data to be fitted had a poor time resolution (20 ms). The assumption is that the saturation phase that has been defined in Sec. II is short enough to give only an initial condition on the amplitude, which is considered as an unknown parameter to fit. The parameter  $R_0$  that is determined by fitting the data to Eq. (10) is a rough estimate of the value  $R_L$  that is reached very soon by the oscillator of Eq. (18) at the end of the short saturation phase. The result of this procedure is shown in Fig. 8, where the SOAE lines are grouped and sorted by their characteristic decay time. Some lines are approximately stationary in the 80-ms interval. For these lines, no fit is shown, because it would not give any useful information.

A visual inspection of the decaying lines is sufficient to appreciate that the best fitted Van der Pol model is inadequate for describing the experimental decay whenever the amplitude decreases by more than 10 dB in the observed time interval. Such large dynamics require a high SNR to be observed, which is the reason for preferring neonatal data, in which SOAE level and the response to external stimuli are typically larger than in adults.

The model of Eq. (10) fits the data very well, over a wide range of characteristic decay times. The best fit values of parameter  $b$ , which is twice the inverse of the line characteristic decay time, are listed in Table I for the considered spectral lines. The value of the parameter  $a$  of Eq. (18) has

been roughly estimated by making the following consideration: after a short time the excitation drops to a value such that the constant damping term becomes dominant. This happens when  $\langle X^2 \rangle \sim b/a$ , at a time which is much smaller than our time strip length, 20 ms, so it can be assumed that the value of  $R_0$  found by fitting the data with Eq. (10) be such that:

$$a \approx \frac{b}{2R_0^2}. \quad (23)$$

It is important to check if the parameters given by the best fit are consistent with the experimental evidence that the response is strongly saturated over a wide amplitude range. The solution of Eq. (18) has been computed using for each line the best-fit values of the parameters  $b$  and  $c$  and the value of the parameter  $a$  determined using Eq. (23), as a function of the stimulus intensity, finding the rms amplitude in three successive 5-ms windows. In Fig. 9, the computed stimulus/response curves are shown both for the linear and the derived nonlinear acquisition mode for lines #6, 10 and 15. The curves are computed for a click effective duration  $\tau_s = 0.2$  ms. The stimulus amplitude  $h_s$  is given in units of  $(\text{krad/s})^2$ , such that for  $\omega_{\text{cav}} = 500$  rad/s,  $P_{\text{dr}}(\text{mPa}) = 0.08h_s$ . These lines are the largest spectral components in the first 20 ms of the studied subjects, and so would give the dominant contribution to their experimental stimulus/response curves. The ability to do such stimulus/response measurements on the present subjects was not possible, so the results were compared qualitatively to analogous experimental measurements available in the literature. This comparison showed that the current analytical results are very similar to measurements (Grandori *et al.*, 1994) of saturation phenomena, showing a linear growth of the response for low stimuli and a saturation above a characteristic stimulus level, which is quite variable between the different subjects. This variability is easily explained in our model by the variability of the parameters  $a$ ,  $b$  of the SOAEs giving the main contribution to the evoked signal. The present model is capable of reproducing both the linear and the saturated parts of the experimental stimulus/response curves. It should be stressed again that all the parameters used in Eq. (18) have not been chosen to fit this saturation behavior, but have been independently found, through a fit of the slow decaying phase. This fit had been performed by putting all the details of the fast saturation behavior into the fitted initial value  $R_0$ .

#### IV. CONCLUSIONS

The technique of following the time evolution of the spectral lines associated to SOAEs after presenting a click stimulus, which was investigated in this work, is a very effective tool for determining the correct functional form of the nonlinear oscillator equation describing the cochlear resonances. The model of Eq. (18) that has been proposed here is capable of describing, with very few parameters, both the saturation phase and the slow decaying phase that are experimentally observed in the time evolution of OAEs after an



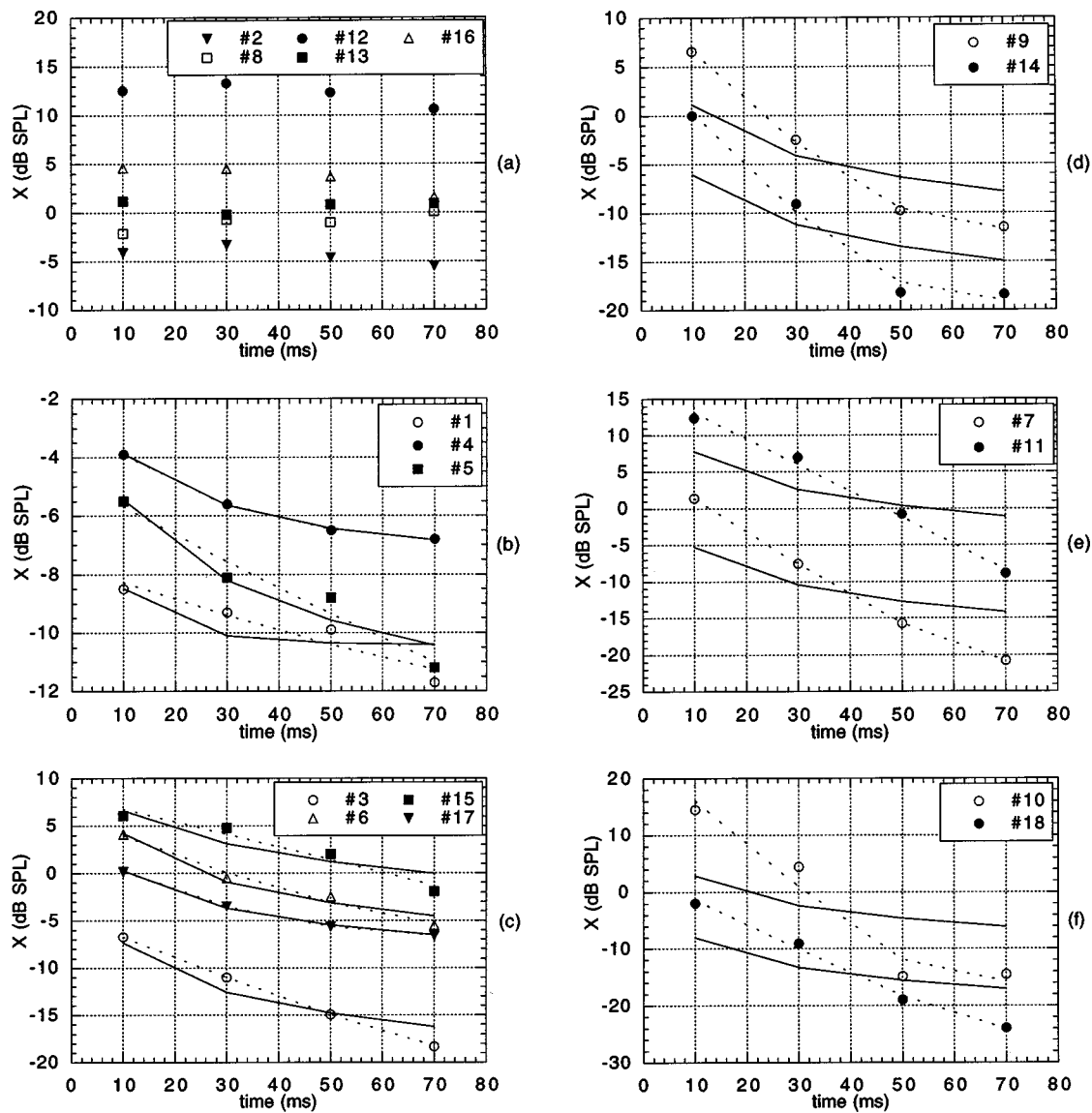


FIG. 8. Comparison between the experimental data and the best fitted predictions of the Van der Pol model (solid line) and of the model of Eq. (10) (dotted line). SOAE lines are grouped and sorted by decreasing decay time to make it evident that the best fitted Van der Pol model is not adequate for describing the experimental decay whenever the amplitude decreases by more than 10 dB in the 80-ms time interval.

impulsive stimulus. The model predicts the observed exponential decay of the lines of frequency corresponding to measurable SOAEs after excitation by a click stimulus, which is not compatible with a Van der Pol oscillator model. The model has also been able to reproduce the CEOAE experimental stimulus/response curves described in the literature, both for the linear and nonlinear standard modes of acquisition.

Of course, the model is also capable of explaining the existence of SOAEs, which are associated to the oscillators with a high value of  $c/b$ . The major advantage of this model with respect to other oscillator models is that the nonzero equilibrium amplitude, associated with SOAEs, arises from the competition between a constant damping term and an anti-damping term which becomes important only near the SOAE level, and which is almost negligible at higher excitation levels. This is in agreement with the high sensitivity of the ear to very low intensity sound, and implies a decay law

that is approximately exponential over a wide amplitude range, in agreement with the observations presented here. The static and dynamical properties of SOAE suppression by external tones and subsequent recovery to the SOAE equilibrium level are very similar to those predicted by the Van der Pol model.

The model is rather simplified with respect to full cochlear models, whose important feature is the capability of predicting the OAE spectral properties. Nevertheless, it has the interesting aspect of providing an unitary explanation of the phenomenology of the single oscillator, with respect to the response to high-level stimuli (saturation), the relaxation dynamics, and the presence of asymptotic equilibrium states (i.e., SOAEs). The results show that a correct modeling of the elementary oscillator should be considered as a very important requisite for the making of more complex cochlear models.

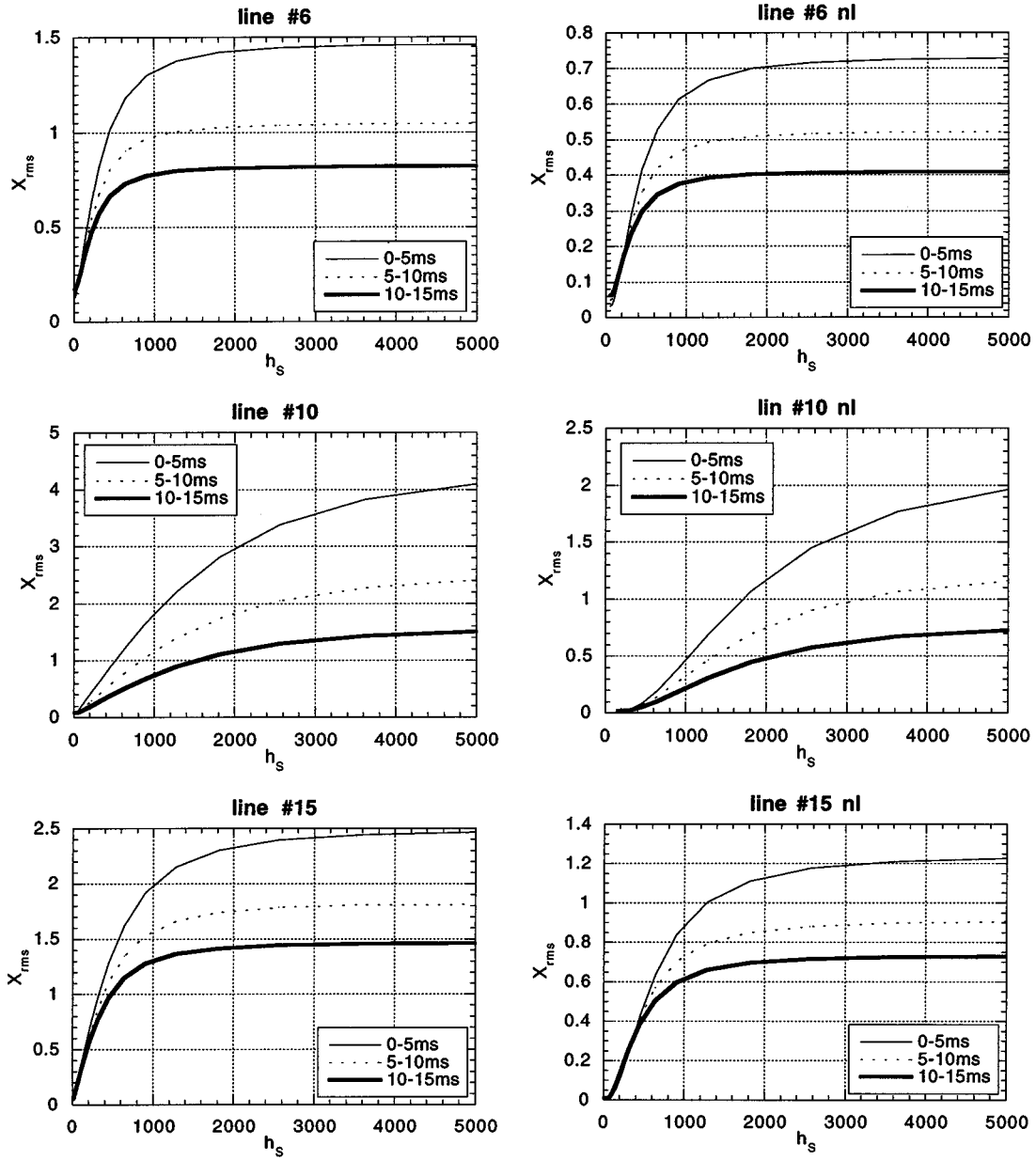


FIG. 9. Computed stimulus/response curves for the linear and the derived nonlinear acquisition mode for lines # 6, 10, and 15, reproducing both the linear and the saturated parts of the experimental stimulus/response curves found in the literature. The three curves are the rms response amplitude in the time intervals 0–5, 5–10, and 10–15 ms after the click stimulus. For  $\omega_{\text{cav}}^2 = 0.5$  krad/s, the driver pressure is given by  $4h_s$ , in units of  $p_0 = 2 \times 10^{-5}$  Pa.

## ACKNOWLEDGMENTS

We wish to thank the Azienda Ospedaliera “San Giovanni in Laterano,” where the presented data were recorded, and where an extensive neonatal hearing screening project is in progress. Special thanks to Dr. D. Tufarelli and A. Ingresso for providing the experimental data and for helpful discussions.

## APPENDIX: MULTI-SCALE ANALYTICAL SOLUTION OF NONLINEAR DIFFERENTIAL EQUATIONS

In the equation:

$$\ddot{X} + b \left[ 1 - \frac{c}{b \langle X^2 \rangle} \right] \dot{X} + \omega_0^2 X = 0 \quad (\text{A1})$$

with the initial conditions given by a click stimulus of amplitude  $h_s$  and duration  $\tau_s$ :

$$X(0,0) = \frac{1}{2} h_s \tau_s^2, \quad \dot{X}(0,0) = h_s \tau_s, \quad (\text{A2})$$

it is possible to clearly identify two time scales: a “fast” scale associated to the proper oscillation frequency  $\omega_0$  and a “slow” scale that describes the time evolution of the envelope waveform. A dimensionless time variable  $\omega_0 t$  is introduced for simplicity as  $t$ . Equation (A1) becomes:

$$\ddot{X} + b \left[ 1 - \frac{c}{b \langle X^2 \rangle} \right] \dot{X} + X = 0, \quad (\text{A3})$$

where  $b$  and  $c$  are now dimensionless parameters, that are not renamed, for simplicity. To solve perturbatively Eq. (A1), the multiple-scale analysis method (Bender and Orszag,

1978) is used. If  $\epsilon$  is a perturbative expansion parameter ( $\epsilon \ll 1$ ), in the multiple-scale approach, functions depend on two independent variables:  $t$  which represents variations on a ‘‘fast’’ time scale and  $\tau \approx \epsilon t$ , which defines a ‘‘long’’ time scale because  $\tau$  is not negligible when  $t$  is of the order  $1/\epsilon$ . In this framework the total derivative with respect to the time becomes:

$$\frac{d}{dt} = \partial_t + \partial_\tau, \quad (\text{A4})$$

$$\frac{d^2}{dt^2} = \partial_t^2 + 2\partial_t\partial_\tau + 0(\epsilon^2), \quad (\text{A5})$$

where  $\partial_\tau \approx \epsilon$ . The solution  $X$  is a function of  $t$  and  $\tau$  and can be perturbatively expanded in power series of the expansion parameter  $\epsilon$ :

$$X(t, \tau) = X^{(0)}(t, \tau) + X^{(1)}(t, \tau) + 0(\epsilon^2). \quad (\text{A6})$$

In Eq. (A3),  $b$  and  $c$  are both of the order  $\epsilon$  ( $b, c \ll 1$ ), while  $\langle \dots \rangle$  represents the average with respect to the fast time scale. Equation (A3) admits a limit-cycle amplitude:

$$X_{\text{SOAE}} = \sqrt{\frac{2c}{b}}. \quad (\text{A7})$$

Applying the multiple-scale technique to the solution of Eq. (A3) results in the following system of equations:

$$\partial_t^2 X^{(0)} + X^{(0)} = 0, \quad (\text{A8})$$

$$\partial_t^2 X^{(1)} + X^{(1)} = -2\partial_t\partial_\tau X^{(0)} - b \left( 1 - \frac{c}{b\langle X^{(0)2} \rangle} \right) \partial_t X^{(0)}. \quad (\text{A9})$$

The leading order solution  $X^{(0)}$  is of the form:

$$X^{(0)} = A(\tau)e^{jt} + A^*(\tau)e^{-jt}, \quad (\text{A10})$$

where  $A(\tau)$  is a complex amplitude which can be put into the form:

$$A(\tau) = R(\tau)e^{j\theta(\tau)}, \quad (\text{A11})$$

with

$$\langle X^{(0)2} \rangle = 2R^2(\tau). \quad (\text{A12})$$

The explicit form of the functions  $R(\tau)$  and  $\theta(\tau)$  can be obtained by imposing that the solution  $X(t, \tau)$  keeps finite at every order of the perturbative expansion, i.e., that there is no secularity into the equations for  $X$  in any perturbative order. At first order in the  $\epsilon$  expansion, for example, the term on the right hand side of Eq. (A9) is considered as an ‘‘external’’ term in the  $X(t)$  evolution equation and indicated as  $F^{(1)}$ . In order to prevent that secular solutions exist, into the  $F^{(1)}$  term the coefficients of the oscillating terms  $e^{jt}$  and  $e^{-jt}$ , that are solutions of the homogeneous equation associated to Eq. (A9), must be zero. Imposing annulment condition of such coefficients, the complex equation is:

$$2 \frac{dA}{d\tau} + b \left( 1 - \frac{c}{2bR^2(\tau)} \right) A = 0, \quad (\text{A13})$$

which, using Eq. (A11), and separating the real and the imaginary part, translates into the two following independent equations:

$$2 \frac{dR}{d\tau} + b \left( 1 - \frac{c}{2bR^2} \right) R = 0, \quad 2 \frac{d\theta}{d\tau} R = 0. \quad (\text{A14})$$

The phase solution is:

$$\theta(\tau) = \theta_0. \quad (\text{A15})$$

The solution for the amplitude is:

$$R^2(t) = R_\infty^2 + (R_0^2 - R_\infty^2)e^{-bt}, \quad (\text{A16})$$

where the dimensional time scale has been reintroduced and:

$$R_\infty^2 = \frac{c}{2b}. \quad (\text{A17})$$

Finally, imposing the initial conditions Eq. (A2), the solution is:

$$X^{(0)} = 2R(t)\sin(\omega_0 t + \theta_0), \quad (\text{A18})$$

where  $R(t)$  is given by Eq. (A16) with  $R_0 = (h_s \tau_s)/2\omega_0$ , and  $\theta_0 = (\omega_0 \tau_s)/2$ . With the multiple-scale technique it is also possible to solve equations of the form:

$$\ddot{X} + \left( b - \frac{c}{\langle X^2 \rangle^m} \right) \dot{X} + \omega_0^2 X = 0, \quad (\text{A19})$$

with  $m$  positive integer or fractional coefficient. With the initial conditions Eq. (A2), the solution is given by Eq. (A18) with:

$$R^{2m}(t) = R_\infty^{2m} + (R_0^{2m} - R_\infty^{2m})e^{-bmt}, \quad (\text{A20})$$

where:

$$R_\infty^{2m} = \frac{c}{2^m b}. \quad (\text{A21})$$

With the multiple-scale analysis the Van der Pol equation can also be perturbatively solved:

$$\ddot{X} + (\eta_1 X^2 - \eta_2) \dot{X} + \omega_0^2 X = 0, \quad (\text{A22})$$

for which, with the initial condition Eq. (A2), the solution is obtained that is given by Eq. (A18) with:

$$R^2(t) = \frac{R_\infty^2}{(1 - \alpha e^{-\eta_2 t})}, \quad (\text{A23})$$

where:

$$R_\infty^2 = \frac{\eta_2}{2\eta_1} \quad \text{and} \quad \alpha = 1 - \frac{R_\infty^2}{R_0^2}. \quad (\text{A24})$$

Details about the solution of the oscillator of Eq. (18) are:

$$\ddot{X} + \left( aX^2 + b - \frac{c}{\langle X^2 \rangle} \right) \dot{X} + \omega_0^2 X = 0, \quad (\text{A25})$$

with  $a, b, c > 0$ . Following the steps previously illustrated to solve Eq. (A1), starting from the solution given by Eqs. (A10), (A11), the parameters evolving over the slow time scale are determined solving the equation system:

$$2 \frac{dR}{d\tau} + \left( 2aR^2 + b - \frac{c}{2R^2} \right) R = 0, \quad 2 \frac{d\theta}{d\tau} R = 0. \quad (\text{A26})$$

The solution for the phase parameter is given by Eq. (A15). Introducing the new variable:  $y = 2\sqrt{a}R^2 + b/2\sqrt{a}$ , the solution for the amplitude is obtained solving the integral equation:

$$\int_{y(0)}^{y(t)} \frac{1}{2\sqrt{a}} \frac{dy}{y^2 - \left( \frac{b^2}{4a} + c \right)} = - \int_0^t \frac{d\tau}{2}. \quad (\text{A27})$$

Defining  $\beta^2 = (b^2/4a) + c$ , as  $y \geq \beta$ , the solution is:

$$\frac{1}{2\beta\sqrt{a}} \ln \left( \frac{y - \beta}{y + \beta} \right) \Bigg|_{y(0)}^{y(t)} = -t. \quad (\text{A28})$$

In the limit  $(4ac/b^2) \ll 1$ ,  $\beta \approx (b/2\sqrt{a})(1 + (2ac/b^2))$ , defining:  $R_\infty^2 \approx c/2b$ ;  $R_L^2 = b/2a$ , the following is obtained:

$$\frac{R^2(t) - R_\infty^2}{R^2(t) + R_L^2 + R_\infty^2} = \frac{R_0^2 - R_\infty^2}{R_0^2 + R_L^2 + R_\infty^2} e^{-b't}, \quad (\text{A29})$$

where:  $b' = b + (2ac/b)$ . Finally, with the definitions

$$r_1 = \frac{R_\infty^2}{R_0^2}; \quad r_2 = \frac{R_L^2}{R_0^2}; \quad r_3 = \frac{R_\infty^2}{R_L^2}; \quad r_4 = \frac{1 - r_1}{1 + r_1 + r_2};$$

the following equation holds:

$$R^2(t) = \frac{R_\infty^2 + R_L^2(1 + r_3)r_4 e^{-b't}}{1 - r_4 e^{-b't}}. \quad (\text{A30})$$

Generalizing Eq. (A25), the equation becomes:

$$\ddot{X} + \left( aX^{2m} + b - \frac{c}{\langle X^2 \rangle^m} \right) \dot{X} + \omega_0^2 X = 0. \quad (\text{A31})$$

Following the procedure previously illustrated, introducing the new variable:  $u = 2^m \sqrt{a}R^{2m} + (b/2\sqrt{a})$ , the solution for the amplitude is obtained by solving the integral equation:

$$\int_{u(0)}^{u(t)} \frac{du}{u^2 - \beta^2} = -m\sqrt{a} \int_0^t d\tau, \quad (\text{A32})$$

where  $\beta^2 = (b^2/4a) + c$ . As  $u \geq \beta$ , the solution is:

$$\frac{1}{2\beta\sqrt{a}} \ln \left( \frac{u - \beta}{u + \beta} \right) \Bigg|_{u(0)}^{u(t)} = -mt. \quad (\text{A33})$$

In the limit  $(4ac/b^2) \ll 1$ ,  $\beta \approx (b/2\sqrt{a})(1 + (2ac/b^2))$ , defining:  $R_\infty^{2m} \approx c/2^m b$ ;  $R_L^{2m} = b/2^m a$ , we obtain:

$$\frac{R^{2m}(t) - R_\infty^{2m}}{R^{2m}(t) + R_L^{2m} + R_\infty^{2m}} = \frac{R_0^{2m} - R_\infty^{2m}}{R_0^{2m} + R_L^{2m} + R_\infty^{2m}} e^{-mb't}, \quad (\text{A34})$$

where:  $b' \approx b + (2ac/b)$ . Finally, with the definitions:

$$\tilde{r}_1 = \frac{R_\infty^{2m}}{R_0^{2m}}; \quad \tilde{r}_2 = \frac{R_L^{2m}}{R_0^{2m}}; \quad \tilde{r}_3 = \frac{R_\infty^{2m}}{R_L^{2m}}; \quad \tilde{r}_4 = \frac{1 - \tilde{r}_1}{1 + \tilde{r}_1 + \tilde{r}_2};$$

the following equation holds:

$$R^{2m}(t) = \frac{R_\infty^{2m} + R_L^{2m}(1 + \tilde{r}_3)\tilde{r}_4 e^{-mb't}}{1 - \tilde{r}_4 e^{-mb't}}. \quad (\text{A35})$$

- Allen, J. B., Shaw, G., and Kimberley, B. P. (1995). "Characterization of the nonlinear ear canal impedance at low sound levels," in Abstracts of the 18th Midwinter Research Meeting of the ARO, edited by R. Popelka (Association for Research in Otolaryngology, Des Moines), pp. 190.
- Ashmore, J. F. (1987). "A fast motile response in guinea pig OHC: The cellular basis of the cochlear amplifier," *J. Physiol. (London)* **388**, 323–347.
- Bender, C. M., and Orszag, S. A. (1978). *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York), pp. 544–566.
- Bialek, W. S., and Wit, H. P. (1984). "Quantum limits to oscillator stability: Theory and experiments on otoacoustic emissions from the human ear," *Phys. Lett. A* **104**, 173–178.
- Brownell, W. E. (1983). "Observations of a motile response in isolated outer hair cells," in *Mechanism of Hearing*, edited by W. R. Webster and L. M. Aitkin (Monash University Press, Clayton, Australia), pp. 5–10.
- Brownell, W. E. (1990). "Outer hair cell electromotility and otoacoustic emissions," *Ear Hear.* **11**, 82–92.
- Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (1985). "Evoked mechanical responses of isolated cochlear outer hair cells," *Science* **227**, 194–196.
- Burns, E. M., Arehart, K. H., and Campbell, S. L. (1992). "Prevalence of spontaneous otoacoustic emissions in neonates," *J. Acoust. Soc. Am.* **91**, 1571–1575.
- Burns, E. M., Campbell, S. L., and Arehart, K. H. (1994). "Longitudinal measurements of spontaneous otoacoustic emissions in infants," *J. Acoust. Soc. Am.* **95**, 385–394.
- Burns, E. M., Keefe, D. H., and Ling, R. (1998). "Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies," *J. Acoust. Soc. Am.* **103**, 462–474.
- Evans, B. N., Hallworth, R., and Dallos, P. (1991). "Outer hair cell electromotility: The sensitivity and vulnerability of the dc component," *Hearing Res.* **52**, 288–304.
- Furst, M., Reshef, I., and Attias, J. (1992). "Manifestations of intense noise stimulation on spontaneous otoacoustic emission and threshold microstructure: Experiment and model," *J. Acoust. Soc. Am.* **91**, 1003–1014.
- Gold, T. (1948). "Hearing II. The physical basis of the action of the cochlea," *Proc. R. Soc. London* **135**, 492–498.
- Grandori, F., Ravazzani, P., Tognola, G., and Hatzopoulos, S. (1994). "'Derived Nonlinear' versus 'Linear' click evoked otoacoustic emissions," in *Advances in Otoacoustic Emissions—Vol. I—Fundamentals and Clinical Applications*, edited by F. Grandori (Commission of the European Communities, Directorate-General for Science, Research and Development, Biomedical and Health Research Program), pp. 48–64.
- Hanggi, P., and Riseborough, P. (1983). "Dynamics of nonlinear dissipative oscillators," *Am. J. Phys.* **51**, 347–352.
- Holton, T., and Hudspeth, A. J. (1986). "The transduction channels of the hair cells of the bullfrog characterized by noise analysis," *J. Physiol. (London)* **375**, 195–227.
- Kemp, D. T. (1978). "Stimulated otoacoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.* **64**, 1386–1391.
- Kulawiec, J. T., and Orlando, M. S. (1995). "The contribution of spontaneous otoacoustic emissions to the click evoked otoacoustic emissions," *Ear Hear.* **16**, 515–520.
- Long, G. R., and Tubis, A. (1988a). "Investigation into the nature of the association between threshold microstructure and otoacoustic emissions," *Hearing Res.* **36**, 125–138.
- Long, G. R., and Tubis, A. (1988b). "Modification of spontaneous and evoked otoacoustic emissions and associated psychoacoustic microstructure by aspirin consumption," *J. Acoust. Soc. Am.* **84**, 1343–1353.
- Long, G. R., Tubis, A., and Jones, K. L. (1991). "Modeling synchronization and suppression of spontaneous otoacoustic emissions using Van der Pol oscillators: Effects of aspirin administration," *J. Acoust. Soc. Am.* **89**, 1201–1212.
- Murphy, W. J., Talmadge, C. L., Tubis, A., and Long, G. R. (1995a). "Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. I. Response to pulsed single-tone suppressors," *J. Acoust. Soc. Am.* **97**, 3702–3710.
- Murphy, W. J., Talmadge, C. L., Tubis, A., and Long, G. R. (1995b). "Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. II. Suppression of interacting emissions," *J. Acoust. Soc. Am.* **97**, 3711–3720.
- Murphy, W. J., Talmadge, C. L., Tubis, A., Long, G. R., and Krieg, E. F. (1996). "Relaxation dynamics of spontaneous otoacoustic emissions per-



- turbed by external tones. III. Response to a single tone at multiple suppression levels," *J. Acoust. Soc. Am.* **100**, 3979–3982.
- Probst, R., Lonsbury-Martin, B. L., and Martin, G. L. (1991). "A review of otoacoustic emissions," *J. Acoust. Soc. Am.* **89**, 2027–2067.
- Probst, R., Coats, A. C., Martin, G. K., and Lonsbury-Martin, B. L. (1986). "Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears," *Hearing Res.* **21**, 261–275.
- Schloth, E. (1983). "Relation between spectral composition of spontaneous otoacoustic emissions and fine structure of threshold in quiet," *Acustica* **53**, 250–256.
- Talmadge, C. L., Long, G. R., Murphy, W. J. and Tubis, A. (1990). "Quantitative evaluation of limit-cycle oscillator models of spontaneous otoacoustic emissions," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele, *Lecture Notes in Biomathematics* **87** (Springer, Berlin), pp. 235–242.
- Talmadge, C. L., Tubis, A., Wit, H. P., and Long, G. R. (1991). "Are spontaneous otoacoustic emissions generated by self-sustained cochlear oscillators?," *J. Acoust. Soc. Am.* **89**, 2391–2399.
- Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (1998). "Modeling otoacoustic emission and hearing threshold fine structures," *J. Acoust. Soc. Am.* **104**, 1517–1543.
- Tubis, A., and Talmadge, C. L. (1998). "Ear canal reflectance in the presence of spontaneous otoacoustic emissions. I. Limit-cycle oscillator model," *J. Acoust. Soc. Am.* **103**, 454–461.
- Uppenkamp, S., and Kollmeier, B. (1994). "Narrow band stimulation and synchronization of otoacoustic emissions," *Hearing Res.* **78**, 210–220.
- Van der Pol, B. (1927). "Forced oscillations in a circuit with nonlinear resistance," *Philos. Mag.* **3**, 65–80.
- van Dijk, P., and Wit, H. P. (1988). "Phase-lock of spontaneous otoacoustic emissions to a cubic difference tone," in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic, London), pp. 101–105.
- van Dijk, P., and Wit, H. P. (1990a). "Synchronization of spontaneous otoacoustic emissions to a  $2f_1 - f_2$  distortion product," *J. Acoust. Soc. Am.* **88**, 850–856.
- van Dijk, P., and Wit, H. P. (1990b). "Amplitude and frequency fluctuations of spontaneous otoacoustic emissions," *J. Acoust. Soc. Am.* **88**, 1779–1793.
- van Hengel, P. W. J., Duifhuis, H., and van den Raadt, M. P. M. G. (1996). "Spatial periodicity in the cochlea: The result of interaction of spontaneous emissions?," *J. Acoust. Soc. Am.* **99**, 3566–3571.
- Zenner, H. P. (1986). "Motile responses in outer hair cells," *Hearing Res.* **22**, 83–90.
- Zweig, G. (1991). "Finding the impedance of the organ of Corti," *J. Acoust. Soc. Am.* **89**, 1229–1254.
- Zweig, G., and Shera, C. A. (1995). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," *J. Acoust. Soc. Am.* **98**, 2018–2047.

# Detection and intensity discrimination of Gaussian-shaped tone pulses as a function of duration

Thomas Baer, Brian C. J. Moore, and Brian R. Glasberg

*Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England*

(Received 26 January 1999; accepted for publication 14 June 1999)

Van Schijndel *et al.* [J. Acoust. Soc. Am. **105**, 3425–3435 (1999)] proposed that the auditory system partitions the spectro-temporal domain into frequency-time (f-t) windows and that the characteristics of these windows could be explored by measuring intensity discrimination for Gaussian-shaped tone pulses presented just above their detection threshold in noise. They reasoned that for a long-duration tone pulse, the auditory representation would be maximally compact in the frequency domain, but would spread across several f-t windows in the time domain. For a very-short-duration tone pulse, the auditory representation would be maximally compact in the time domain, but would spread across several f-t windows in the frequency domain. There should be some intermediate duration at which the auditory representation is compact in both the time and frequency domains and for which intensity-discrimination performance should worsen, due to the limited opportunity for multiple looks. Their data for signal frequencies of 1 and 4 kHz were consistent with this expectation; intensity discrimination was poorest at a duration of about 3–5 ms at 1 kHz and 1 ms at 4 kHz (durations are specified between 6.8-dB-down points on the envelope). This experiment attempted to replicate those results and to extend them to a wider range of frequencies and levels. Intensity discrimination of Gaussian-shaped tone pulses was measured at three levels: 10 dB above absolute threshold or above masked threshold in a pink noise with a spectrum level of either 15 or 40 dB at 1 kHz. The signal frequency was 0.25 kHz (durations from 2 to 320 ms), 1 kHz (durations from 0.5 to 80 ms), or 4 kHz (durations from 0.1 to 20 ms). Three normally hearing subjects were tested. At 1 and 4 kHz, performance was poorest overall for the 15-dB pink noise level, and thresholds showed a peak at intermediate durations (about 3–5 ms at 1 kHz and 1 ms at 4 kHz). Such peaks were still apparent, but smaller in the no-noise condition and were almost absent at the higher noise level. For the 0.25-kHz signal frequency, peaks were not observed consistently at any level, although two subjects showed small peaks for durations around 10 ms. An explanation is offered for the results in terms of the level and frequency dependence of basilar-membrane input–output functions. © 1999 Acoustical Society of America.

[S0001-4966(99)00110-1]

PACS numbers: 43.66.Dc, 43.66.Fe, 43.66.Mk [SPB]

## INTRODUCTION

Van Schijndel *et al.* (1999) have proposed that the auditory system uses a “multiple looks” mechanism (Viemeister and Wakefield, 1991) to combine information about stimuli across both time and frequency. Effectively, this means that it partitions the spectro-temporal domain into frequency-time (f-t) windows and combines information across these windows. For long-duration narrow-band signals (e.g., sustained tones), good discrimination performance is achieved by combining noisy estimates across multiple looks in time, while for short-duration wideband signals (clicklike stimuli), performance is improved by combining noisy estimates across frequency. In principle, it might be possible to find a signal whose representation is as compact as possible in the auditory system, i.e., whose frequency-time spread is minimal. Such a signal would reduce the opportunity for multiple looks, which might result in poor discrimination performance.

Van Schijndel *et al.* (1999) applied these ideas to an intensity discrimination task, using Gaussian-windowed tones with varying durations,  $D$ , of the Gaussian window

(which are specified here between 6.8-dB-down points on the envelope). These stimuli have Gaussian-shaped envelopes in the time domain and Gaussian-shaped spectra in the frequency domain. The spectral bandwidth is inversely proportional to the duration. Van Schijndel *et al.* presented their stimuli in a background of pink noise. Six values of  $D$  were used, corresponding to less than 1 to more than 25 periods of the carrier tone. In their first experiment, the carrier frequency was 1 kHz, the overall noise level was 35 dBA, and the reference sound (the pedestal) was presented at a level of 0, 10, 20, and 30 dB, relative to its masked threshold. They found that the Weber fraction,  $\Delta I/I$ , hardly varied with  $D$  for stimuli presented 0 and 30 dB above masked threshold. However, at 10 dB, and, to some extent, 20 dB, there was a marked increase in the Weber fraction for  $D$ 's of 3–7 ms, corresponding to three to seven periods of the 1-kHz carrier tone. In a second experiment, the carrier frequency was 4 kHz, and the signal was presented at 10 dB above masked threshold in a 35-dBA pink noise, while in a third experiment the carrier frequency was 1 kHz, and the signal was presented at 10 dB above masked threshold in a 55-dBA

pink noise. The results for these experiments were similar to those for experiment 1;  $\Delta I/I$  showed a maximum at a value of  $D$  corresponding to three to five periods of the carrier.

Van Schijndel *et al.* noted that the peak in  $\Delta I/I$  occurred roughly when the spectral bandwidth of the Gaussian-shaped tone pulse was comparable to the critical bandwidth or equivalent rectangular bandwidth (ERB) of the auditory filter. This is consistent with the idea that there is a specific value of  $D, D_{\min}$ , for each carrier frequency at which the frequency-time spread of the signal in the auditory system is minimal. To explain why  $\Delta I/I$  did not vary with  $D$  when the signal was at 0 or 30 dB above masked threshold, they distinguished three cases.

- (1) When the signal is very close to the masked threshold, performance is limited by the inherent fluctuations in the masking noise. Assume that  $\Delta I/I$  is measured for the signal duration  $D_{\min}$  giving minimal spectro-temporal spread in the auditory system. If the signal duration is now either increased or decreased, there will be a spread over several windows. However, if  $\Delta I/I$  is fixed (and  $I$  decreases with increasing duration), the signal-to-noise ratio decreases in each window, so combining information across windows provides roughly the same total information, and performance remains roughly the same.
- (2) When the signal is 10–20 dB above masked threshold, performance is controlled by “coding noise” in the auditory system. They argued that the level of this noise in each frequency-time window should be proportional to the excitation in that window, i.e., Weber’s law holds within each window. Thus, the “effective” signal-to-noise ratio is approximately the same at each point in the excitation pattern that is above threshold. If the duration is either increased or decreased from the value  $D_{\min}$ , there will be a spread over several frequency-time windows. In this case, the opportunity for multiple looks leads to improved performance, as the “extra” looks are obtained without a worsening of the signal-to-noise ratio in each window.
- (3) When the signal is 30 dB or more above masked threshold, multiple frequency-time windows are always excited, regardless of the signal duration, as there is a considerable spread of detectable excitation across the tonotopic array. Thus the dependence of  $\Delta I/I$  on  $D$  disappears again.

At 4 kHz, the value of  $D_{\min}$  found by van Schijndel *et al.* (1999) was only about 1 ms. This is much shorter than the equivalent rectangular duration of the sliding temporal integrator (window) derived from experiments on forward and backward masking (Festen *et al.*, 1977; Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994), gap detection (Moore *et al.*, 1993a), and increment and decrement detection (Buunen and van Valkenburg, 1979; Plack and Moore, 1991; Moore *et al.*, 1993b; Peters *et al.*, 1995; Moore *et al.*, 1996), which is typically in the range 5–15 ms. The temporal window derived from these experiments on temporal resolution has often been assumed to reflect processes occurring after the auditory nerve (Moore, 1997). It has also been assumed that discrimination of auditory stimuli

is based on the *output* of the window; the rapid time variations that might be present at the input to the window are smoothed and are no longer available for auditory processing. In contrast, the duration of the hypothetical f-t window derived by van Schijndel *et al.* (1999) was argued to depend largely on peripheral filtering processes. It seems contradictory to assume that, at high frequencies, a very short f-t window might be used in the intensity discrimination task of van Schijndel *et al.* (1999), but that such short windows could not be used in tasks such as increment and decrement detection. Therefore, it seems prudent to examine the paradigm of van Schijndel *et al.* (1999) in more detail, and to determine whether there might be another plausible explanation for the pattern of the results. Such an explanation is offered in Sec. III of this paper.

Our experiment attempted to replicate the results of van Schijndel *et al.* (1999), using a similar condition, and to extend them to a wider range of frequencies and levels. The 35-dBA noise used in two of their three experiments would have produced only a small amount of masking, so the levels of the stimuli would have been relatively low. For example, at 1000 Hz, the spectrum level of their noise was only 2 dB. If the maximum in the value of  $\Delta I/I$  at a specific duration is caused by the signal having a maximally compact representation in the auditory system at that duration, then a similar maximum should be observed over a wide range of overall noise levels, provided that the signal is always presented at 10 dB above its masked threshold in the noise. Also, a maximum should be observed when the signal is presented at 10 dB above its absolute threshold. Van Schijndel *et al.* (1999) suggested that the spectral width of the elementary f-t window was “roughly a critical band.” To test this idea more rigorously, we included a much lower carrier frequency (250 Hz) than used by van Schijndel *et al.* At 250 Hz, the critical bandwidth, as specified by the ERB of the auditory filter, is only about 52 Hz (Glasberg and Moore, 1990), so the maximum in the value of  $\Delta I/I$  should occur at a larger value of  $D$  than for the higher carrier frequencies.

## I. METHOD

The experiment was performed in two stages. In the first stage, absolute or masked thresholds were measured for each duration used at each carrier frequency. In the second stage, intensity discrimination thresholds (DLs) were measured using pedestal levels 10 dB above the detection thresholds measured in stage one.

### A. Stimuli

The stimuli to be detected or discriminated were Gaussian-windowed sinusoids, whose waveforms are specified by

$$s(t) = A \cos(2\pi F_c t - \pi/4) \exp(-\pi t^2/D^2), \quad (1)$$

where  $A \cos(-\pi/4)$  is the instantaneous amplitude at the temporal center of the stimulus,  $F_c$  is the carrier frequency,  $D$  is a measure of the duration of the Gaussian window, and  $t$  is time, specified relative to the temporal center of the stimulus. Stimulus level is specified in terms of  $A$ , the peak amplitude of the carrier. If the Gaussian function was an

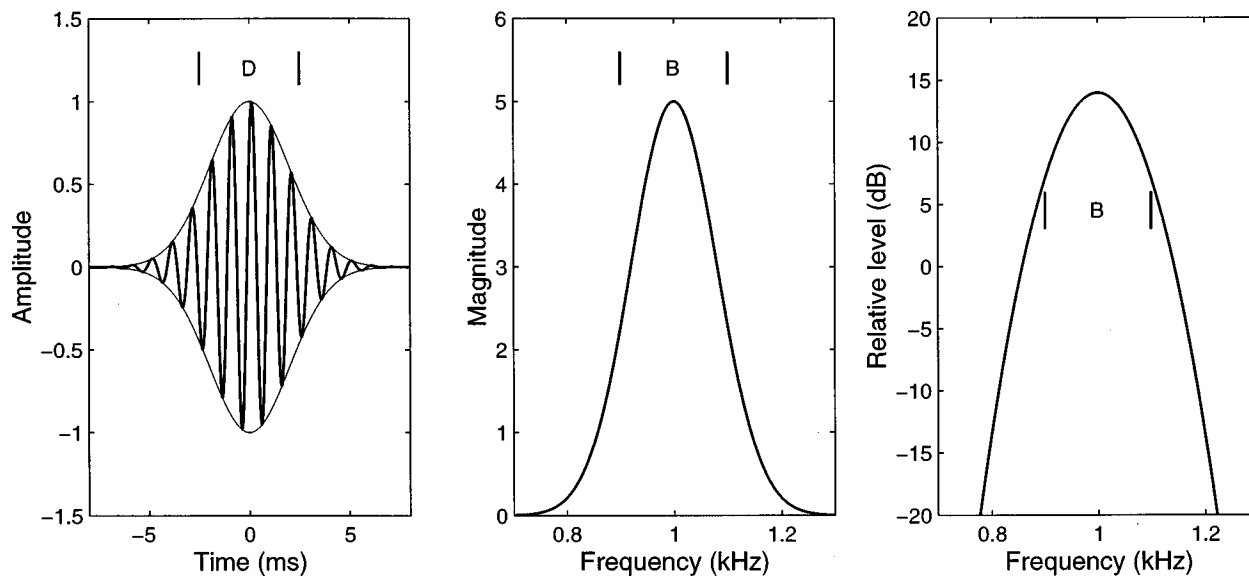


FIG. 1. Illustration of the waveform of a Gaussian-shaped tone pulse (left panel) and its spectrum. The waveform is scaled so that the peak value of the Gaussian envelope is 1 [in Eq. (1), this corresponds to setting  $A=1$ ]. The envelope is shown as the thin line. The spectrum is plotted both with a linear ordinate (amplitudelike units—middle panel) and a dB ordinate (right panel).

intensity-weighting function, then  $D$  would be its equivalent rectangular duration. However, in Eq. (1), the Gaussian function is actually an amplitude-weighting function. Here  $D$  is the duration at which the envelope is 6.82 dB down from its peak value. The bandwidth,  $B$ , is given by the equation

$$B = 1/D. \quad (2)$$

Again, if the Gaussian function was an intensity-weighting function, then  $B$  would be its equivalent rectangular bandwidth. Here  $B$  is the bandwidth at which the spectrum is 6.82 dB down from its peak value. The time waveform and spectrum of a typical stimulus are illustrated in Fig. 1. The spectrum is shown both with a linear amplitude ordinate and a dB ordinate.

In theory, the Gaussian function extends indefinitely in time. In practice, the value of the function becomes negligible outside a certain range, and we truncated the function outside the range  $-1.6D \leq t \leq 1.6D$ . The window was about 70 dB below its peak value at  $t = \pm 1.6D$ . The phase shift  $\pi/4$  was included so that, for a given level, the energy in the pulse varied in direct proportion with  $D$ . In pilot experiments, the phase shift was 0, so that the pulse was symmetric in time but the energy departed somewhat from this relationship at very short durations. This did not appear to change the results substantially.

Stimuli were calculated with 16-bit resolution at a 50-kHz sampling rate on a PC with a Tucker-Davis Technologies (TDT) AP2 array processor and were generated with TDT System II equipment, which also collected subjects' responses and provided feedback. The test stimuli were presented either in quiet or in a continuous pink-noise background. The noise was obtained from an Ivie IE-20B noise generator, low-pass filtered at 15 kHz, and routed through a manual attenuator before being mixed with the test stimuli. Subjects listened monaurally through HD414 headphones in a double-walled sound-isolated booth.

There were three values of  $F_c$ , 250, 1000, and 4000 Hz. There were eight values of  $D$  for each  $F_c$ , chosen so that the number of carrier periods within the time  $D$  varied from 0.4 or 0.5 to 80. The values of  $D$  were 2, 5, 10, 20, 40, 80, 160, and 320 ms for  $F_c = 250$  Hz; 0.5, 1, 2, 5, 10, 20, 40, and 80 ms for  $F_c = 1000$  Hz; and 0.1, 0.25, 0.5, 1, 2, 5, 10, and 20 ms for  $F_c = 4000$  Hz. Detection thresholds and intensity discrimination were measured in quiet, and in pink noise with a spectrum level at 1000 Hz of 15 or 40 dB (*re* 20  $\mu$ Pa). The lower noise level produced masked thresholds that were typically about 20 dB above the absolute threshold. The higher noise level produced thresholds that were about 25 dB higher than for the lower noise level, but usually did not exceed 90 dB SPL, even for the shortest duration used. For the intensity-discrimination conditions, the pedestal level was set 10 dB above the absolute or masked threshold, so the pedestal level did not usually exceed 100 dB SPL.

## B. Subjects

Three normal-hearing subjects were used. Two were paid for their services. The other was author BG. All had absolute thresholds better than 15 dB HL at all audiometric frequencies. Subject BG was highly experienced in psychoacoustic tasks, while the other two subjects were not. All subjects received 2–4 h of practice on a selection of conditions, after which their performance appeared to be stable. No systematic improvements in performance were observed during the course of the experiment.

## C. Procedure

A trial contained two observation intervals indicated by lights on the response box and separated by a 500-ms interval. For each of the 72 conditions (in quiet and two noise levels, at three frequencies, with eight durations per frequency), the detection threshold was measured using a three-



down one-up two-interval forced-choice (2IFC) adaptive procedure tracking 79% correct. The signal was presented at random in either the first or the second interval, and the subject was required to indicate the interval containing the signal. Feedback was provided by lights on the response box. Twelve turnpoints were obtained. The step size was 5 dB until four turnpoints had occurred and 2 dB thereafter. The threshold was taken as the mean level across the last eight turnpoints. If the standard deviation across turnpoints was greater than 3 dB, the run was discarded. At least three valid runs were obtained. If the standard deviation across runs was greater than 2 dB, more runs were obtained until the standard deviation fell below 2 dB.

The following procedure was used to select appropriate levels at which to present stimuli in stage two. For each of the nine noise-frequency conditions for each of the three subjects, a linear regression was performed on the thresholds plotted as a function of  $\log D$ . The measured thresholds were generally within 1–2 dB of the regression lines, and the difference never exceeded 3 dB. The pedestal levels for intensity discrimination were set 10 dB above the values determined from the fitted regression lines.

Intensity-discrimination thresholds were measured using a similar three-down one-up 2IFC adaptive procedure tracking 79% correct. The pedestal was presented at intensity  $I$  in both observation intervals, and an increment in intensity,  $\Delta I$ , was added in one of the intervals, selected at random. The subject was required to indicate the interval containing the more intense stimulus. The initial step size, in terms of  $\Delta I/I$ , was 3 dB, and the final step size was 1 dB. If the standard deviation across turnpoints was greater than 3 dB, the run was discarded. At least six valid runs were obtained for each condition.

## II. RESULTS

### A. Detection thresholds

Figure 2 shows threshold levels as a function of  $D$  (log scale). Patterns for individual subjects were very similar, and only the mean across subjects is shown. Standard deviations (s.d.'s) of the thresholds across subjects were nearly always smaller than 2 dB. The largest s.d. was 2.8 dB. In general, the curves can be well fitted by straight lines. Linear regression lines were fitted to the individual data. Table I gives the means and standard deviations of the slopes across subjects. The mean slopes were essentially the same as for lines fitted to the data averaged across subjects. Slopes are roughly  $-3$  dB/doubling, which is consistent with constant energy at threshold across pulse duration,  $D$ , as has been found in earlier studies (Hughes, 1946; Garner and Miller, 1947; Plomp and Bouman, 1959; Florentine *et al.*, 1988; van Schijndel *et al.*, 1999). However, there is a small but systematic increase of slope with masker level.

### B. Intensity DLs

Figures 3 and 4 show the intensity DLs, expressed as  $10 \log(\Delta I/I)$ , plotted as a function of  $D$  (log scale), for each carrier-masker condition. The standard deviation of the DLs across runs for a specific subject and condition ranged from

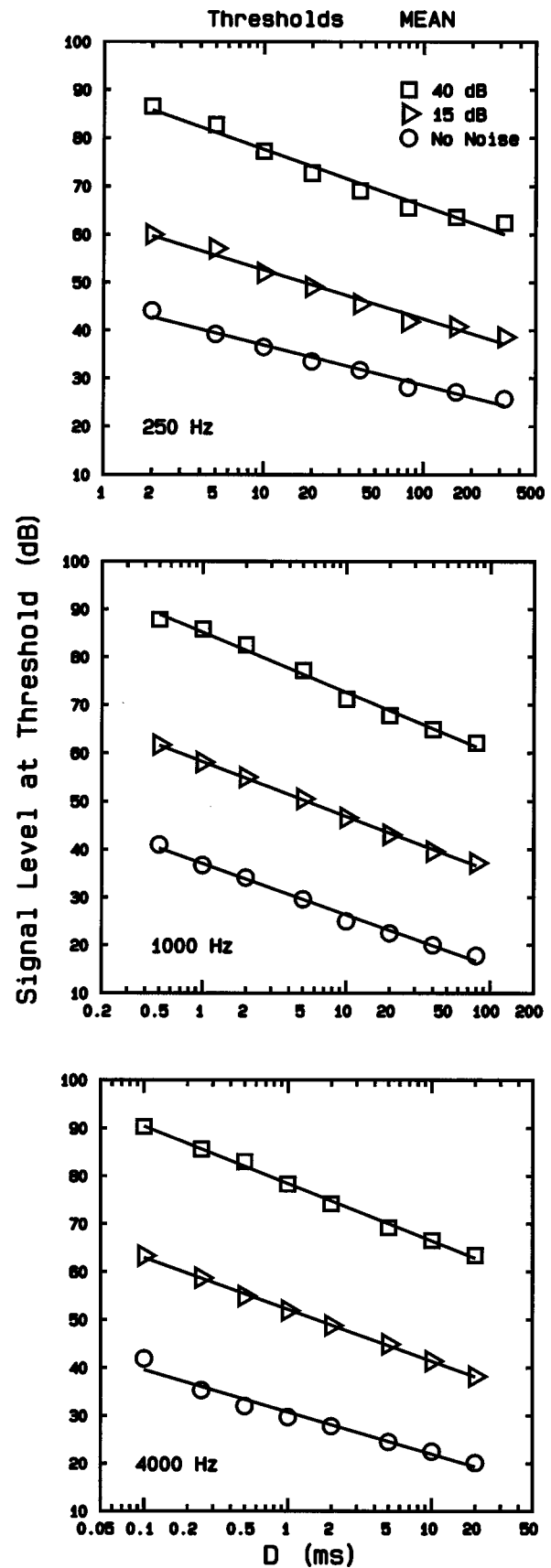


FIG. 2. Data averaged across subjects showing absolute thresholds (circles) and masked thresholds in pink noise with a spectrum level at 1000 Hz of 15 dB (triangles) or 40 dB (squares), plotted as a function of equivalent rectangular duration,  $D$ . Error bars are not shown as almost all of the standard deviations across subjects were smaller than 2 dB, so the error bars would be smaller than the symbols used in the figure.

TABLE I. Means (standard deviations) of the threshold-curve regression slopes fitted to the data of the individual subjects, in dB per doubling of  $D$ .

	No noise	15 dB	40 dB
250 Hz	-2.5 (0.2)	-3.1 (0.2)	-3.5 (0.1)
1000 Hz	-3.2 (0.3)	-3.4 (0.1)	-3.8 (0.1)
4000 Hz	-2.7 (0.3)	-3.2 (0.1)	-3.6 (0.0)

0.4 to 4.1 dB, but was typically in the range 1–3 dB. The standard error was typically 0.4–1.2 dB, and was never greater than 1.7 dB. Separate plots are shown for each subject (Fig. 3) and for the mean across subjects (Fig. 4). For the condition closest to that of van Schijndel *et al.* (1999) (1000-Hz carrier, 15-dB masker—middle column of each plot, triangles), all three subjects show a marked peak. However, the location of the peak varies across subjects; it is at 2 ms for BG, 1 ms for JH, and 5 ms for RK. The peak for RK is markedly higher than that for the other two subjects, so the mean across subjects shows a peak at 5 ms, corresponding to five periods of the carrier. Although the peak in the mean data occurs at about the same duration as found by van Schijndel *et al.*, our subjects perform less consistently than theirs.

For the 1000-Hz carrier with no masking noise (middle column, circles), each subject shows a peak, but again the results vary across subjects. In the mean data, there is a peak at 2–5 ms, but it is less marked than the peak for the 15-dB masker. For the 1000-Hz carrier, 40-dB masker (middle column, squares), small maxima occur in the functions, but they

are much less distinct than for the 15-dB noise or no-noise conditions. Intensity DLs are generally smallest with this masker.

For the 4-kHz carrier (right column), the mean patterns were similar to those for the 1-kHz carrier, and there was a similar range of variation across subjects. However, in this case the peak was typically close to 1 ms, corresponding to four periods of the carrier. This corresponds well to the results of van Schijndel *et al.* (1999). The peak was clearest for the mid-level (15-dB carrier) condition. For the no-noise condition, the peak was smaller and less consistent across subjects. For the 40-dB masker, performance was generally better than for the other two conditions. There were small peaks in the functions relating  $\Delta I/I$  to  $D$ , but they were much less regular and less distinct than for the other two conditions.

For the 250-Hz carrier (left column), there was generally no peak in the middle of the range of  $D$  values used. If van Schijndel *et al.* (1999) were correct in assuming that the spectral width of the elementary f-t window was roughly one critical bandwidth, then the width at 250 Hz should have been about a factor of 2.5 smaller than at 1000 Hz (Glasberg and Moore, 1990). Assuming that the duration of the f-t window is proportional to the reciprocal of its spectral width, one would expect to see peaks in the  $\Delta I/I$  values for values of  $D$  around 12.5 ms. Such peaks were not found, although the mean data show a very small peak at  $D = 10$  ms for the no-noise condition. For the 15- and 40-dB noise conditions, the mean data show small peaks at 5–10 ms, corresponding

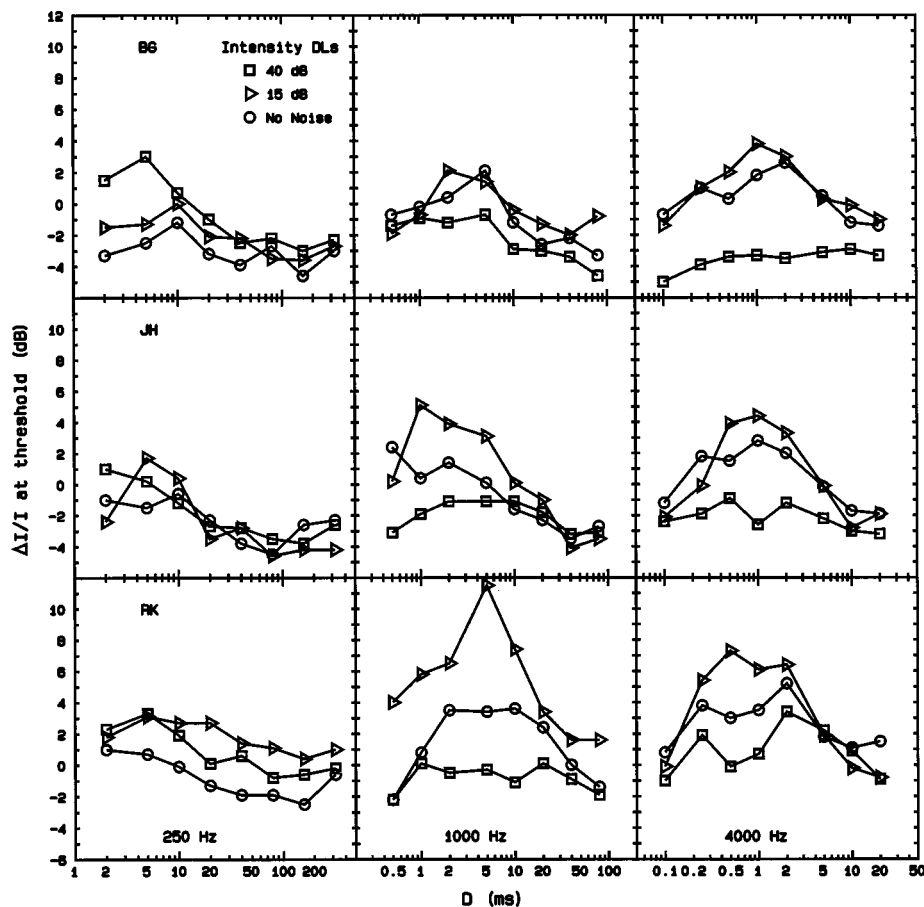


FIG. 3. Intensity discrimination thresholds, expressed as  $\Delta I/I$  in dB, plotted as a function of equivalent rectangular duration,  $D$ . Each row shows results for one subject, and each column shows results for one frequency. The pedestal levels were 10 dB above the absolute threshold (circles) or above the masked threshold in pink noise with a spectrum level at 1000 Hz of 15 dB (triangles) or 40 dB (squares).

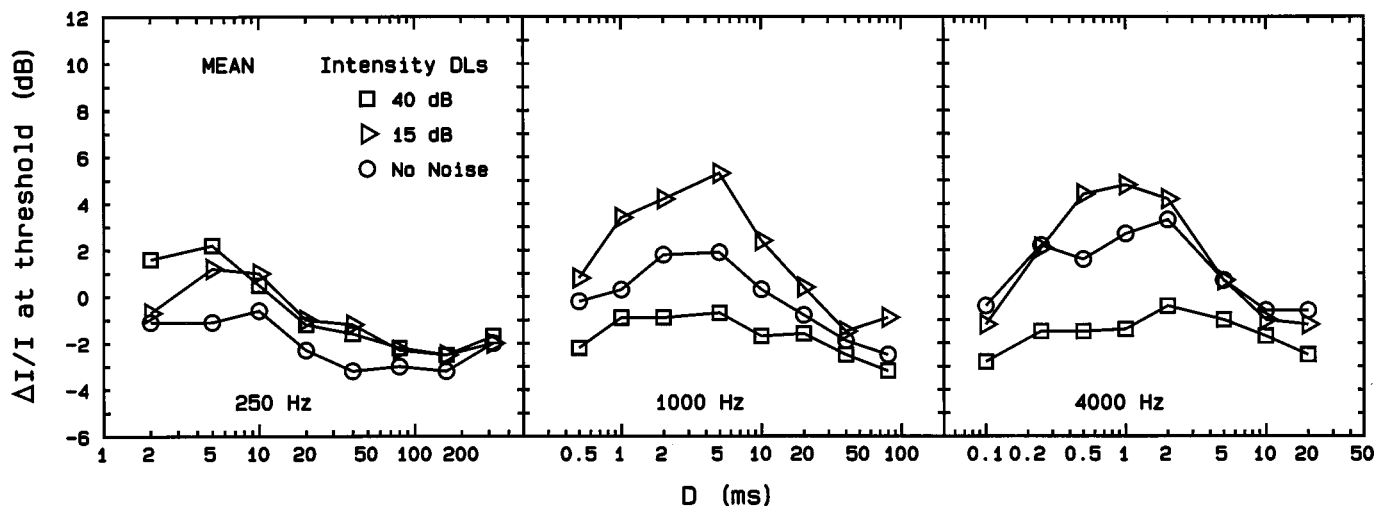


FIG. 4. As in Fig. 3, but showing the values of  $\Delta I/I$  averaged across subjects.

to two to four periods of the carrier. Unlike the data for the 1000- and 4000-Hz carriers, performance was not best for the 40-dB noise condition. In the mean data, performance was best for the no-noise condition.

### III. DISCUSSION

#### A. Intensity-discrimination data

We will consider separately two aspects of the results: first, the variations in overall performance with level and second, the variations in performance with duration. For the 1000- and 4000-Hz carriers, intensity-discrimination performance was generally worst for the 15-dB masker and best for the 40-dB masker, with the no-noise condition giving intermediate thresholds. It is not clear how this could be accounted for within the framework of the theory proposed by van Schijndel *et al.* (1999). A possible explanation for the variation in overall performance with level comes from consideration of input-output functions on the basilar membrane, which, for frequencies close to the characteristic frequency (CF), are highly compressive for moderate sound levels, but more linear for very low and perhaps for very high sound levels (Sellick *et al.*, 1982; Robles *et al.*, 1986; Ruggero *et al.*, 1997).<sup>1</sup> The effect at very high levels is somewhat controversial, and recent data suggest that, for cochleae in good condition (showing high sensitivity), the response may remain highly compressive at high levels. However, the physiological data are usually presented for a single place on the basilar membrane, with the frequency of the stimulating tone fixed at the CF measured at low levels. The position of the peak of the response on the basilar membrane, which is probably the position relevant for psychophysical detection of a tone in noise, may shift more basally with increasing level. As pointed out by Oxenham and Plack (1997), if the basilar-membrane input-output function is measured for a tone frequency which tracks the position of the peak, then the function becomes somewhat more linear at high levels even in the most sensitive cochleae.

In the 15-dB noise condition of our experiment, most of the pedestal levels fell in the range 45–68 dB SPL, with the higher levels occurring for the shorter durations. However,

for the shortest durations, the energy in the pulses was spread over a range considerably greater than 1 ERB. The level per ERB was typically in the range 45–59 dB, where the basilar membrane input-output function is highly compressive. In contrast, for the 40-dB noise condition, the pedestal levels fell in the range 70–100 dB SPL, and the level per ERB fell in the range 70–91 dB, where the input-output function is probably less compressive. In the no-noise condition, the pedestal levels per ERB were within the range 27–43 dB SPL, at 1000 and 4000 Hz, a range where processing is nearly linear (although some compression would have been applied to the pedestal plus increment).

The results for the 1000- and 4000-Hz carriers are consistent with the idea that overall performance was determined by the amount of basilar-membrane compression; more compression led to poorer intensity discrimination. Thus, performance was worse for the 15-dB noise level than for the other conditions. Compression may make performance worse for two reasons. First, it reduces the effective differences in level between the stimuli to be discriminated. Second, it can have different effects on the signal to be discriminated and on the background noise, a point which is covered in more detail below.

We turn now to consideration of the changes in performance with duration. The model of van Schijndel *et al.* (1999) predicts that the duration at which intensity discrimination is poorest should be inversely related to the auditory filter bandwidth. This bandwidth increases markedly with increasing level for medium and high frequencies, but the change with level is reduced or absent for lower frequencies (Moore and Glasberg, 1987; Rosen and Stock, 1992; Baker *et al.*, 1998; Hicks and Bacon, 1999; Glasberg *et al.*, 1999). Thus, their model predicts that, at medium and high frequencies, the peaks in the functions relating intensity discrimination to duration should move to shorter durations with increasing level, while at low frequencies the peaks should occur at a duration that is independent of level. The data do not provide any support for these predictions.

The improvement in performance as the duration was decreased below about 5 ms at 1000 Hz and 1 ms at 4000 Hz

could have been caused by basilar-membrane responses being more linear for very brief stimuli. Recio *et al.* (1998) reported that, for click stimuli applied to the region of the chinchilla cochlea with characteristic frequency (CF) in the range 8–10 kHz, responses were nearly linear (input–output functions with slopes greater than 0.8 dB/dB) within 250  $\mu$ s of the onset of the response (corresponding to about two to three periods of the CF), but became highly compressive within 500 to 700  $\mu$ s (corresponding to five to seven periods). If the response at lower frequencies is approximately linear over the first two to three periods of presentation of a brief tone pulse, this could explain the improvement in intensity discrimination performance for very short durations. The improvement would be expected to be greatest in cases where strong compression led to the poorest performance for intermediate durations, which is consistent with our data.

Several mechanisms may contribute to the improvement in intensity discrimination performance with increasing duration beyond 5 ms at 1000 Hz and 1 ms at 4000 Hz. One has to do with the ability of the auditory system to combine information over time, either via multiple looks or via a true integration process (Florentine, 1986; Viemeister and Wakefield, 1991; Eddins and Green, 1995). Another mechanism follows from extension of an argument proposed by von Klitzing and Kohlrausch (1994) to explain the fact that the overshoot effect (Zwicker, 1965) is greatest for mid-range levels (Bacon, 1990), and also to explain the deterioration in intensity discrimination that occurs at mid-range levels for clicks and for high-frequency tones presented in noise (Raab and Taub, 1969; Carlyon and Moore, 1984; Plack and Viemeister, 1992). The argument assumes that performance is partly limited by fluctuations in the background noise. The effects of the fluctuations can be either increased or decreased by basilar-membrane compression, depending on whether the signal or the background is compressed more. Note that we are assuming here that fluctuations in the portions of the noise preceding and following the signal may adversely affect performance, as the listener does not know the exact time of occurrence of the signal. Also, across-channel level comparisons may play a role, in which case noise fluctuations in frequency regions away from the signal frequency could affect performance.

Consider, as an example, a 1000-Hz tone burst with  $D = 5$  ms, presented in the 15-dB noise. The pedestal level in this case had a mean value of about 60 dB, well within the range where the basilar-membrane response is highly compressive. The effective noise level within a 1-ERB range around 1000 Hz was about 36 dB, a level where the basilar-membrane input–output function is usually only weakly compressive. The pedestal and pedestal-plus-increment would be subject to strong compression, while the noise would be subject to less compression. The net effect would be that the responses to the tone pulses would be brought closer in magnitude to the responses to the peaks in the noise, resulting in impaired intensity discrimination. As the value of  $D$  was increased, the level of the pedestal was decreased, and became closer to the effective level of the noise background. In this case, the compression of the signal

would be reduced, leading to improved intensity discrimination.

For the higher noise level, the situation would have been almost reversed; the noise level per ERB fell in the range where the basilar-membrane response is highly compressive, while the pedestal levels were in the higher level range where the response becomes more linear. Thus, the noise would have been compressed more than the signals to be discriminated. This could explain why intensity discrimination at 1000 and 4000 Hz was markedly better at the 40-dB noise level than at the 15-dB noise level or in the no-noise condition. It should be noted that van Schijndel *et al.* (1999) also used two noise levels, for a signal frequency of 1000 Hz. They found similar results for the two levels. This is consistent with our explanation, as for both noise levels the pedestal level for the 5-ms duration (the one giving the poorest intensity discrimination) fell in the range where the basilar membrane response would have been highly compressive.

Our failure to find distinct peaks in the functions relating  $\Delta I/I$  to duration for the 250-Hz carrier is hard to explain in terms of the theory based on f-t windows proposed by van Schijndel *et al.* (1999). However, it could be explained by reduced basilar-membrane compression at low frequencies (Cooper and Yates, 1994; Yates, 1995; Rhode and Cooper, 1996; Hicks and Bacon, 1999). The reduced compression could also explain the fact that, for the 15-dB noise condition, performance was generally better at 250 Hz than at 1000 or 4000 Hz. It could also explain why, at 250 Hz, performance was similar for the 15- and 40-dB noise levels. It is noteworthy that, for the 250-Hz carrier, performance was best for the no-noise condition. This may have happened because the “internal noise” that determines the absolute threshold has different statistical properties from the external noise. The latter, after passing through the auditory filter, would have relatively slow amplitude fluctuations, which might make intensity discrimination especially difficult.

In summary, the main features of the intensity discrimination results can be explained in terms of the input–output function of the basilar membrane and the way that this function changes with level and frequency. The results appear more consistent with this explanation than with an explanation based on f-t windows and multiple looks, as proposed by van Schijndel *et al.* (1999). However, it remains possible that both explanations are valid to some extent; the theory of van Schijndel *et al.* might account for the basic finding of a peak in  $\Delta I/I$  at an intermediate duration, while basilar-membrane compression might account for changes in the magnitude of the peak with noise level.

## B. Detection data

Comparison of the masked thresholds across frequency reveals an interesting effect. In theory, a pink noise should produce a roughly equal power output from each auditory filter (ignoring the effects of irregularities in the frequency response of the earphone and the transmission through the middle ear, which would affect the signal and masker to a similar extent in a given frequency region). If the signal threshold in noise corresponded to a constant ratio,  $R$ , of



signal energy to noise power at the output of the auditory filter (Fletcher, 1940; Patterson and Moore, 1986), one would expect the signal threshold to be roughly constant across frequency for a given duration. This is not, in fact, the case. For example, for a signal duration of 10 ms and a pink noise spectrum level of 15 dB at 1000 Hz, the mean threshold is about 52, 47, and 41 dB at 250, 1000, and 4000 Hz, respectively. The higher value at 250 Hz could be caused partly by the fact that the relative bandwidth of the auditory filter (bandwidth divided by center frequency) is slightly greater at 250 than at 1000 or 4000 Hz. However, changes in relative ERB across frequency cannot account for the 6-dB difference in threshold at 1000 and 4000 Hz. Even for long-duration tones, the value of  $R$  appears to decrease with increasing frequency (Patterson *et al.*, 1982; Patterson and Moore, 1986; Moore *et al.*, 1997). However, the decrease is small for frequencies above 1000 Hz. The change across frequency that we observed for our short duration stimuli appears to be greater than the effect observed for longer stimuli.

Comparable effects have been observed previously using a white noise rather than a pink noise masker, although many studies of temporal integration at masked threshold presented only the changes in threshold with duration, and not the absolute values of the thresholds (Plomp and Bouman, 1959; Florentine *et al.*, 1988). Dai and Wright (1996) presented extensive data using signal frequencies from 125 to 4000 Hz and durations from 2 to 1000 ms. The noise spectrum level was either 25 or 40 dB. They found that the threshold increased with increasing signal frequency for long, but not for short, signal durations. They reviewed several previous studies showing a similar trend (Garner and Miller, 1947; Gengel, 1972). For a white noise, the level per ERB increases with increasing center frequency, so the finding of roughly constant thresholds across frequency for short durations indicates that the value of  $R$  decreases markedly with increasing frequency, which is consistent with our results.

Dai and Wright attempted to account for their results using an energy-detector model (Green and Swets, 1974) in which listeners were assumed to use a time window that was matched to the signal duration and to use the output of a single auditory filter centered at the signal frequency. The model took into account the effect of the inherent fluctuations in the background noise and the fact that, at short durations, some of the signal's energy fell outside of the pass-band of the auditory filter centered at the signal frequency. The model predicted the general form of the results for longer signal durations reasonably well, but the observed thresholds were about 6 dB higher than the predicted thresholds. For short durations (2 and 8 ms), the model predicted that thresholds should decrease with increasing signal frequency, whereas the obtained thresholds were roughly constant across frequency.

A partial explanation for the discrepancies between the predictions of their model and the obtained results is that the time window used by the listeners may not be exactly matched to the signal duration. Presumably, there is some minimum duration over which integration occurs, which may

be comparable to the equivalent rectangular duration of the sliding temporal integrator estimated from experiments on temporal resolution (Festen *et al.*, 1977; Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994). Use of a time window longer than the signal duration would result in worse performance than for a "matched" time window and this could partly explain the 6-dB discrepancy in overall value between the obtained and predicted thresholds. A second explanation is connected with their assumption that listeners make use of the output of a single auditory filter centered at the signal frequency. For small signal durations, such as 2 ms, the bandwidth of the signal may greatly exceed the ERB of the auditory filter, for low center frequencies. Van den Brink and Houtgast (1990) have shown that listeners can integrate across frequency, i.e., they can combine information from the outputs of different auditory filters. Indeed, across-frequency integration is more efficient for short-duration signals than for long-duration signals. Dai and Wright (1996) discussed the possibility of such across-frequency integration, but did not take it into account in their model, which might explain why the model erroneously predicted that thresholds should decrease with increasing frequency for short-duration signals. Even for brief signals, the across-frequency integration is not equivalent to perfect energy integration. Thus, the fact that the signal bandwidth relative to the ERB of the auditory filter increased as the frequency decreased could have contributed to the across-frequency differences in threshold for the brief signals used in our experiment.

In theory, it might be expected that temporal integration at masked threshold would be affected by basilar-membrane compression; greater compression should lead to a steeper function relating threshold to duration (Penner, 1978; Carlyon *et al.*, 1990; Moore and Oxenham, 1998; Moore, 1998). Oxenham *et al.* (1997) presented some evidence in support of this idea. They measured thresholds for detecting 6.5-kHz sinusoidal tone bursts (duration 2 to 200 ms) in white noise with a spectrum level of -10, 20, or 50 dB. They found that, for durations up to 20 ms, the functions relating threshold to duration were steeper (had a more negative slope) for the medium noise level than for the higher or lower noise levels, consistent with greater basilar-membrane compression at medium levels. However, the slope difference was not apparent at longer durations.

In a second experiment, Oxenham *et al.* again used a 6.5-kHz signal, but measured the signal-to-masker ratio,  $S/N_0$ , necessary to detect the signal in white noise as a function of the noise spectrum level, for signal durations of 2, 10, and 200 ms. For the shortest duration, the value of  $S/N_0$  showed a local maximum for spectrum levels in the range 20–30 dB (signal levels in the range 66–76 dB). Thus, temporal integration was slightly greater at these noise levels than at higher or lower levels. The noise spectrum levels used in our experiment fell on either side of this range; at 4 kHz, the spectrum levels were 9 and 34 dB. Thus, we probably "missed" the steepening of the temporal integration function. However, the linear regression lines fitted to the data did have slightly more negative slopes at the higher

noise level, which is what would be expected from the data of Oxenham *et al.* (1997).

#### IV. CONCLUSIONS

- (1) Thresholds for the detection of brief Gaussian-shaped tone pulses (frequency 250, 1000, or 4000 Hz) in quiet or in a pink noise background (spectrum level at 1000 Hz either 15 or 40 dB) declined with increasing duration at a rate of about 3 dB per doubling of duration.
- (2) For a given duration, and for durations less than 20 ms, the ratio of the signal energy to the noise power at the output of the auditory filter centered at the signal frequency declined markedly with increasing signal frequency. This might reflect several factors including: (a) a reduced influence of the inherent fluctuations in the noise with increasing frequency; and (b) for very short durations, the proportion of the signal energy falling within the passband of one auditory filter increases with increasing frequency, and across-frequency integration may not be optimal.
- (3) Thresholds for the intensity discrimination of Gaussian-shaped tone pulses of various durations presented at 10 dB above the detection threshold in quiet or in a pink noise background showed a pattern of results that varied across frequency. At 1000 and 4000 Hz, performance was generally poorest for the lower noise spectrum level. For the lower noise spectrum level the thresholds showed a distinct peak (indicating poorer performance) at durations corresponding to about five periods of the carrier. Such peaks were still apparent but less clear in the no-noise condition and were almost absent at the higher noise level. At 250 Hz, performance was similar for the two noise levels, and there were not distinct peaks in the thresholds at a specific duration; performance tended to worsen with decreasing duration.
- (4) The pattern of results can be explained in terms of the compressive input–output function of the basilar membrane, and the way that the function changes with level and with frequency. This explanation appears more satisfactory than one based on the concept of elementary f-t windows and multiple looks, although both explanations may apply to some extent.

#### ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). We thank Kristin Rutter for gathering pilot data in connection with this experiment. We also thank Sid Bacon, Andrew Oxenham, and an anonymous reviewer for helpful comments on an earlier version of this paper.

<sup>1</sup>It is likely that the input–output function in response to a tone would be affected by the presence of background noise. However, neural rate versus level functions suggest that the slope of the input–output function is not increased by background noise, i.e., the response is not less compressive in the presence of background noise (Young *et al.*, 1983).

Bacon, S. P. (1990). "Effect of masker level on overshoot," *J. Acoust. Soc. Am.* **88**, 698–702.

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).

Buunen, T. J. F., and van Valkenburg, D. A. (1979). "Auditory detection of a single gap in noise," *J. Acoust. Soc. Am.* **65**, 534–537(L).

Carlyon, R. P., and Moore, B. C. J. (1984). "Intensity discrimination: a severe departure from Weber's Law," *J. Acoust. Soc. Am.* **76**, 1369–1376.

Carlyon, R. P., Buus, S., and Florentine, M. (1990). "Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners," *J. Acoust. Soc. Am.* **87**, 260–268.

Cooper, N. P., and Yates, G. K. (1994). "Nonlinear input-output functions derived from the responses of guinea-pig cochlea nerve fibers: Variations with characteristic frequency," *Hearing Res.* **78**, 221–234.

Dai, H., and Wright, B. A. (1996). "The lack of frequency dependence of thresholds for short tones in continuous broadband noise," *J. Acoust. Soc. Am.* **100**, 467–472.

Eddins, D. A., and Green, D. M. (1995). "Temporal integration and temporal resolution," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Festen, J. M., Houtgast, T., Plomp, R., and Smoorenburg, G. F. (1977). "Relations between interindividual differences of auditory functions," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Fletcher, H. (1940). "Auditory patterns," *Rev. Mod. Phys.* **12**, 47–65.

Florentine, M. (1986). "Level discrimination of tones as a function of duration," *J. Acoust. Soc. Am.* **79**, 792–798.

Florentine, M., Fastl, H., and Buus, S. (1988). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," *J. Acoust. Soc. Am.* **84**, 195–203.

Garner, W. R., and Miller, G. A. (1947). "The masked threshold of pure tones as a function of duration," *J. Exp. Psychol.* **37**, 293–303.

Gengel, R. W. (1972). "Auditory temporal integration at relatively high masked-threshold levels," *J. Acoust. Soc. Am.* **51**, 1849–1851.

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.* **47**, 103–138.

Glasberg, B. R., Moore, B. C. J., and Stone, M. A. (1999). "Modelling changes in frequency selectivity with level," in *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier (World Scientific, Singapore).

Green, D. M., and Swets, J. A. (1974). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Hicks, M. L., and Bacon, S. P. (1999). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," *J. Acoust. Soc. Am.* **105**, 326–338.

Hughes, J. W. (1946). "The threshold of audition for short periods of stimulation," *Philos. Trans. R. Soc. London, Ser. B* **133**, 486–490.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).

Moore, B. C. J. (1998). *Cochlear Hearing Loss* (Whurr, London).

Moore, B. C. J., and Glasberg, B. R. (1987). "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns," *Hearing Res.* **28**, 209–225.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). "A model for the prediction of thresholds, loudness and partial loudness," *J. Audio Eng. Soc.* **45**, 224–240.

Moore, B. C. J., and Oxenham, A. J. (1998). "Psychoacoustic consequences of compression in the peripheral auditory system," *Psychol. Rev.* **105**, 108–124.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1993a). "Detection of temporal gaps in sinusoids: Effects of frequency and level," *J. Acoust. Soc. Am.* **93**, 1563–1570.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1993b). "Effects of frequency on the detection of decrements and increments in sinusoids," *J. Acoust. Soc. Am.* **94**, 3190–3198.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1996). "Detection of decrements and increments in sinusoids at high overall levels," *J. Acoust. Soc. Am.* **99**, 3669–3677.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (1988). "The shape of the ear's temporal window," *J. Acoust. Soc. Am.* **83**, 1102–1116.

Oxenham, A. J., and Moore, B. C. J. (1994). "Modeling the additivity of nonsimultaneous masking," *Hearing Res.* **80**, 105–118.

Oxenham, A. J., and Plack, C. J. (1997). "A behavioral measure of basilar-

- membrane nonlinearity in listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **101**, 3666–3675.
- Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (1997). "Short-term temporal integration: Evidence for the influence of peripheral compression," *J. Acoust. Soc. Am.* **101**, 3676–3687.
- Patterson, R. D., and Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
- Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold," *J. Acoust. Soc. Am.* **72**, 1788–1803.
- Penner, M. J. (1978). "A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts," *J. Acoust. Soc. Am.* **63**, 195–201.
- Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (1995). "Effects of level and frequency on the detection of decrements and increments in sinusoids," *J. Acoust. Soc. Am.* **97**, 3791–3799.
- Plack, C. J., and Moore, B. C. J. (1990). "Temporal window shape as a function of frequency and level," *J. Acoust. Soc. Am.* **87**, 2178–2187.
- Plack, C. J., and Moore, B. C. J. (1991). "Decrement detection in normal and impaired ears," *J. Acoust. Soc. Am.* **90**, 3069–3076.
- Plack, C. J., and Viemeister, N. F. (1992). "The effects of notched noise on intensity discrimination under forward masking," *J. Acoust. Soc. Am.* **92**, 1902–1910.
- Plomp, R., and Bouman, M. A. (1959). "Relation between hearing threshold and duration for tone pulses," *J. Acoust. Soc. Am.* **31**, 749–758.
- Raab, D. H., and Taub, H. B. (1969). "Click-intensity discrimination with and without a background masking noise," *J. Acoust. Soc. Am.* **46**, 965–968.
- Recio, A., Rich, N. C., Narayan, S. S., and Rugero, M. A. (1998). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.* **103**, 1972–1989.
- Rhode, W. S., and Cooper, N. P. (1996). "Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*," *Auditory Neurosci.* **3**, 101–121.
- Robles, L., Ruggero, M. A., and Rich, N. C. (1986). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves, and response phases," *J. Acoust. Soc. Am.* **80**, 1364–1374.
- Rosen, S., and Stock, D. (1992). "Auditory filter bandwidths as a function of level at low frequencies (125 Hz–1 kHz)," *J. Acoust. Soc. Am.* **92**, 773–781.
- 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.
- Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.* **72**, 131–141.
- van den Brink, W. A. C., and Houtgast, T. (1990). "Efficient across-frequency integration in short-signal detection," *J. Acoust. Soc. Am.* **87**, 284–291.
- van Schijndel, N. H., Houtgast, T., and Festen, J. (1999). "Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window," *J. Acoust. Soc. Am.* **105**, 3425–3435.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.
- 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.
- Yates, G. K. (1995). "Cochlear structure and function," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
- Young, E. D., Costalupes, J. A., and Gibson, D. J. (1983). "Representation of acoustic stimuli in the presence of background sounds: Adaptation in the auditory nerve and cochlear nucleus," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin).
- Zwicker, E. (1965). "Temporal effects in simultaneous masking by white-noise bursts," *J. Acoust. Soc. Am.* **37**, 653–663.



# Intensity discrimination and detection of amplitude modulation

Magdalena Wojtczak<sup>a)</sup> and Neal F. Viemeister<sup>b)</sup>

Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 3 August 1998; accepted for publication 22 June 1999)

Thresholds for detection of low-rate sinusoidal amplitude modulation and for detection of intensity increments were measured over a wide range of levels in an examination of the relationship between these fundamental aspects of intensity processing. As expected, thresholds measured with a continuous 1-kHz tone decrease with increasing carrier/pedestal level. For levels between 6 and 85 dB SPL the data are well described by  $10 \log \Delta I/I = 0.44 \cdot (20 \log m) + D(f_m)$ , where  $\Delta I/I$  is the Weber fraction for increment detection,  $m$  is the modulation index at threshold, and  $D(f_m)$  depends on modulation rate ( $f_m$ ). The relationship between the psychometric functions for modulation and increment detection is also consistent with this equation. The data indicate a clear relationship between modulation and increment detection and thus provide an important additional consideration for models of modulation processing. No existing models provide an adequate account of this relationship. © 1999 Acoustical Society of America. [S0001-4966(99)02310-3]

PACS numbers: 43.66.Fe, 43.66.Mk [RVS]

## INTRODUCTION

It seems clear that intensity discrimination and the detection of amplitude modulation (AM) are related in some fundamental way in that both require the resolution of intensity or amplitude changes. Certainly, other factors are likely to distinguish between these tasks. Amplitude modulation detection at high modulation frequencies is limited by the ‘‘sluggishness’’ of the auditory system (Viemeister, 1979); intensity discrimination, at least for long-duration signals, is not. Amplitude modulation detection may involve the detection of periodicities, intensity discrimination does not. Spectral cues can be involved in AM detection when sinusoidal carriers are used but such cues generally are not involved in intensity discrimination. Different mechanisms may be involved in AM detection and intensity discrimination. For example, AM processing may involve channels that are tuned to specific modulation frequencies (Dau *et al.*, 1997a, b). Intensity discrimination with continuous pedestals (increment detection) may involve ‘‘change detectors’’ (Macmillan, 1971; Hafter *et al.*, 1997).

Despite these potentially important differences, performance in both situations is determined, at least partially, by intensity resolution and thus the tasks are conceptually related. The empirical evidence strongly suggests a relationship. The level effects seen in intensity discrimination have counterparts in AM detection. Weber’s law holds over a wide range of levels for intensity discrimination of broadband noise (for a recent review see Plack and Carlyon, 1995). Similarly, modulation thresholds (expressed as modulation depth, a relative measure) are constant for broadband noise carriers over a wide range of levels (Viemeister, 1979; Bacon and Viemeister, 1985). The ‘‘near-miss’’ to Weber’s law seen for intensity discrimination of tones (McGill and Goldberg, 1968) has a counterpart in the decrease in modulation thresholds with increasing carrier level for tonal carriers (see Kohlrausch, 1993).

In his classic study, Riesz (1928) measured beat detection thresholds, essentially AM thresholds, as a function of carrier level and carrier frequency. To compare these thresholds to ‘‘classical’’ intensity jnd’s (just-noticeable differences), he proposed a formula relating the Weber fraction ( $\Delta I/I$ ) to the relative level of the two beating components. This formula and its validity will be considered in Sec. III. For now, it should be noted that Riesz’s formula was not based on evidence, but was simply a reasonable conjecture about the relationship between AM and increment detection.

Stiegel (1977) compared intensity discrimination and AM detection using 500-ms, 1-kHz gated carriers and pedestals. He concluded, essentially, that none of the threshold measures he considered, including Riesz’s, adequately described the relationship between intensity discrimination and modulation detection. Bacon and Viemeister (1985) measured AM thresholds and increment detection thresholds with hearing-impaired subjects and with subjects with normal hearing. Continuous noise carriers were used and, because a wide variety of conditions were examined, there was a fairly large range of thresholds. As in Stiegel’s study, however, the range of thresholds was not large enough to thoroughly evaluate the validity of Riesz’s conjecture, and the variability of the data precluded strong statements about the relationship between AM and increment detection.

The present study is similar to Bacon and Viemeister (1985), but, unlike that study, is specifically designed to examine the relationship between AM and increment detection. To obtain a large range of thresholds, sinusoidal carriers/pedestals are used with a very wide range of levels. (As noted above, the change in AM and increment thresholds with level is much larger for tonal carriers and pedestals than for broadband noise.) Low modulation frequencies (2, 4, and 16 Hz) are used to ensure that AM detection is minimally affected by temporal resolution or by the use of spectral cues. Unlike in Stiegel (1977), continuous carriers and pedestals are used, primarily because the temporal modulation transfer function (TMTF) measured with gated sinusoidal carriers and low modulation frequencies shows a high-pass segment (Viemeister, 1979). This effect is poorly understood

<sup>a)</sup>On leave from Institute of Acoustics, Adam Mickiewicz University, Poznan, Poland; electronic mail: magda@nextear.psych.umn.edu

<sup>b)</sup>Electronic mail: nfv@nextear.psych.umn.edu



(see Yost and Sheft, 1997) and the use of gated carriers in the present study thus would have complicated interpretation of the results. To examine the generality of our findings, AM and increment detection thresholds at several carrier/pedestal frequencies were measured, as were psychometric functions for 1-kHz carrier/pedestals.

The general question addressed in this study is whether there is a well-defined relationship between envelope processing and intensity discrimination, two fundamental aspects of hearing. This question has not been thoroughly addressed either experimentally or theoretically. We show that, indeed, there is a strong empirical relationship, one that generalizes across frequency and to the psychometric functions. Overall, these results provide additional clues about the processes involved in envelope detection and they point to areas in which current models of such processes need refinement.

## I. METHOD

### A. Stimuli and procedure

Increment detection thresholds and AM-detection thresholds were measured using an adaptive two-down one-up 3IFC procedure that estimated the 70.7% correct point on the psychometric function (Levitt, 1971).

For most of the conditions in the increment detection task, a 1-kHz pedestal was presented continuously throughout a block of trials. Its intensity was incremented in the signal interval by adding a 510-ms, 1-kHz tone with 5-ms raised-cosine ramps in phase with the pedestal. Three 510-ms observation intervals, separated by 500 ms, were indicated on the computer monitor. The relative intensity increment [ $10 \log(\Delta I/I)$ ] was decreased by 4 dB after two correct responses and increased by the same amount after each incorrect response until four reversals were obtained. The step size was then reduced to 2 dB. The threshold estimate was the mean of the last eight reversals (12 reversals per block). A block was discarded when the standard deviation of the last eight reversals exceeded 3.5 dB. This happened on fewer than 10% of the runs. Thresholds from three blocks were averaged to produce the final threshold estimate. The experimental conditions were run in order from the highest to the lowest pedestal level. Replications of selected conditions indicated no order effects.

In the AM-detection task, the carrier was a 1-kHz tone which was presented continuously through the entire block of trials. It was amplitude modulated for 500 ms in one of the three observation intervals by a sinusoid starting in sine phase plus a dc component. The modulation frequency for most conditions was 4 Hz, corresponding approximately to the rate at which modulation thresholds are at a minimum for continuous tonal carriers as reported in other studies (e.g., Yost and Sheft, 1997). Thresholds for modulation frequencies of 2 and 16 Hz also were obtained, but only at four carrier levels. The tracking procedure was the same as that described above for the increment detection task, except that the modulation depth ( $20 \log m$ ) was adjusted, in 4- then 2-dB steps. The intensity was not adjusted to compensate for the increase due to amplitude modulation. This did not affect performance—in all cases the intensity increase produced by

modulation was much smaller than the increment threshold measured for that condition. The criterion for discarding runs was the same as in the increment detection task. Fewer than 10% of the runs were discarded.

In addition to the extensive data obtained with a 1-kHz carrier, a modulation rate of 4 Hz was used to measure AM detection at carrier frequencies ranging from 250 Hz to 8 kHz with the carrier presented at a level of 60 dB SL. Likewise, a pedestal level of 60 dB SL was used to measure increment detection for the same pedestal frequencies. The procedures used to estimate the thresholds were identical to those described above.

Psychometric functions for increment detection and AM detection were measured using a 3IFC procedure with a continuous 1-kHz pedestal/carrier presented at a level of 65 dB SPL. Psychometric functions for 4-Hz AM detection were also measured with the carrier gated on only during the observation intervals. Percent correct [ $P(C)$ ] was measured for eight increment levels, selected in steps of 0.5 dB [ $10 \log(\Delta I/I)$ ] and eight modulation depths, selected in steps of 1 dB ( $20 \log m$ ). The increments and modulation depths were chosen to produce a range of performance from near chance to near perfect. For every increment and modulation depth, eight blocks consisting of 50 trials were run. Thus, 400 responses per data point were obtained. Each listener first was run in the easiest condition, yielding almost perfect performance. This condition was treated as practice and the data were discarded, so seven points contributed to the measured psychometric functions. Increments and modulation depths were presented in descending order across several experimental sessions.

All stimuli were generated digitally on a NeXT computer using a 16-bit D/A and a sampling rate of 44.1 kHz. A fixed analog attenuator was used to achieve the desired pedestal/carrier level. Stimuli were presented monaurally through Sony MDR-V6 headphones.

### B. Subjects

Data were collected from seven listeners, not all of whom participated in every condition. Four listeners, two of whom had previous experience in psychoacoustic experiments, performed the AM- and increment-detection task using a 1-kHz pedestal/carrier. Psychometric functions were obtained from three listeners, one of whom had earlier experience in AM and increment detection. Three listeners, one of whom did not participate in any earlier testing, were used to measure AM and increment detection across pedestal/carrier frequencies. The inexperienced subjects were given practice until asymptotic performance was observed. The subjects [except S1 (author MW)] were paid for their service. All subjects had normal hearing at all audiometric frequencies.

## II. RESULTS

Increment-detection thresholds for four listeners are presented in Fig. 1(a). Three listeners (S1–S3) showed the “near miss” to Weber’s law, i.e., their Weber fractions decreased as the pedestal level increased. For these listeners the

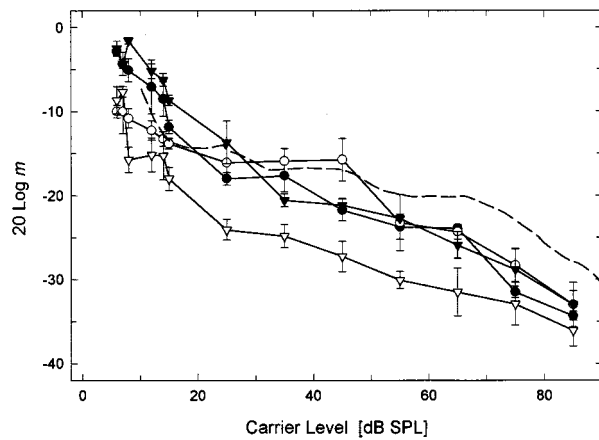
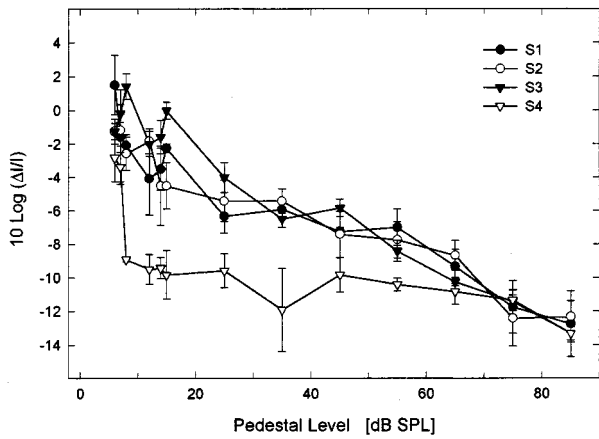


FIG. 1. Increment thresholds (a, upper panel) and 4-Hz modulation thresholds (b) versus the level of the 1-kHz, continuous pedestal or carrier. The dashed line in (b) is from Zwicker and Graf (1987).

between-subject differences also were very small over the whole range of levels and their increment-detection thresholds are very similar, considering the performance difference between 2IFC and 3IFC, to those shown by Viemeister and Bacon (1988). The linear fit to the averaged data for pedestal levels between 15 and 85 dB SPL gives a slope of  $-0.11$ , which is similar to the slope of  $-0.09$  reported by Viemeister and Bacon for a continuous pedestal. Surprisingly, S4 did not show the near-miss: his thresholds were almost equal across levels between 8 and 85 dB SPL and at low levels were considerably better than those for the other listeners. We have no explanation for this unusual result.

Figure 1(b) shows AM detection thresholds for the individual listeners. In this task all four listeners showed a decrease in AM thresholds with increasing carrier level. As in increment detection, the thresholds for S4 were generally lower, sometimes substantially, than those for the other listeners. Furthermore, for listener S4 the form of the modulation-detection function was different from that for the increment-detection function.

The data shown in Fig. 1(b) are in good agreement with most of the earlier data on AM detection using continuous 1-kHz carriers (see Kohlrausch, 1993, for a summary). The most complete set of data obtained under conditions similar to ours are those of Zwicker and Graf (1987) and are indi-

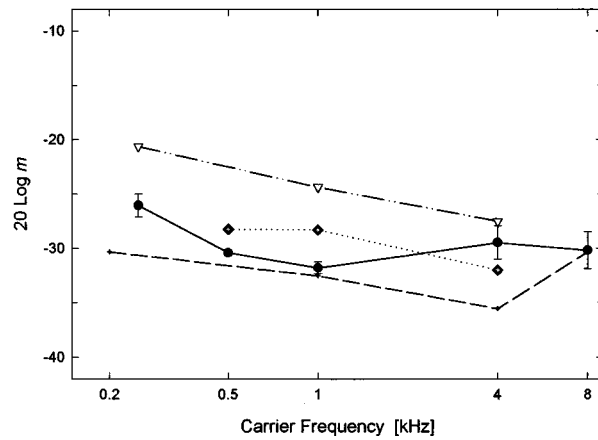
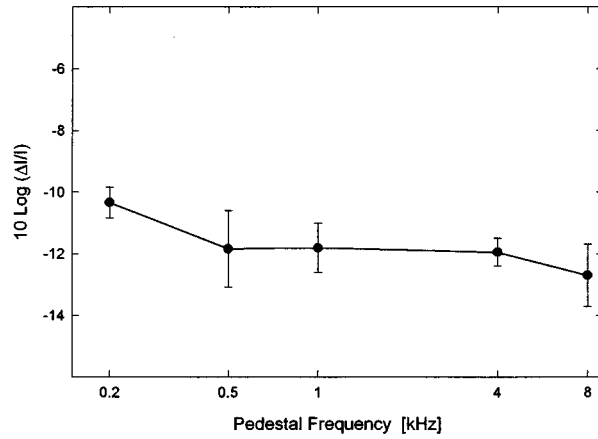


FIG. 2. Mean increment detection thresholds (upper panel) and mean 4-Hz modulation detection thresholds (lower panel, filled circles) plotted as a function of pedestal/carrier frequency. The dashed line in the bottom panel also shows thresholds from Riesz (1928) measured with a carrier level of 60 dB SL, the dotted line shows thresholds from Yost and Sheft (1997) for a carrier level of 70 dB SPL, and the dashed-dotted line shows thresholds from Zwicker and Graf (1987) for a carrier level of 70 dB SPL.

cated by the dashed line in Fig. 1(b). The agreement in the form of these functions is good. In contrast, Riesz's (1928) data, when converted back to  $20 \log m$ , disagree with these results in that they show approximately constant sensitivity for levels greater than about 60 dB SL.<sup>1</sup>

The AM-detection thresholds for modulation frequencies of 2 and 16 Hz were on average about 3 dB higher than those shown in Fig. 1(b). This is consistent with the data of Yost and Sheft (1997). For both frequencies the decrease in thresholds with increasing carrier level is similar to that shown for 4 Hz.

Figure 2 shows the effects of carrier/pedestal frequency when the carrier is presented at 60 dB SL. For increment detection (top panel) there is a small decrease in the Weber fraction from 250 to 500 Hz. From 500 Hz to 8 kHz, the Weber fraction is approximately constant, consistent with data obtained with gated pedestals (Jesteadt *et al.*, 1977). The AM-detection thresholds (lower panel, filled circles) decrease from 250 to 1 kHz and then may increase slightly. Also shown are data obtained with continuous carriers from Riesz (1928), from Yost and Sheft (1997), and from Zwicker and Graf (1987). With gated carriers Yost and Sheft (1997)

TABLE I. Slopes of psychometric functions fitted to the data using linear regression on coordinates  $\log d'$  vs  $10 \log(\Delta I/I)$  for increment detection, and  $\log d'$  vs  $20 \log m$  for AM detection. Goodness-of-fit measures are shown in parentheses.

Subject no.	Increment detection continuous pedestal	AM detection continuous carrier	AM detection gated carrier
1	0.18 ( $r^2=0.97$ )	0.09 ( $r^2=0.94$ )	0.11 ( $r^2=0.93$ )
2	...	0.11 ( $r^2=0.93$ )	0.14 ( $r^2=0.95$ )
3	0.18 ( $r^2=0.89$ )	0.09 ( $r^2=0.92$ )	...

showed approximately constant AM-detection thresholds.

Psychometric functions were constructed from the percentage of correct responses assuming unbiased observers. This assumption was reasonable as each interval was chosen by every listener approximately in  $\frac{1}{3}$  of all trials. The psychometric functions were well fitted by linear functions in  $\log d'$  vs  $10 \log(\Delta I/I)$  or  $\log d'$  vs  $20 \log m$ . The slopes and a measure of goodness-of-fit are given in Table I.

### III. DISCUSSION

The basic question addressed in this paper is whether there is a well-defined relationship between increment detection and AM detection. The data shown in Fig. 3 suggest that there is. In this scatter plot, the average increment detection thresholds for the four listeners are plotted against the average AM-detection thresholds measured for a 4-Hz modulation rate. For each point, the carrier and pedestal levels are equal and, in general, the higher thresholds were obtained at lower levels.

The linear function fitted to these data<sup>2</sup> has a slope of 0.39 and a rms deviation of 0.5 dB ( $r^2=0.97$ ). Although this fit is reasonably good, it should be noted that for low pedestal levels the average Weber fractions are driven down by the unusually low values produced by listener S4. The slopes for individual subjects together with goodness-of-fit measures are given in Table II. Excluding the data from S4 leads to

$$10 \log \frac{\Delta I}{I} = 0.44 \cdot (20 \log m) + 1.7 \quad (1)$$

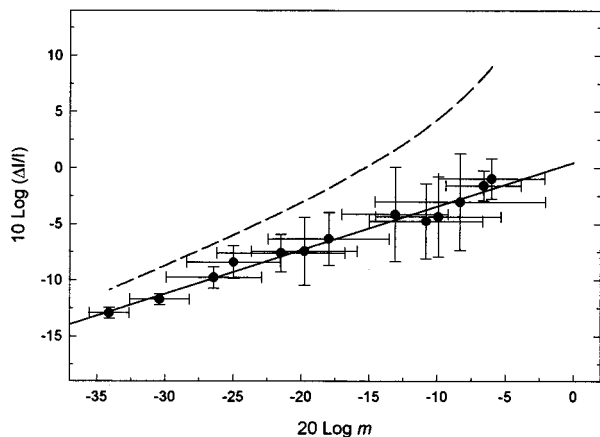


FIG. 3. Scatterplot showing the relation between increment thresholds and modulation thresholds. Each point is the average of the thresholds for four subjects at a given carrier/pedestal level. The solid line is the regression fit, the dashed line is based on Riesz's formula.

TABLE II. Slopes of regression fits on coordinates  $10 \log(\Delta I/I)$  vs  $20 \log m$ , and goodness-of-fit measures for individual subjects.

Subject	Slope of regression fit	rms deviation (dB)	$r^2$
S1	0.38	1.0	0.93
S2	0.50	0.95	0.92
S3	0.42	1.1	0.95
S4	0.30	1.5	0.71

with a rms deviation of 0.5 dB. In our opinion this equation most accurately describes the relationship between increment detection and AM detection at 4 Hz for the typical listener.

### A. Interpretations

A simple interpretation of these results is that for AM the just-detectable intensity increment ( $\Delta I$ ) corresponds to half the difference between the maximum and the minimum intensity of the modulated waveform, while the reference intensity ( $I$ ) corresponds to the carrier intensity. This produces the following:

$$\frac{\Delta I}{I} = 2m,$$

which in logarithmic units yields

$$10 \log \frac{\Delta I}{I} = 0.5 \cdot (20 \log m) + 3. \quad (2)$$

This equation provides only a slightly worse fit than that provided by the linear regression analysis. A similar relation can be obtained by considering relative *amplitude* changes and assuming  $\Delta A/A = m$ . This leads to  $\Delta I/I = 2m + m^2$ .

An alternative interpretation was provided by Riesz (1928). He proposed that the change in intensity necessary to detect beats corresponded to the difference between the maximum intensity and the minimum intensity of the just-detectable fluctuations, and that the minimum intensity corresponds to the reference intensity. Given that the beats produced approximately sinusoidal envelope fluctuations, it seems justifiable to rewrite Riesz's formula in terms of the modulation index<sup>3</sup> to facilitate comparison between the present data and his conjecture. This yields

$$\frac{\Delta I}{I} = \frac{(1+m)^2 - (1-m)^2}{(1-m)^2} = \frac{4m}{(1-m)^2}. \quad (3)$$

The dashed line in Fig. 3 is from Eq. (3). At the lower thresholds, Eq. (3) overpredicts the Weber fraction by about 3 dB. This could be corrected by assuming, as we did above, that the jnd corresponds to half the difference between the maximum and the minimum intensity of the modulated waveform. More serious is the discrepancy at higher thresholds where the form of the function appears to be considerably different from that describing the data. This, of course, is a consequence of the denominator of Eq. (3) approaching zero, i.e., when the reference intensity ( $I$ ) approaches zero. A reasonable additional modification of Riesz's conjecture would be to include an additive constant corresponding to absolute threshold. Then, at low carrier levels the minimum of the fluctuations and the reference intensity would be de-

terminated by absolute threshold and the predicted Weber fraction would be decreased. Since the major discrepancy between Eq. (3) and the data occurs at low carrier levels (where the modulation thresholds are large), this “correction” for absolute threshold will yield predictions in better agreement with the data.

None of these interpretations is especially compelling because they only hint at the fundamental aspects of the relationship between increment detection and modulation detection. What is missing is a sound theoretical basis for these interpretations. In current terminology, the issue is whether a decision variable that can account for AM detection can also account for intensity discrimination. Various decision variables have been proposed for AM detection (see Strickland and Viemeister, 1996) and while none are entirely satisfactory, it appears that the “max/min” statistic proposed by Goldstein (1967) and further considered by Forrest and Green (1987) provides a reasonable account of a variety of AM detection data. Although this decision variable is based on amplitudes, it yields a prediction identical to Riesz’s (see the Appendix) and thus suffers the same inadequacies. Possible modifications are similar to those suggested above. A problem with the max/min statistic is that its underlying premise is only slightly less contrived than those used to derive Eqs. (2) and (3)—how are envelope maxima and minima extracted; why a max/min *ratio*? Perhaps more importantly, the max/min statistic does not adequately account for the psychometric function for modulation detection (see below).

## B. Generality

### 1. Modulation frequency

Modulation thresholds depend on modulation frequency and therefore the descriptions of the data for 4 Hz will not generalize to other modulation frequencies. It appears, however, that the linear relation between the Weber fraction and modulation threshold (in the decibel coordinates of Fig. 3) is valid for other modulation frequencies. Figure 4 shows that the modulation thresholds at 2 and 16 Hz were greater than those at 4 Hz by a constant amount (in dB) independent of carrier level. Generalizing Eq. (1),

$$10 \log \frac{\Delta I}{I} = 0.44 \cdot (20 \log m) + D(f_m), \quad (4)$$

where  $D(f_m)$  is in units of  $20 \log m$  and depends only on the modulation frequency,  $f_m$  [ $D(f_m) = 1.7$  at  $f_m = 4$  Hz]. The lines shown in Fig. 4 are based on Eq. (4) and provide a good description of the data. This suggests that the effects of modulation frequency are equivalent to a frequency-dependent attenuation of the effective modulation depth. Although the reason for the changes in modulation thresholds at low modulation frequencies is not understood, it is somewhat reassuring that they appear to be equivalent to a simple attenuation and thus consistent with the usual interpretations of TMTFs as attenuation characteristics.

### 2. Carrier frequency

Figure 5 shows the mean increment-detection thresholds replotted from Fig. 2 (filled circles) and the thresholds pre-

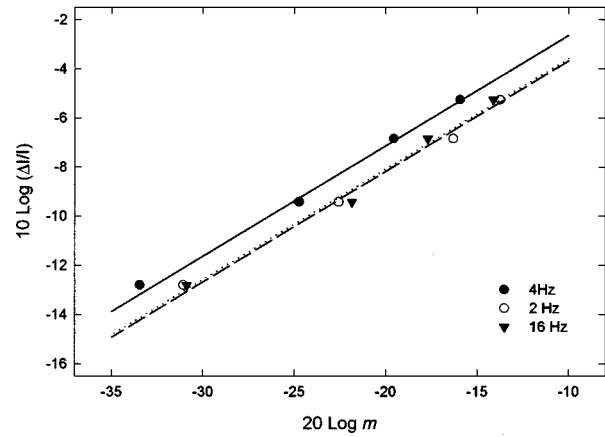


FIG. 4. As in Fig. 3 but for additional modulation frequencies of 2 and 16 Hz. Data are also shown for 4 Hz measured at the same carrier/pedestal level used for 2 and 16 Hz. The solid line is from Eq. (1). The data were obtained for three subjects and exclude the 4-Hz data for S4.

dicted from the mean AM-detection thresholds based on Eq. (1) (open circles). The difference between the measured and predicted thresholds is less than 1 dB for all pedestal frequencies except 8 kHz, for which that difference is 1.1 dB. This suggests that Eq. (1) generalizes across carrier/pedestal frequency. Specifically,  $D(f_m)$  in Eq. (4) does not appear to depend on carrier/pedestal frequency, at least at 60 dB SL and over the frequency range examined. The effects of level were not examined at carrier/pedestal frequencies other than 1 kHz.

### 3. Psychometric functions

For intensity discrimination using *gated* pedestals  $d'$  is approximately proportional to  $\Delta I$  (cf. Buus and Florentine, 1991), but for modulation detection using *gated* carriers  $d'$  is proportional to  $m^2$  (Stiegel, 1977; Moore and Sek, 1992; Edwards and Viemeister, 1994). This is not consistent with our summary of the present data [Eq. (4)] which predicts that  $d'$  should be proportional to  $(\Delta I/I)^{1/0.44}$ , assuming that for modulation detection  $d'$  is proportional to  $m^2$ . Thus, the empirical relationship between increment and modulation

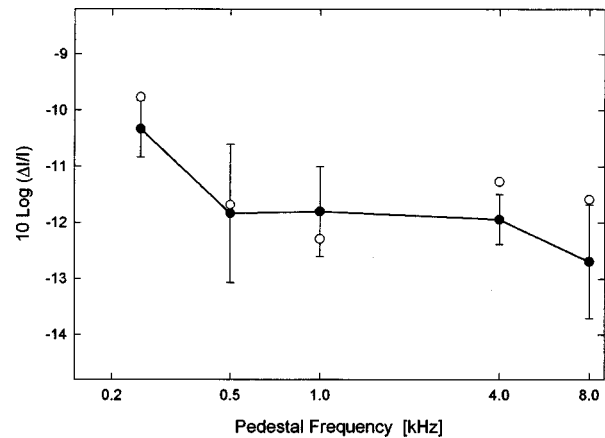


FIG. 5. Mean increment detection thresholds replotted from Fig. 2 (filled circles) and the increment detection thresholds predicted from AM detection based on Eq. (1).



detection depends on the level of performance used to define threshold. In the context of Eq. (4),  $D(f_m)$  should be generalized to  $D(f_m, d')$ . This is not very satisfying because it is *post hoc* and ignores what may be a fundamental difference between intensity discrimination and modulation detection for gated pedestals and carriers.

Recall, however, that for reasons discussed in the Introduction, the present experiments used *continuous* carriers and pedestals. For continuous pedestals (increment detection) it appears that  $d'$  is not proportional to  $\Delta I$ , as it is for gated pedestals, but is approximately proportional to  $\Delta I^2$  (Green and Nachmias, 1979). Although there are several possible explanations (uncertainty, small-signal suppression, overintegration, adaptation, change detection), the reason for this difference, and, more generally, for gated-continuous differences in both intensity discrimination and modulation detection, remains unclear.

Table I shows the slopes of the psychometric functions. For increment detection (column 2) and for modulation detection with gated carriers (column 4) the results are consistent with those discussed above. The new result is that the slope of the psychometric function for modulation detection with continuous carriers (column 3) is similar to that for gated carriers. This result seems to disagree with the observation by Sheft and Yost (1996), in which the psychometric functions measured with a continuous carrier were steeper than those measured with a gated carrier. It should be noted, however, that Sheft and Yost used noise carriers whereas the present study used tonal carriers. Also, for all modulation rates they tested, except a 2-kHz modulation rate, the difference in the slopes observed with gated and continuous carriers was small (e.g., 1.19 vs 1.49 for a 5-Hz modulation rate).

The slopes of the psychometric functions measured in the present study using continuous carriers and pedestals suggests that Eq. (4) generalizes to different levels of performance, i.e., no correction for different  $d'$ 's is necessary. Perhaps more importantly, Eq. (4) predicts that the slope of the psychometric function for modulation detection should be less than that for increment detection by a factor of 0.44. This is close to that observed in the data of Table I and suggests that Eq. (4) is tapping something fundamental about the relationship between intensity discrimination and modulation detection for continuous pedestals/carriers.

It appears that the form of the psychometric function for intensity discrimination using gated pedestals is the unusual result. The most plausible account, in our opinion, is that this results from memory limitations: intensity discrimination with gated pedestals presumably involves an across-interval comparison and therefore involves memory; in the other conditions good performance can be achieved using within-interval cues and thus the memory requirement is less (Harris, 1963). This is closely related to the distinction between sensory-trace and context coding (Braidia *et al.*, 1984) and the "change detector" hypothesis (see Hafter *et al.*, 1997), but does not appear to explain why the slope of the psychometric function is less for intensity discrimination using gated pedestals.

## C. Other observations

There are additional reports that permit some degree of assessment of the applicability of Eq. (4) across conditions that were not studied here.

### 1. Gated carriers

Stiegel (1977) compared intensity discrimination and AM detection using gated 1-kHz carriers and pedestals. An analysis of his data using the approach used for Fig. 3 indicated no clear relationship. This, in part, reflects the limited range of levels (15 to 60 dB SPL): the changes in thresholds were relatively small over this range and so the regression analysis was not informative. It is our opinion that Eq. (4) may generalize to gated carriers/pedestals but, as discussed above, it is not consistent with the psychometric functions for these conditions.

### 2. Noise carriers

As mentioned in the Introduction, Bacon and Viemeister (1985) examined the relationship between AM and increment detection using continuous noise carriers. Their regression analysis indicated that a description of the data ( $r^2=0.72$ ) was provided by the equation  $10 \log(\Delta I/I) = 0.31(20 \log m) - 1.9$ , where  $20 \log m$  was the sensitivity to AM, computed as the average of modulation thresholds in dB at 4 and 8 Hz. A reanalysis of their data using Eq. (1) yielded a rms error of 1.2 dB (vs 0.74 dB for their fit). We conclude that Eq. (1) may describe the relationship between AM and increment detection shown by Bacon and Viemeister, but that the variability in the data precludes strong conclusions.

It should be noted that Bacon and Viemeister's regression analysis was based on data from a wide variety of conditions, including lowpass noise carriers and thresholds from hearing-impaired subjects. A far less complicated, but less general, situation involves comparison using broadband noise carriers in subjects with normal hearing. Here, Weber's law holds for both increment and AM detection and thus is consistent with Eq. (1). Furthermore, Viemeister's (1979) data show a low-rate modulation threshold of  $-24.5$  dB. Equation (1) predicts that the increment threshold should be  $-9.1$  dB. This is very close to the observed  $-10$  dB.

## D. Effects of duration

Although there are no data that permit direct assessment of the generality of Eq. (4) to account for the effects of duration (at a fixed level), this equation implies that the slope of the temporal integration function (dB per decade change in duration) should be a factor of 0.44 less for increment detection than for modulation detection. Sheft and Yost (1990) found slopes of  $-7.46$  and  $-9.36$  dB/decade for gated and continuous noise carriers, respectively. The predicted slopes of  $-3.28$  and  $-4.11$  dB/decade compare with observed slopes of  $-3.6$  and  $-6.5$  dB/decade shown by Viemeister (1988) for intensity discrimination using gated and continuous noise pedestals.

The data using pure-tone carriers are limited. Yost and Sheft (1997) measured modulation thresholds for durations

of 125 and 500 ms. For gated carriers and low modulation rates (8 and 16 Hz) this fourfold increase in duration resulted in approximately a 10-dB decrease in threshold, independent of carrier frequency. This yields a predicted decrease in the Weber fraction of 7.3 dB/decade, a value somewhat larger than the 4.4 dB/decade shown by Florentine (1986) for 1-kHz pedestals. For continuous pure-tone carrier Yost and Sheft showed decreases ranging from 13 to 5 dB, depending on carrier frequency. The corresponding predicted decreases are 9.5 and 3.7 dB/decade, respectively. Oxenham (1997) showed that increment detection for 4- and 8-kHz quasi-continuous pedestals is essentially independent of duration for durations longer than 16 ms. This is clearly inconsistent with Eq. (4). It also appears to be inconsistent with the data of Garner and Miller (1944) which showed for increments longer than 30 ms, an average decrease of 3.8 dB/decade for 1-kHz continuous pedestals. The Oxenham data also are inconsistent with the effects of duration on increment detection using noise pedestals discussed above.

Overall, it appears that for a fixed level, Eq. (4) can roughly account for the effects of duration except perhaps for continuous pure-tone carriers, the condition which is the focus of this paper. An explanation is that increment detection is based on change detection (Oxenham, 1997; see Hafter *et al.*, 1997), and thus is essentially duration independent, while modulation detection involves multiple looks over modulation cycles (Viemeister, 1979) or some form of long-term integration (Sheft and Yost, 1990). Equation (4) can, of course, be generalized by allowing  $D(f_m)$  to depend on the duration of modulation. It is not clear, however, whether such a simple modification will be consistent with the effects of level over a wide range of durations.

#### IV. SUMMARY AND CONCLUSIONS

The present results indicate that intensity discrimination using continuous pedestals and modulation detection using continuous carriers are closely related. In particular, we found the following.

- (1) For a modulation frequency of 4 Hz, Eq. (1) provides an excellent description of the data for three of the four subjects over a wide range carrier/pedestal levels and modulation thresholds.
- (2) For modulation frequencies of 2 and 16 Hz, modulation thresholds were higher than those at 4 Hz by a constant amount (in dB), indicating that Eq. (1) can be simply generalized to other modulation frequencies [Eq. (4)]. The range of modulation frequencies for which Eq. (4) applies is not known. It will not be valid at high modulation frequencies when modulation detection with 1-kHz carriers involves spectral resolution of the sidebands.
- (3) Equation (4) holds for different points on the psychometric functions chosen to define increment-detection and AM-detection thresholds. More importantly, it describes the relationship between the psychometric functions for these situations: the slope of the AM function should be approximately half that for increment detection, as observed in the present study. However, Eq. (4) does not

predict the relationship between the psychometric functions for modulation detection and intensity discrimination with gated carriers/pedestals

- (4) Equation (4) appears to be valid for carrier/pedestal frequencies ranging from 250 to 8000 Hz, at least at 60 dB SPL.
- (5) Based upon other published data, it appears that Eq. (4) may be valid for noise carriers/pedestals. It also roughly describes the effects of duration except for conditions involving continuous pure-tone carriers/pedestals.
- (6) Riesz's (1928) quantitative conjecture about the relationship between beat (AM) detection and intensity discrimination is incorrect. It leads to significantly higher intensity  $jnd$ 's than those observed, especially at low carrier levels. Also, the level effects shown by Riesz differ from those of the present study and from other published data. Considering this discrepancy, and the discrepancy between Riesz's formula and our empirical summary of the relationship between intensity  $jnd$ 's and AM detection [Eq. (4)], we believe that Riesz's classic study should not be used as a description of intensity discrimination (cf. Allen and Neely, 1997).

Finally, although we have shown empirically that there is a strong relationship between increment and AM detection, we cannot offer a compelling theoretical description of this relationship. Several interpretations of the detection process are approximately consistent with Eq. (4), but none are entirely satisfactory, even after modifications to include limits related to absolute threshold. Of those considered, a modified version of the max/min decision statistic is perhaps the most appealing, primarily because it has been used with moderate success to account for other data on AM detection. However, it does not describe the psychometric function for AM detection and, in our opinion, it suffers, as do the alternate interpretations, from its contrived nature. Clearly, we need a deeper theoretical understanding of AM processing and, now, how it relates to the fundamental ability to detect intensity changes in quasi-continuous sounds.

#### ACKNOWLEDGMENTS

We thank Dr. Armin Kohlrausch for his help in translating and Dr. Mark Stellmack and Dr. Martin Rickert for stimulating discussions about this research. We also thank the reviewers for their constructive suggestions. This work was supported by Grant No. DC00683 from NIDCD.

#### APPENDIX: PREDICTIONS BASED ON MAX/MIN

For the max/min statistic, the decision variable is the ratio of the maximum of the (processed and noisy) envelope to its minimum. For modulation detection with deterministic stimuli, the sensitivity index  $d'$  is

$$d' = \frac{((1+m)/(1-m)) - 1}{\sigma}, \quad (A1)$$

where the first term is the mean of the max/min ratio when AM is present, and  $\sigma$  is the variability produced by an additive Gaussian, internal noise. For increment detection,

$$d' = \frac{((A + \Delta A)/A) - 1}{\sigma}, \quad (\text{A2})$$

where  $A$  is the amplitude corresponding to the intensity of the pedestal and  $\Delta A$  is the increment in amplitude produced by  $\Delta I$ . For equal  $d'$ ,

$$\frac{A + \Delta A}{A} = \frac{(1 + m)}{(1 - m)},$$

where  $\Delta A$  and  $m$  now are the threshold amplitude increment and modulation depth, respectively. Squaring both sides and expressing in intensity units,

$$\frac{I + \Delta I}{I} = \frac{(1 + m)^2}{(1 - m)^2}, \quad \frac{\Delta I}{I} = \frac{4m}{(1 - m)^2}.$$

The last equation is identical to Eq. (3), that proposed by Riesz.

Equation (A1) simplifies to  $d' = 2m/(1 - m)\sigma$ , which, as previously noted, is inconsistent with observed psychometric functions for modulation detection. Equation (A2) simplifies to  $d' = \Delta A/A\sigma \approx \Delta I/2I\sigma$ . This also is not consistent with the psychometric functions obtained for increment detection.

<sup>1</sup>Harris (1963) attempted a "repetition" of Riesz's study and, although he concludes that his data are in close correspondence with Riesz's, there is a decrease in threshold between 60 and 80 dB, the highest level tested. The decrease is obscured by his use of the decibel-difference measure that was also used by Riesz.

<sup>2</sup>The data were fitted using a least-squares procedure that takes into account the fact that both of the measured variables are subject to error (Mandel, 1984). For the present data, the fits are negligibly different from those using the typical procedure which assumes no measurement error in the independent variable. The rms deviation reported here is based on the distance between each data point and the fitted function measured on a line orthogonal to the function.

<sup>3</sup>We are defining the modulation index for beats to be half the peak-trough difference in the envelope divided by the average amplitude. This is equivalent to ratio of the amplitudes of the two components,  $r$  ( $r \leq 1$ ). The envelope of beats is  $k\text{Abs}[\sqrt{1 + [2r/(1 + r^2)]}\text{Cos}(2\pi\Delta ft)]$ , where  $\Delta f$  is the frequency difference between the components and  $k$  is a constant. This, of course, is different from the envelope for sinusoidal AM and so the modulation index for beats is somewhat ambiguous. An alternate definition, one that is perhaps more relevant to perception, is the ratio of the amplitude of the envelope spectral component at  $\Delta f$  to the dc component:  $m' = A(\Delta f)/A(0)$ . Using the Hilbert envelope to extract these components yields the following values for  $(20 \log r, 20 \log m')$ :  $(-20, -20.03)$ ;  $(-10, -10.32)$ ;  $(0, -3.52)$ . For modulation depths less than  $-10$  dB, the differences between these measures are negligible.

Allen, J. B., and Neely, S. T. (1997). "Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wide-band noise," *J. Acoust. Soc. Am.* **102**, 3628–3646.

Bacon, S. P., and Viemeister, N. F. (1985). "Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners," *Audiology* **24**, 117–134.

Baird, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (1984). "Intensity perception. XIII. Perceptual anchor model of context coding," *J. Acoust. Soc. Am.* **76**, 722–731.

Bus, S., and Florentine, M. (1991). "Psychometric functions for level discrimination," *J. Acoust. Soc. Am.* **90**, 1371–1380.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). "Modeling auditory processing of AM. I. Detection and masking with narrowband carriers," *J. Acoust. Soc. Am.* **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). "Modeling auditory processing of AM. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* **102**, 2906–2919.

Edwards, B. W., and Viemeister, N. F. (1994). "Modulation detection and discrimination with three-component signals," *J. Acoust. Soc. Am.* **95**, 2202–2212.

Florentine, M. (1986). "Level discrimination of tones as a function of duration," *J. Acoust. Soc. Am.* **79**, 792–798.

Forrest, T. G., and Green, D. M. (1987). "Detection of partially filled gaps in noise and the temporal modulation transfer function," *J. Acoust. Soc. Am.* **82**, 1933–1943.

Garner, W. R., and Miller, G. A. (1944). "Differential sensitivity to intensity as a function of the duration of the comparison tone," *J. Exp. Psychol.* **34**, 450–463.

Goldstein, J. L. (1967). "Auditory spectral filtering and monaural phase perception," *J. Acoust. Soc. Am.* **41**, 458–479.

Green, D. M., and Nachmias, J. (1979). "Intensity discrimination with gated and continuous sinusoids," *J. Acoust. Soc. Am.* **66**, 1051–1056.

Haftner, E. R., Bonnel, A.-M., and Gallun, E. (1997). "A role for memory in divided attention between two independent stimuli," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 228–237.

Harris, J. D. (1963). "Loudness discrimination," *J. Speech Hear. Disord. Monogr. Suppl.* **11**, 1–59.

Jesteadt, W., Wier, C. C., and Green, D. M. (1977). "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 169–177.

Kohlrausch, A. (1993). "Comment on Temporal modulation transfer functions in patients with cochlear implants [*J. Acoust. Soc. Am.* **91**, 2086–2104 (1992)]," *J. Acoust. Soc. Am.* **93**, 1649–1650.

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

Macmillan, N. A. (1971). "Detection and recognition of increments and decrements in auditory intensity," *Percept. Psychophys.* **10**, 233–238.

Mandel, J. (1984). "Fitting straight lines when both variables are subject to error," *J. Quality Technology* **16**, 1–14.

McGill, W. J., and Goldberg, J. P. (1968). "Pure-tone intensity discrimination and energy detection," *J. Acoust. Soc. Am.* **44**, 576–581.

Moore, B. C. J., and Sek, A. (1992). "Detection of combined frequency and amplitude modulation," *J. Acoust. Soc. Am.* **92**, 3119–3131.

Oxenham, A. N. (1997). "Increment and decrement detection in sinusoids as a measure of temporal resolution," *J. Acoust. Soc. Am.* **102**, 1779–1790.

Plack, C. J., and Carlyon, R. P. (1995). "Loudness perception and intensity coding," in *Hearing*, edited by B. C. J. Moore (Academic, New York), pp. 123–160.

Riesz, R. R. (1928). "Differential sensitivity of the ear for pure tones," *Phys. Rev.* **31**, 867–875.

Sheft, S., and Yost, A. W. (1990). "Temporal integration in amplitude modulation detection," *J. Acoust. Soc. Am.* **88**, 796–805.

Sheft, S., and Yost, A. W. (1996). "AM detection with one- and two-tone modulators," *J. Acoust. Soc. Am.* **99**, 2540(A).

Stiegel, M. S. (1977). "Intensity discrimination with three signal ensembles," unpublished dissertation, University of Pittsburgh.

Strickland, E. A., and Viemeister, N. F. (1996). "Cues for discrimination of envelopes," *J. Acoust. Soc. Am.* **99**, 3638–3646.

Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.

Viemeister, N. F. (1988). "Psychophysical aspects of auditory intensity coding," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 213–241.

Viemeister, N. F., and Bacon, S. P. (1988). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," *J. Acoust. Soc. Am.* **84**, 172–178.

Yost, W. A., and Sheft, S. (1997). "Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions," *Auditory Neuroscience* **3**, 401–414.

Zwicker, E., and Graf, L. (1987). "Modulationsschwellen bei Verdeckung," *Acustica* **64**, 148–154.



# Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees

Astrid van Wieringen<sup>a)</sup> and Jan Wouters<sup>b)</sup>

Lab. Exp. ORL, Katholieke Universiteit Leuven, Kapucijnenvoer 33, 3000 Leuven, Belgium

(Received 19 January 1999; accepted for publication 14 June 1999)

Gap-detection thresholds were determined for different complex patterns of electrical stimulation in four postlingually deafened LAURA cochlear implantees, to examine the nature of within- and across-channel auditory processes in more detail. Gap detectability was examined as a function of stimulus complexity (one, two, or three channels), channel distance within and across multichannel pre- and post-gap markers, stimulus asymmetry, and pulse rate. All markers varied in duration from 200 to 500 ms to ensure that subjects were not using overall stimulus duration as a cue. Gap-detection thresholds for all subjects were short (<5 ms) when the pre- and post-gap markers stimulated the same single or multiple channels, even when the distance between simultaneously stimulated channels was large (exp. 1). For some subjects, gap detectability was more difficult in the across-channel condition, when the pre- and post-gap markers each stimulated different channels, although performance improved substantially in most subjects after extensive training (exp. 2). Additional tests with random maskers also suggest that neural interaction only affects performance mildly, and that the magnitude of the gap-detection threshold probably depends more on the subject's cognitive (in)ability to attend to the temporal gap than on the temporal acuity of their auditory system. Other stimulus conditions showed a difference in performance related to the order of the markers: gap thresholds were longer when the pre-gap marker stimulated one channel and the post-gap marker stimulated two or more channels, than vice versa (exp. 3). In addition, gap thresholds of three of the subjects increased with decreasing pulse rate from 1250 to 400 pps, a finding which may be related to the rate of the speech processing strategies used by each individual implantee (exp. 4). © 1999 Acoustical Society of America. [S0001-4966(99)01010-3]

PACS numbers: 43.66.Mk, 43.66.Ts, 43.64.Me [SPB]

## INTRODUCTION

In order to be able to exploit the present technology of cochlear implant (CI) devices to its full extent, it is necessary to understand the biophysical limitations of electrical stimulation related to the coding of (speech) sounds by the auditory system. This knowledge can aid us in choosing an optimal number of electrode pairs and/or an optimal pulse rate for each individual implantee, and it may even aid us in predicting speech recognition performance. In order to understand the basic perceptual capabilities of the auditory system regarding auditory temporal acuity for electrical stimulation, gap-detection experiments have first been carried out with very simple markers, such as electrical sinusoidal stimuli or pulse trains presented to pairs of electrodes (e.g., Moore and Glasberg, 1988; Preece and Tyler, 1989; Shannon, 1989). This condition is often referred to as a "within-channel" condition, because the two markers, i.e., the stimuli surrounding the gap, both stimulate the same location in the cochlea, and hence cause the same neural population to respond. Under these circumstances, gap-detection thresholds are on the order of a few milliseconds. More recently, however, studies have dealt with more complex patterns of stimulation, to examine how gap detectability is affected by across-channel cues, i.e., when different neural populations

are made to respond (Hanekom and Shannon, 1998). Not only do gap detection thresholds increase markedly in this condition, but the magnitude of these thresholds also depends on the distance between two pairs of electrodes. For the sake of clarity, a pair of electrodes will be referred to here as a channel. As shown by Hanekom and Shannon (1998), average gap-detection thresholds increased by about a factor of 10 as the channel separation between the markers increased. The authors state that their gap-detection data, which are visualized as "tuning curves," can be used as a measure of channel interaction.

Although the within- and across-channel modes of stimulation elicit different gap thresholds, the nature of the mechanisms causing this difference in magnitude remains a matter of debate. While the within-channel task merely requires detection of a discontinuity in the stimulus, the across-channel task requires a relative timing operation between different stimulated neural regions. However, within-channel cues cannot be understood merely in terms of stimulating the same region in the cochlea. Chatterjee *et al.* (1998) presented markers which differed in rate to the same channel. Interestingly, gap thresholds of three Nucleus implantees increased as the difference in rate between the markers increased, despite the same location in the cochlea being stimulated by both markers. Presumably, these data do not reflect the temporal resolution of the auditory system, but rather serve as a measure of the perceived difference between the two markers. Although one subject was capable of attending to the gap

<sup>a)</sup> Author to whom correspondence should be addressed; Electronic mail: Astrid.vanWieringen@uz.kuleuven.ac.be

<sup>b)</sup> Electronic mail: Jan.Wouters@uz.kuleuven.ac.be



itself, the discontinuity in rate strongly dominated gap detection.

Studies with normal-hearing listeners using acoustical stimulation have also shown that gap-detection performance with complex stimuli cannot be predicted from that with simple ones. While minimal detectable gaps are on the order of 5 ms for spectrally similar markers, performance decreases by approximately tenfold as the difference between the center of the sinusoidal markers (e.g., Williams and Perrott, 1972; Formby and Forrest, 1991; Formby *et al.*, 1996) or noise markers (e.g., Formby *et al.*, 1993; Philipps *et al.*, 1997, 1998) becomes larger. In addition, performance also depends on the number and position of the frequency components before and after the gap (Formby *et al.*, 1998). It is suggested that a relative timing operation is performed centrally due to a lack of lateral neural connections between cochlear output fibers in the auditory periphery (Phillips *et al.*, 1997, 1998).

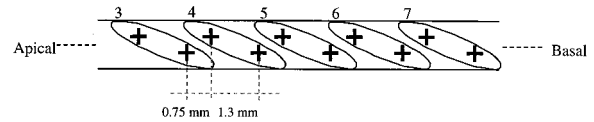
The impoverished gap detectability in the across-channel condition has also been explained in terms of auditory attentional processes (Phillips *et al.*, 1997). Selective attention paid to the first marker reduces the subject's ability to attend to the onset of the second marker, resulting in an imprecise time-stamping of events in the nonattended channels. This suggests that gap thresholds may improve if subjects could be trained to ignore spectral differences between the markers and pay selective attention to the gap only.

The main goal of this study was to explore further the temporal processing abilities of cochlear implantees for different complex patterns of electrical stimulation, in order to learn more about the way temporal speech cues are processed through electrical stimulation. Four experiments were carried out. In experiment 1, gap detectability was examined as a function of stimulus complexity in symmetric single- and multiple-channel pre- and post-gap markers. In experiment 2, gap thresholds were determined for pre- and post-gap markers presented to different channels to gain more insight into the nature of the processing mechanism(s) in the across-channel task. Moreover, a specific aim was to examine whether gap thresholds in this across-channel condition could be improved by training subjects to ignore the perceptual differences between the pre- and post-gap markers on different channels. This may allow us to understand which aspects are limited by peripheral mechanisms and which by central mechanisms. In addition, the extent of neural interaction between adjacent and distant channels was examined in more detail by adding a masker to the stimulus configuration.

In experiment 3, the patterns of stimulation bounding the gap were asymmetric (as in experiment 2), yet partly overlapping. For different stimulus configurations, it was examined how gap detectability was affected by the number of channels in either the pre- or the post-gap marker. These stimulus configurations may be relevant to speech perception performance, particularly with regard to the perception of consonant-vowel (CV) and vowel-consonant (VC) combinations.

Experiment 4 dealt with the effect of pulse rate on gap detectability in within-channel tasks. For acoustical stimulation there is no clear effect of center frequency (Hall *et al.*,

**Type 5.6**



**Type 5.8**

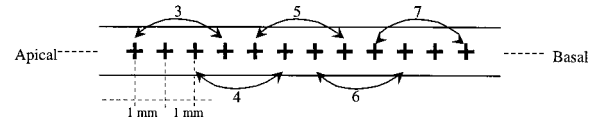


FIG. 1. Schematic representation of the two different electrode arrays of the LAURA implant device (not to scale). The arrays differ with regard to the magnitude of the ball contacts, the positioning of the electrodes on the array, and the combination of the bipolar channels (indicated by the digits). See text for details.

1996). However, for electrical stimulation, sensitivity to pulse rate may be related to the transmission capabilities of the cochlear implant device: speech processing strategies of the Nucleus cochlear implant device stimulate at a much lower rate than those of the LAURA device. In the present experiment this notion was studied by measuring gap-detection thresholds of LAURA implantees at two different rates.

## I. GENERAL METHODS

### A. Electrode designs and electrical stimulation of the LAURA device

The LAURA cochlear implant system was developed by Philips Hearing Implants (formerly Antwerp Bionic Systems), together with the University of Antwerp, in the late 1980's (Peeters *et al.*, 1993). Its internal device consists of an array of 16 electrodes, of which pairs of electrodes are defined as bipolar channels. The overall stimulation rate for the standard biphasic current pulses of 40  $\mu\text{s}/\text{phase}$  is fixed at 10 000 pps. The rate per channel depends on the number of active channels. For this standard setting, a biphasic current pulse is sent to one of the active channels each 100  $\mu\text{s}$ . Hence, the stimulation rate for eight active channels is 1250 pps per channel. Besides the standard pulse width of 40  $\mu\text{s}/\text{phase}$ , biphasic pulses of 100 and 200  $\mu\text{s}/\text{phase}$  can also be used, but then only at lower stimulation rates. For experimental purposes, the current amplitude can vary from 0 to 1500  $\mu\text{A}$  in steps smaller than 1  $\mu\text{A}$ .

Currently, two different types of electrode configurations are in use: type 5.6 and type 5.8 (Fig. 1). On both arrays the electrodes are placed on the most apical 17 of a 25-mm silicon carrier with a slightly curved 3-mm tip. Type 5.8, the most recent design, mainly differs from type 5.6 in the orientation of the electrode contacts within the channels. It was changed from offset-longitudinal (type 5.6) to longitudinal (type 5.8), and the distance between pairs of electrodes, also referred to as channels, was increased from 1.3 to 3 mm. Another difference between the two designs is the diameter of the ball contacts, which is 0.24 and 0.34 mm for

TABLE I. Biographical data of the four LAURA implantees. The table includes their sex (M/F), etiology, approximate duration of profound deafness (>90 dB) in years (D), age at implantation (CI), number of implanted years (# CI), type of internal device (INT=5.6 or 5.8), number of channels active for speech processing (# CH), pulse width in  $\mu\text{s}/\text{phase}$  (PW), vowel (V) and consonant (C) recognition scores in percentage correct. See text for details.

Subject	M/F	Etiology	D	CI	# CI	INT	# CH	PW	V	C
JH	M	meningitis	4	13	4	5.6	8	40	53	45
SG	M	unknown	7	26	4	5.6	8	40	73	58
HS	M	cong. progr.	12	43	2	5.8	8	40	40	35
EC	F	unknown progressive	<2	46	3	5.6	6	100	33	38

types 5.6 and 5.8, respectively. Only one of the four subjects in this study uses type 5.8.

## B. Subjects

Four postlingually deafened subjects, three males and one female, participated in this study. They were implanted at the University ENT Dept. of St. Augustinus Hospital in Antwerp. Details concerning their etiology, duration of profound deafness, age at implantation, number of implanted years, type of internal device, number of channels active for speech processing (out of a total of eight), and speech results are given in Table I. The channels under test in this study are given in Table II. The subjects are considered moderate to good performers. Subjects JH and SG were among the best performers in a phoneme recognition study of 25 implantees with the LAURA implant and LAURA flex speech processor, while subjects HS and EC performed slightly below average in the vowel recognition test ( $X=42\%$ ,  $SD=19\%$ )

and slightly above average in the consonant recognition test ( $X=33\%$ ,  $SD=13\%$ ; van Wieringen and Wouters, 1999).

The number of tested channels varied per subject, depending on their availability. Subjects SG and JH listened to many channel combinations, subjects EC and HS to somewhat fewer. However, in the multiple-channel conditions, each subject detected gaps in markers containing both adjacent and distant channels (Table II).

Before data collection, each subject practiced gap detection for each different stimulus condition until performance did not improve anymore. The exact amount of practice varied from experiment to experiment. Moreover, subjects SG and JH had experience with other psychophysical experiments and needed less practice than subjects HS and EC. Subjects were tested once every one or two weeks at their homes for about 2–2 1/2 hrs (=1 testing session), and 8–10 testing sessions were required of each subject to complete the study. They were not paid a wage, but they received a

TABLE II. Selected channels with 1 = most apical, and 8 = most basal. The stimuli are represented as follows: the gap is indicated with “–” and the numbers before and after “–” indicate the channel(s) under test. Continuous stimulation (condition 1d) is indicated between brackets and the masker channels are printed in bold.

	JH	Channel number	SG	Channel number
Exp. 1	1a	2-2, 4-4, 5-5, 6-6, 7-7	1a	2-2, 3-3, 4-4, 5-5, 6-6
	1b	67-67, 27-27	1b	23-23, 28-28, 36-36
	1c	246-246, 267-267	1c	246-246
	1d	4-4(26)	1d	4-4(26)
Exp. 2	2a	2-4, 2-6, 2-7, 6-7	2a	2-3, 2-4, 2-6, 3-4, 3-5, 3-8
	2b	2-6 (4), 2-7 (4), 6-7 (4)	2b	2-4 (6), 2-6 (4), 3-5 (4), 3-5 (6), 3-8 (5)
Exp. 3	3a	26-2, 67-6	3a	34-3, 35-3, 38-3
	3b	2-26, 6-67	3b	3-34, 3-35, 3-38
	3c	246-4	3c	246-4, 356-5
	3d	4-246	3d	4-246, 5-356
Exp. 4	4a	2-2, 4-4, 6-6, 7-7	4a	2-2, 3-3, 4-4, 5-5
	HS		EC	
Exp. 1	1a	2-2, 5-5, 6-6, 7-7	1a	2-2, 4-4, 5-5, 6-6
	1b	27-27, 67-67	1b	26-26, 56-56
	1c	246-246	1c	246-246
	1d	4-4(26)	1d	4-4(26)
Exp. 2	2a	2-6, 6-7	2a	2-4, 2-6, 4-5
	2b	2-6 (4), 6-7 (4)	2b	
Exp. 3	3a	26-2, 67-6	3a	24-2, 25-2, 26-2, 45-4, 56-5
	3b	2-26, 6-67	3b	2-24, 2-25, 2-26, 4-45, 5-56
	3c	246-4, 267-6	3c	246-4
	3d	4-246, 6-267	3d	4-246
Exp. 4	4a	2-2, 5-5, 6-6, 7-7	4a	2-2, 5-5, 6-6

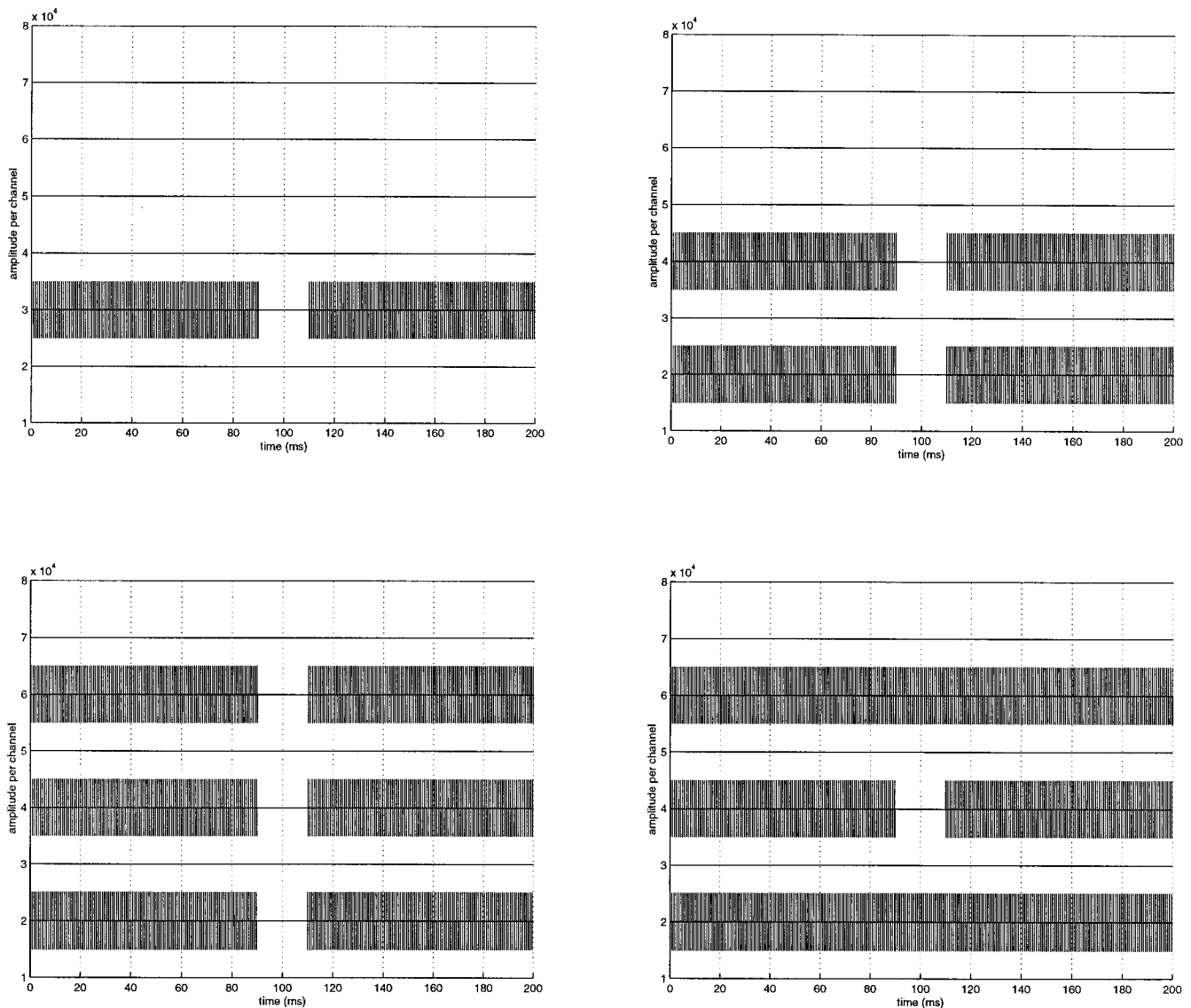


FIG. 2. Schematic representation of the stimuli in experiment 1. In the figure, and as an example, using our definition of stimulus configuration, the single-channel configuration (1a, top left) is denoted as 3-3, the two-channel configuration (1b, top right) is denoted as 24-24, the three-channel configuration with the gap in all channels (1c, bottom left) is denoted as 246-246, and the three-channel configuration with gap in one channel (1d, bottom right) is denoted as 4-4(26). The duration of the markers is roved (not shown).

gift for participating. The different stimulus conditions of each experiment were always tested together in the same session. As a result, differences in performance within an experiment cannot be attributed to different testing moments.

### C. Stimuli

The stimuli were biphasic electrical pulse trains with 40  $\mu$ s/phase pulse widths for subjects SG, JH, and HS, and 100  $\mu$ s/phase pulse widths for subject EC. The standard pulse rate was 1250 pps (800  $\mu$ s interpulse interval), but for some conditions for subject EC (1c and 1d) the pulse rate was decreased to 1111 pps (900  $\mu$ s), in order to be able to stimulate three channels quasi-simultaneously with 100  $\mu$ s/phase stimuli.

The duration of each of the markers varied randomly between 200 and 500 ms to prevent subjects from using overall duration as a cue. The step size was a multiple of 40 or 60 ms. The initial gap was 140 ms, after which the step

size was decreased by a factor of (approximately) 1.2. This factor could not always be exactly 1.2, because the duration of the gap between markers had to be a multiple of the interpulse interval (800 or 900  $\mu$ s), especially in condition 1d (Fig. 2). As a result, the shortest possible gap was 0.8 ms for 1250 pps stimuli, and 0.9 ms for 1111 pps stimuli. All channels were stimulated at the same rate. However, each different channel was delayed by 100  $\mu$ s, as simultaneous stimulation of two or more channels is not possible. All stimuli were created in MATLAB 5.1, checked visually on a digital scope, and stored on the hard disc of a portable computer. For each stimulus condition, two series of stimuli with different roving durations of the markers were created. Also, three different reference stimuli without a gap were created for each stimulus condition. These stimuli, which were 400, 740, and 1200 ms in duration, were presented randomly during stimulus presentation. Although many different stimuli were created, the same 40  $\mu$ s/phase stimuli could be used for

subjects SG, JH, and EC, because the selection of the channels and their current levels were controlled by the software which transmitted the pulse trains to the internal device of the cochlear implant (see procedure).

#### D. Procedure

Stimulus presentation and response collection were controlled with a software interface which was specifically designed at our lab to send the MATLAB 5.1 generated stimuli to the internal part of the LAURA device by means of a DSP-TIC30 board on a 486DX4 portable personal computer. Prior to gap detection, threshold and most comfortable levels (MCL) were determined for each of the subject's channels, and the different stimulus configurations of the different stimulus conditions were balanced in loudness (see Sec. 1E). The threshold is the lowest level at which a sound is detected, the MCL is the level at which the stimulus sounds "normal."

Subsequently, gap thresholds were determined using an adaptive, two-interval forced-choice procedure, in which one of the two intervals contained a gap between the markers. The interstimulus interval was 500 ms. The observation intervals were indicated by two blocks on a computer screen; the "signal" interval was chosen by a mouse click. Gap duration was varied adaptively via a transformed two-down, one-up procedure (Levitt, 1971). The first presentation contained the stimulus with the 140-ms gap. During the first four reversals, two stimuli were skipped after two correct responses or one incorrect response, so that gap was increased or decreased by a factor of 1.73 instead of 1.2. During the remaining eight reversals the gap was varied by a factor of 1.2. Visual feedback was provided after each response. Each threshold run was stopped after 12 reversals, and the mean of the last 6 reversals was taken as the estimate of the minimal detectable gap. After training, each gap-detection threshold was based on at least three estimates. Data collection was continued, however, if performance seemed to improve (see experiment 2).

#### E. Loudness balancing

Prior to gap detection, all the stimuli were balanced in loudness by means of an adaptive procedure. This meant that the levels of each channel were adjusted in order to present the stimulus at MCL. Note that each stimulus consisted of two markers, a pre-gap and post-gap one, and that each of these markers could stimulate not only different channels, but also a different number of channels. For the multiple-channel stimuli, the levels of two or more channels had to be adjusted simultaneously. While the threshold remained unchanged, the level of each channel was adjusted by a percent of its dynamic range.

During the loudness balancing procedure, three 500-ms stimuli were presented continuously to the subject, using an automatic procedure. The first and third presentations were the reference signal, for instance a one-channel stimulus configuration (condition 1a, Fig. 2). The second stimulus, the variable one, was a different stimulus, for example a two-channel stimulus configuration (condition 1b, Fig. 2). By

pressing +, ++, or +++ (or -, --, or ---) with a mouse button, the subject could match the loudness of the variable stimulus to the reference signal by making it a little bit or a great deal louder (or softer). After pressing "OK," which indicated that the three stimuli sounded equally loud, the next triad was presented. The subject always matched each variable stimulus twice with the reference one, once with the starting value of the MCL of each channel of the variable stimulus being 100  $\mu$ A lower than that of the reference signal, once with it being 100  $\mu$ A higher than that of the reference signal. The mean of the two adjustments was taken as the MCL of that channel.

Although the different experiments required many different stimulus configurations to be balanced in loudness, the procedure was easily understood by the subject and yielded stable values. Loudness balancing was always done prior to gap detection in new experimental conditions, and not repeated during the course of the sessions. However, subjects were repeatedly asked whether the presentations still sounded "normal." If not, loudness balancing was repeated for the stimulus conditions under test.

## II. EXP. 1: EFFECT OF STIMULUS COMPLEXITY

The purpose of this experiment was to examine how stimulus complexity affected temporal resolution in four LAURA implantees. Gap thresholds of 1250-pps pulse trains were determined for one, two, and three symmetric channels stimulated quasi-simultaneously by both the pre- and post-gap markers. Moreover, in one stimulus configuration, the silent interval occurred only in one of the three channels.

#### A. Method

Figure 2 illustrates the different stimulus conditions of experiment 1. Gap-detection thresholds were measured in conditions where the pre- and post-gap markers were presented to the same channel (1a), to the same two channels (1b), and to the same three channels (conditions 1c and 1d). In conditions 1b and 1c, the gap is present in all two or three channels, respectively, while it is only presented in one channel in condition 1d. Acoustical experiments with narrow bands of noise have shown that gap detection can be disrupted by the presence of a synchronous stimulus in another channel that does not contain the gap (Grose and Hall, 1993). In experiment 1 very little practice was required, about 5–8 experimental runs to reach asymptotic performance.

For this experiment, condition 1a was matched in loudness with condition 1b, condition 1b was matched in loudness with condition 1c, and condition 1c with 1d. This was done for each combination of channels used in the experiments.

#### B. Results experiment 1

Mean gap thresholds and standard deviations of each of the four different stimulus configurations (1a, 1b, 1c, 1d) are plotted for each subject and each (combination of) channel(s) separately in Fig. 3(a)–(d). For single-channel stimulation, gap thresholds vary between 1 and 2.5 ms for SG and JH, and between 2 and 4.5 ms for subjects HS and EC. These



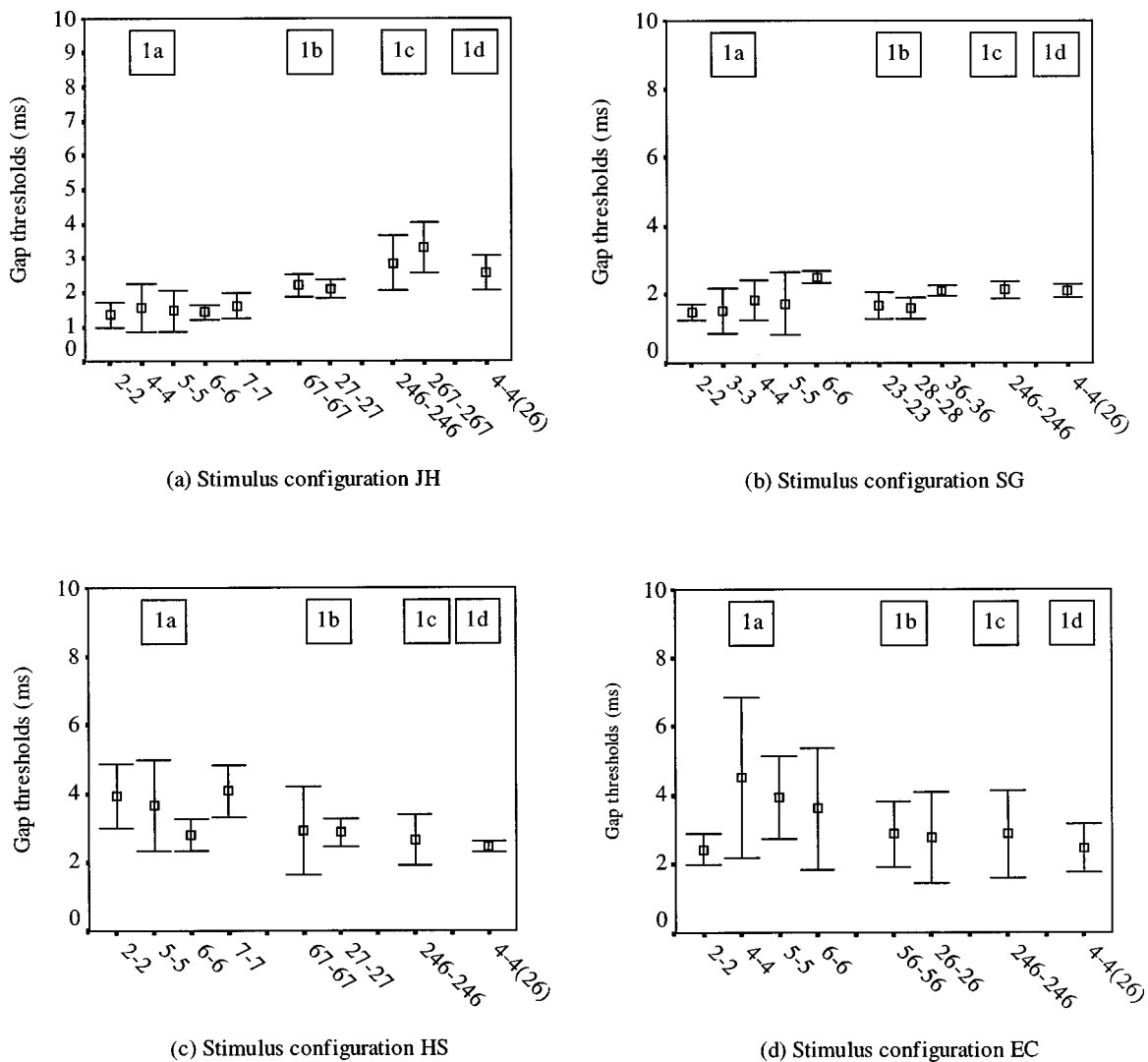


FIG. 3. Mean and standard deviations of the gap thresholds for one-channel (1a), two-channel (1b), and three-channel (1c, 1d) stimulus configurations in experiment 1. In the three-channel configurations, the gap either occurs on all three channels (1c) or on one of the three channels (1d), here channel 4. The channels under tests are indicated on the abscissa.

data are consistent with the range of thresholds reported by Shannon (1989), and Hanekom and Shannon (1998), for electrical stimulation.

Gap-detection thresholds are approximately equally short in more speechlike conditions, when the same two or three channels are stimulated quasi-simultaneously by both markers (1b, 1c), or when the gap occurs on only one of the three channels (1d). Note that gap detection in condition 1d was not disrupted by the presence of other continuous pulse trains. Apparently, all the within-channel stimulus conditions in experiment 1 merely required a process of discontinuity detection.

In addition, gap-detection thresholds depend neither on the position of the channel(s), nor on the distance between channels. Previously, some acoustical studies have shown that gap-detection performance improves when the gap information is carried by more than one perceptual channel (Green and Forrest, 1989; Grose, 1991; Hall *et al.*, 1996). Note, however, that the pressure spectrum level or sensation level of the individual channels is held constant in the acoustical studies, whereas the intensity of each channel declines

as the number of channels increases in the present electrical experiments. Gap-detection performance is comparable for the different within-channel stimuli in experiment 1, because they all sound equally loud, irrespective of the number of channels under test.

A three-way analysis of variance (general linear model, SPSS 7.5, 1997) with “subjects,” “stimulus configuration,” and “position of channel(s)” as fixed effects, yielded “subjects” as a significant main effect ( $p < 0.0001$ ). *Post hoc*, the Tukey honestly significant difference (HSD) method showed this effect to be related to the difference in performance between subjects JH, SG and subjects HS, EC. This difference is also evident in Fig. 3(a)–(d). Neither “stimulus configuration” nor “position of channel(s)” were statistically significant and there were no significant interactions.

### III. EXP. 2: ACROSS-CHANNEL PROCESSING AND EFFECT OF TRAINING

The aim of the second experiment was to examine the nature of across-channel cues in more detail for identical pre-

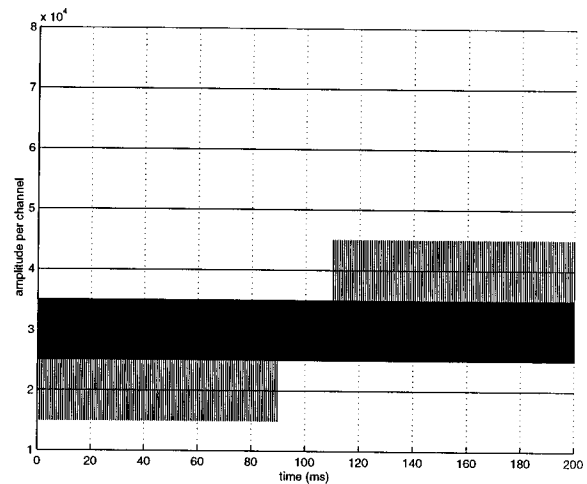
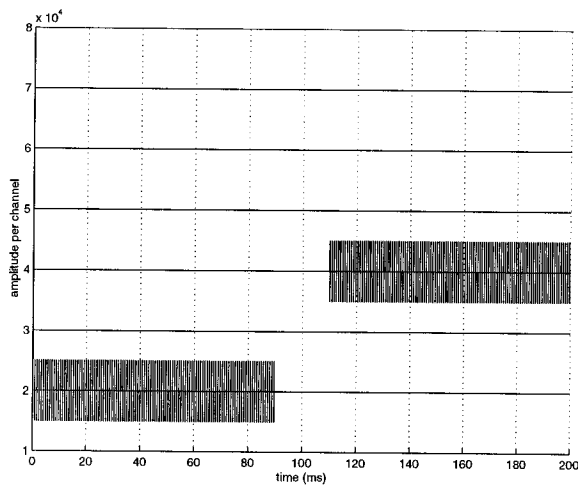


FIG. 4. Schematic representation of the stimuli in experiment 2. Following our definition of stimulus configuration, the two-channel configuration (2a, left) is denoted as 2-4, and the two-channel configuration with masker (2b, right) is denoted as 2-4(3). The duration of the markers is roved (not shown).

and post-gap markers presented to different channels. Previously, it has been shown that gap-detection thresholds increase significantly when two different channels are stimulated, and that thresholds increase with increasing channel distance (Hanekom and Shannon, 1998). These data were interpreted in terms of differences in neural interaction: the smaller the distance between channels, the larger their neural interaction, and hence, the shorter the gap-detection threshold. However, a similar pattern of performance was also found when markers differing in rate were presented to the same channels (Chatterjee *et al.*, 1998). If gap-detection performance merely reflected the effect of neural interaction, gap thresholds on identical channels should always be short, irrespective of the difference in pulse rate. Chatterjee *et al.*'s data, however, suggest that a different mechanism is involved when pulse rates differ, and that this mechanism does not reflect temporal acuity of the auditory system.

In experiment 2, gap thresholds are determined for pre- and post-gap markers presented to different channels. Gap detectability was not examined systematically as a function of channel separation, as Hanekom and Shannon (1998) had already determined threshold functions, i.e., V-shaped "tuning curves" for many combinations of channels in four implantees using the Nucleus cochlear implant device. More important to us was whether the increasing thresholds with increasing channel distance would decrease by (1) training subjects to attend to the silent interval, and (2) by adding a random masker to the stimulus configuration. If, on the one hand, performance in the across-channel task is impoverished as a result of the task being cognitively (too) demanding, gap thresholds should improve after extra training. If, on the other hand, the nature of the underlying mechanism is peripheral, inserting a random masker between adjacent channels should reduce the amount of neural interaction, resulting in longer gap-detection thresholds.

## A. Method

### 1. Subjects

Due to technical constraints, subject EC did not participate in those experimental conditions where a masker was

added to the stimulus configuration. For 100  $\mu$ s/phase pulses the total rate is limited to 3333 pps. Technically, our system was not capable of transmitting 100  $\mu$ s/phase pulse trains with random pulse intervals and 100  $\mu$ s/phase pulse trains with fixed interpulse intervals at the same time.

Subjects SG and JH listened to several different combinations of channels (Table II and also illustrated in Fig. 4), while HS and EC listened to a limited number of stimulus configurations of adjacent or distant channels. For this experiment, subjects were repeatedly asked to attend to the gap only and to ignore other perceptual differences. This experiment was started together with experiment 1, and, as in all experiments, all subjects were trained prior to testing until performance did not improve anymore. Gap-detection thresholds were collected after approximately 20 practice runs, but training was pursued at the beginning of each testing session. At least six to eight estimates were collected for each stimulus configuration, because some of the data displayed large standard deviations and also because performance of two subjects suddenly improved substantially during the course of the testing sessions.

### 2. Stimuli

The stimuli were generated and presented in the same way as in the previous experiment. In experiment 2, the pre- and post-gap markers were presented to two different channels, either with or without a masker (Fig. 4). The pulse rate of the stimuli was 1250 pps, that of the masker varied. Three different maskers, A, B, and C, were included in our study, as it was unclear whether they would produce similar masking effects. The amplitude of all three maskers was constant. The average interpulse time of both masker A and masker B was 0.5 ms (2000 pps), but the distribution of the pulses was uniform for masker A, while the chance of a pulse every 100  $\mu$ s was 0.2 for masker B. For masker C, the pulse rate was 2162 pps, which is periodic, but not related to the standard pulse rate (so subjects could not hear patterns in them). Maskers were generated in MATLAB 5.1 and presented continuously during stimulus presentation to the three subjects with the 40  $\mu$ s/phase stimuli.

### 3. Loudness balancing

Subjects first matched the no-masker stimuli (2a) in this experiment with the symmetric single-channel stimulus configurations (1a). Very often, the across-channel stimuli sounded only slightly louder than these single-channel stimuli. After this adjustment, the across-channel stimuli were matched with the balanced two-channel stimulus configurations (1b) used in experiment 1. Subsequently, the no-masker stimuli were balanced in loudness to the stimuli with a masker (2b). This was done for each of the different combinations of channels (Table II).

## B. Results experiment 2

Figure 5(a)–(d) show mean and standard deviations of gap thresholds for asymmetric stimulus configurations with and without a masker. To illustrate the effect of additional training, gap thresholds were divided into two equal groups. The filled symbols illustrate the means of the first three or four estimates, the unfilled symbols those of the last three or four estimates. The data display considerable differences in performance between subjects in this across-channel task, and these will be discussed separately for the no-masker and masker conditions in the following paragraphs.

### 1. No masker and effect of training

Compared to the data discussed in experiment 1, gap-detection thresholds of asymmetric stimuli without a masker are longer and less consistent (large standard deviations, condition 2a in Fig. 5). In general, our results are similar to those of Hanekom and Shannon (1998), although their data are plotted on a log scale and include the data of the within-channel stimulus conditions. However, the variation in performance within and across subjects illustrates that it is difficult but not impossible to attend selectively to the temporal gap in the presence of other confounding cues. As a result, not all the data show an effect of channel distance: gap-detection thresholds of subject JH were relatively short from the beginning (after prior training), i.e., on average, 3 ms for stimuli presented to adjacent pairs of electrodes and about 8 ms for larger separations. Separate two-way analyses of variance showed no significant effect of channel distance, either for the first half (filled symbols), or for the second half (unfilled symbols) of the data. The results are, however, different for the other three subjects, who had more difficulty detecting the temporal cue. Statistically, the first-order interaction between “channel distance” and “training effect” was significant for SG ( $p < 0.001$ ), and HS ( $p < 0.0001$ ), and this effect of training is visible in Fig. 5(b) (SG) and Fig. 5(c) (HS). After three to four estimates, SG’s thresholds decreased from about 15 to 4 ms, and the effect of channel distance, which was significant in the first half of the data, disappeared. A similar effect was encountered for subject HS, although the magnitude of this threshold improvement was from 50 to 15 ms. It appears that once subjects are capable of detecting the temporal gap in these conditions, the effect of channel distance largely disappears. For some subjects, such as subject HS, it did not disappear completely: his gap-detection thresholds in the within-channel conditions (exp. 1) are still significantly shorter than those in the across-

channel conditions. Still, the effect of channel distance is much more pronounced for subject EC, who was not capable of attending selectively to the temporal gap in the presence of other confounding cues, despite continued training. Statistically, her data showed no significant main effects or interactions.

The effect of training deserves some more attention, because subjects were trained with all the different stimulus conditions in this study prior to data collection. The results of this experiment, however, confirm previous suggestions that gap detection under these conditions is not a measure of temporal resolution, but rather one of the perceived difference between the pre- and post-gap marker (Chatterjee *et al.*, 1998). Even without a (physical) gap, two markers stimulating different channels most likely elicit a “gap percept” as a result of the pitch jump. Hence, despite individual differences, the magnitude of the gap thresholds probably depends more on the subjects’ cognitive (in)ability to attend to the temporal gap than on the temporal acuity of their auditory system. If this is true, the V-shaped “tuning curves” of Hanekom and Shannon (1998), indicating neural interaction, should be interpreted with some caution. Our data suggest that neural interaction or overlap only affects performance mildly, and that the difference in gap threshold magnitude between adjacent and distant electrode pairs should, to a large part, be ascribed to central mechanisms. In our opinion, the differences in performance between the two studies cannot be explained by the different electrode configurations of the Nucleus and LAURA cochlear implant devices. Our electrode configuration was comparable with the BP+2 configuration of the Nucleus device. Moreover, the data of Hanekom and Shannon (1998) did not show a clear relation between electrode configuration and gap detectability: gap-detection thresholds varied more across subjects than across the four different stimulation modes (BP+1, BP+2, BP+2 and pseudo-monopolar).

### 2. Stimulus configuration with a masker

A masker was added to the stimulus configuration to determine whether thresholds would increase as a result of a reduction of neural interaction (condition 2b). These stimulus configurations were presented to the subjects while they were completing the no-masker sessions, so they had already received extensive training in the across-channel task. As a result, only the second half of the no-masker data were compared with the masker data in the statistical analyses. Gap thresholds of the stimulus configurations with masker (2b) are also illustrated in Fig. 5 for three subjects (JH, SG, HS). The position of the masker is indicated separately between brackets. Gap thresholds were averaged over the three different maskers, because a separate three-way analysis of variance (subjects, masker type, and channel distance) containing only the data with the different maskers, showed no effect of masker type. As a result, each value is based on at least nine estimates.

For subjects JH and SG, the masker had no effect on the data. However, for subject HS, masking significantly elevated the gap thresholds ( $p < 0.0001$ ), although perhaps more as a result of increasing difficulty in performing the

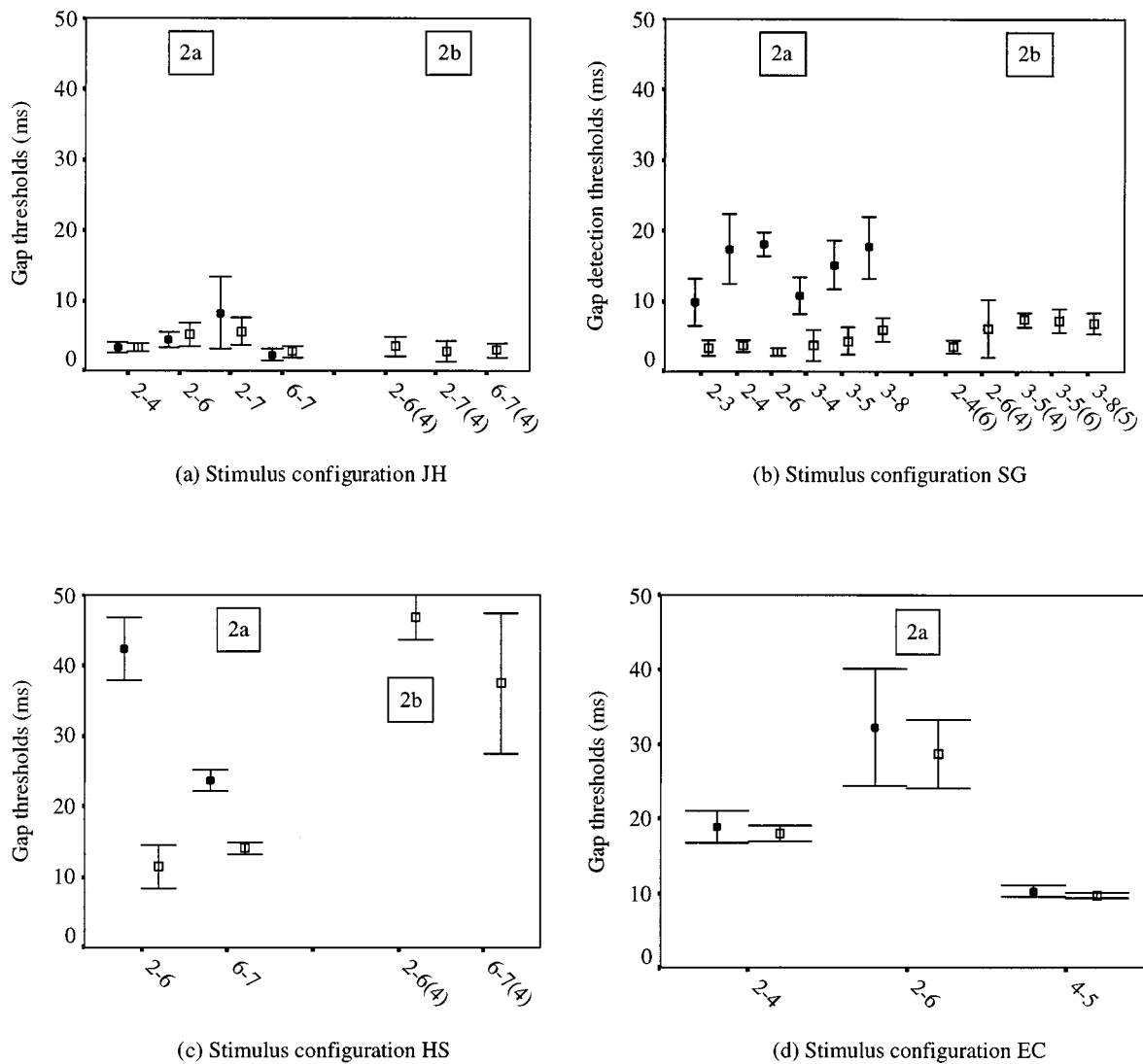


FIG. 5. Mean and standard deviations of the gap thresholds for the no-masker (2a) and masker (2b) conditions (experiment 2). The data of the no-masker stimulus configurations are divided into two equal groups. Filled symbols refer to the first three or four estimates, unfilled symbols to the last three or four estimates. On the abscissa it is indicated which channel is stimulated by the pre- and post-gap markers. The position of the masker in condition 2b is indicated between brackets.

task (the masker acting as a distracter), than from decreased neural interaction. Maskers were presented in different stimulus configurations, either between two stimulating channels or on lower or higher channels. This was done to examine whether the position of the masker affected the amount of neural interaction. Statistical analyses showed no effect related to the position of the masker. Thus, as far as our data are concerned, masking hardly affected performance of subjects who were capable of attending to the gap despite other confounding cues, and it did not reduce neural interaction effects between adjacent channels. However, masking did seem to complicate the task for the subject who was apparently easily confused by the perceptual differences between the pre- and post-gap markers. None of the data of the three subjects in the masking conditions displayed a statistically significant effect of channel separation.

#### IV. EXP. 3: EFFECT OF ASYMMETRIC CONFIGURATIONS ON PERCEPTION

In experiment 1, the gap always occurred on the same channels in the two markers; in experiment 2 it was divided

over different channels. In this experiment, the pre- and post-gap markers carried the silent interval on the same channel, as in experiment 1, but one of the two markers also stimulated one or two other channel(s). It is investigated whether the gap, as a “within-channel” cue, facilitates gap detectability in stimulus configurations where other perceptual channels are also stimulated and whether gap detectability is affected by the order of the markers.

#### A. Method

##### 1. Stimuli

Two different asymmetric stimulus configurations were tested, as shown schematically in Fig. 6. In one, the pre-gap marker stimulated two (3a) or three (3c) channels, after which the post-gap marker stimulated one of those channels. In the other, the markers were reversed (conditions 3b and 3d respectively). In conditions 3a and 3b, the single channel in the pre- or post-gap marker was always the most apical of the two, while the single channel in conditions 3c and 3d was always the middle one (usually channel 4).



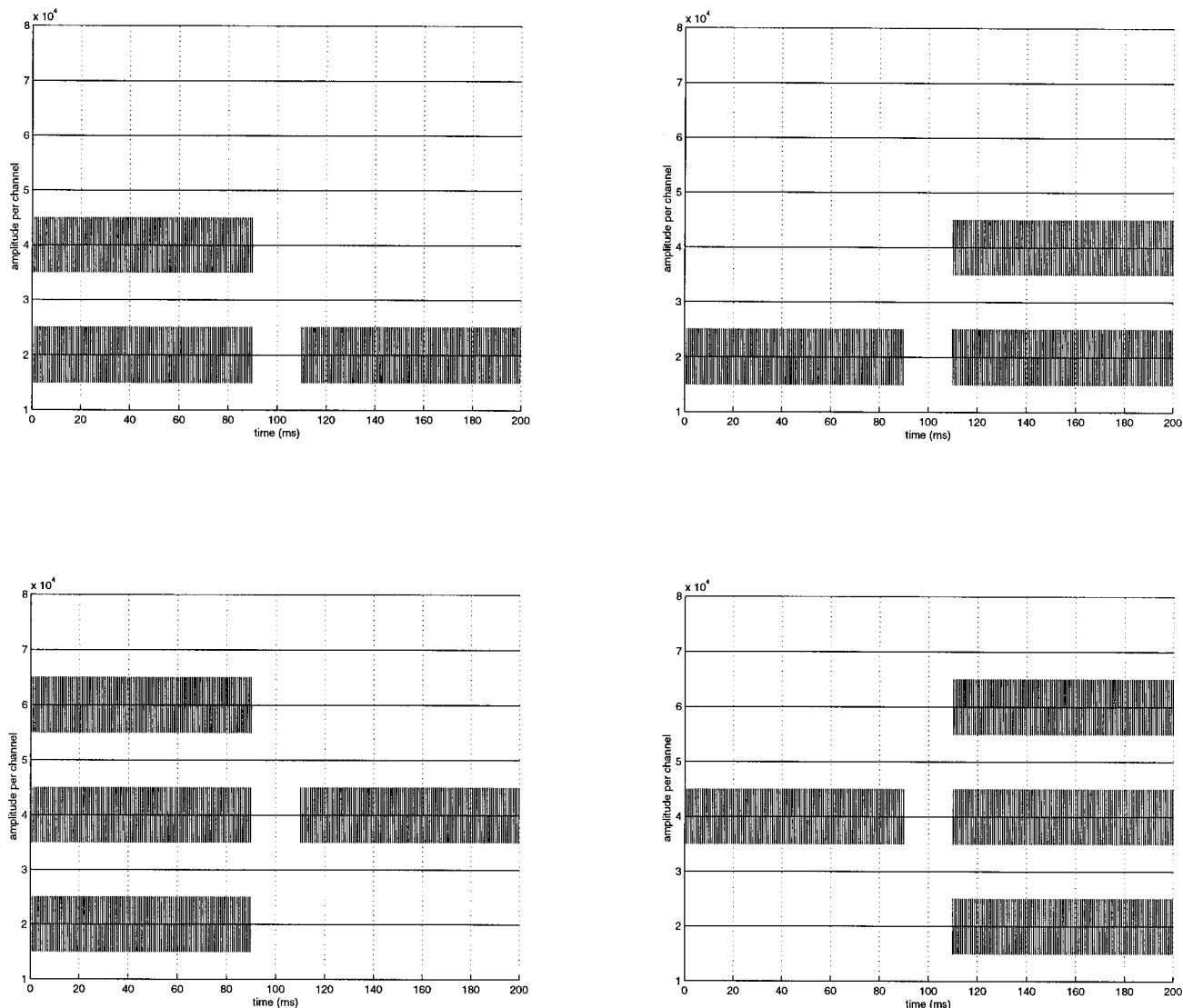


FIG. 6. Schematic representation of the stimuli in experiment 3. Following our definition of stimulus configuration, the two-channel configurations are denoted as 24-2 (3a, top left) and 2-24 (3b, top right), respectively, and the three-channel configurations are denoted as 246-4 (3c, bottom left) and 4-246 (3d, bottom right), respectively. The duration of the markers is roved (not shown).

## 2. Subjects and loudness balancing

All four subjects participated in this experiment. For each of these stimulus configurations, gap detectability is measured both for adjacent and distant channels (Table II). Prior to testing, subjects adjusted the loudness of the two-channel stimulus configurations (3a, 3b) to that of stimulus configuration 1b (Fig. 2), and that of the three-channel stimulus configurations (3c, 3d) to that of 1c (Fig. 2). As this experiment covered several testing sessions, care was taken that the same stimulus configurations of conditions a and b, and c and d, were always tested the same session. These stimulus configurations were administered to the subjects from the beginning of the study, and did not require much training, about 5–10 experimental runs. Each value is based on at least three estimates.

## B. Results experiment 3

Figure 7(a)–(d) show the means and standard deviations of stimulus configurations 3a, 3b, 3c, and 3d for each of the

four subjects separately. Compared to the data of the previous experiment, subjects JH and SG performed similarly ( $<10$  ms) in the two-channel stimulus conditions, and subjects HS and EC performed somewhat better, especially when two channels were stimulated in the pre-gap marker.

The most surprising finding, however, was the difference in performance related to the order of the markers: gap detection was increasingly more difficult when the pre-gap marker stimulated one channel and the post-gap marker stimulated two or three channels than vice versa. The statistical significance of the factors “subjects”, “order”, and “channel distance” were analyzed for the two-channel and three-channel configurations in two separate three-way analyses of variance. The dependence of performance on channel order was evident for both the two- and the three-channel configurations (3a vs 3b and 3c vs 3d); it was significantly different for all subjects ( $p < 0.001$ ) except JH. Although the statistical analyses showed the factors “subjects” and “order” to be significant main effects, “channel

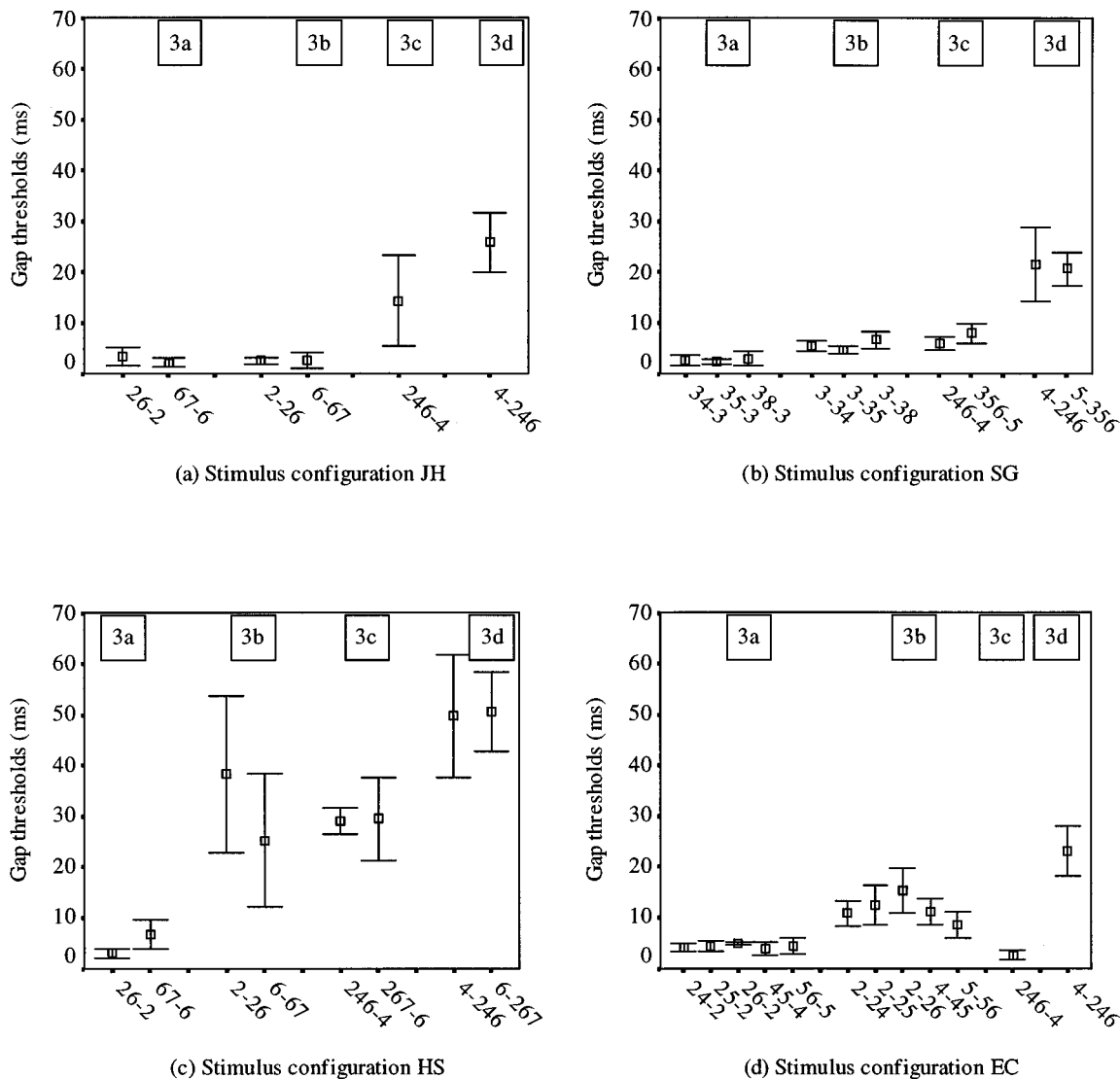


FIG. 7. Mean and standard deviations of the gap thresholds for two-channel (3a, 3b) and three-channel (3c, 3d) stimulus configurations. The abscissa indicates which channels are stimulated by the pre- and post-gap markers.

distance'' within the pre- or post-gap marker [see Fig. 7(b) and (c)] was not significant.

Recently, Formby *et al.* (1998) also reported that gap-detection thresholds of normal-hearing subjects can change significantly by interchanging the temporal position of one of the frequency components before and after the silent interval. For instance, in stimulus configurations where the frequencies of the pre- and post-gap markers F1 and F2 were fixed at 2000 Hz, gap-detection thresholds increased markedly when a remote variable F3 component, ranging from 2000 to 3100 Hz, was added to the post-gap marker. In our study, the magnitude of the threshold did not depend on the position of the electrode, but gap detection clearly became increasingly more difficult when the post-gap marker stimulated more channels than the pre-gap marker. A plausible inference could be the difference in loudness between the pre- and post-gap markers, because the amplitude of the single-channel (usually channel 4 in these experiments) in the pre-(3d) or post-gap (3c) markers was the same as that of the (same) channel in the multiple-channel configuration. In a pilot experiment, this notion was studied in more detail with

two subjects (SG and JH), by enhancing the amplitude of the single-channel marker while the amplitudes in the three-channel marker remained unchanged. As gap detection also depends on the level of the markers (Moore and Glasberg, 1988; Shannon, 1989; Hanekom and Shannon, 1998), performance improved, both for conditions 3c and 3d. However, the perceptual asymmetry remained, even when the amplitude of the single-channel marker was raised by 50%. At that point, gap thresholds of both SG and JH were shorter than 1 ms for condition 3c, and at least 4 ms for condition 3d.

It is also possible that the later occurring information, i.e., the post-gap marker, partly masks detection of the preceding input if the post-gap marker stimulates more channels than the pre-gap marker. The origin of this effect may be cognitive, and not necessarily related to physiological mechanisms. Previously, acoustical studies dealing with intensity discrimination have suggested that the masker (here, the post-gap marker) may interfere with the memory trace, making it difficult for subjects to detect the variable under test (e.g., Plack and Viemeister, 1992; Plack *et al.*, 1995). It is also possible that performance is affected by the marked

change in timbre caused by a sudden onset of additional channels in the post-gap marker.

## V. EXP. 4: EFFECT OF RATE

Our last experiment dealt with the effect of a different pulse rate on gap detectability. One would expect gap thresholds to increase with decreasing rate as a result of the longer time interval between pulses and, hence, increased roughness of the signal. Very recently, Busby and Clark (1999) have shown that pulse rate has no influence on gap detectability with the Nucleus cochlear implant device. Similar findings were reported by Preece and Tyler (1989), but data of Dobie and Dillier (1985) do show that thresholds increase with decreasing pulse rate. It is possible that a rate effect is related to the speech processing strategies of the different cochlear implant devices. As LAURA implantees use speech processing strategies with faster rates than users of the Nucleus cochlear implant device, we studied possible differences in sensitivity for 1250 and 400 pps (which is close to the rate of the Nucleus cochlear implant device).

### A. Method

#### 1. Stimuli

Gap thresholds were determined for symmetric, single-channel markers (cf. experiment 1a) which had a pulse rate of 400 pps. Two sets of stimuli were generated, one for each of the different pulse widths (40 and 100  $\mu$ s/phase). As with the 1250 pps stimuli, the duration of the markers roved from interval to interval and the gap size was always a multiple of the pulse rate. However, as the shortest gap, i.e., 2.5 ms, was considered too large to measure gap-detection performance accurately, its size was reduced to 1 ms.

#### 2. Subjects and loudness balancing

All four subjects participated in this experiment. Prior to testing, most comfortable levels of each of the channels under test were determined. No additional loudness balancing was necessary. The 400-pps stimuli were administered in the same sessions as the 1250-pps single-channel stimuli (experiment 1a).

### B. Results experiment 4

Average gap thresholds and standard deviations of the single-channel 400-pps stimuli are plotted for each subject in Fig. 8(a)–(d), together with those of the 1250-pps stimuli of experiment 1a [Fig. 3(a)–(d)]. The figures show that lowering the rate from 1250 to 400 pps resulted in longer gap thresholds for subjects JH, SG, and HS, but not for subject EC. A three-way analysis of variance (general linear model) showed that “subjects” and “rate” were statistically significant main effects ( $p < 0.0001$ ). The position of the channel, the third factor, had no effect on the data (cf. experiment 1). *Post hoc* Tukey HSD tests on the “subjects” factor yielded significant differences in performance for subjects JH, SG, and HS, but not for subject EC.

The difference in performance may be related to the rate as used daily with the speech processing strategy of each implantee, which, in turn, is related to the pulse width and

the number of active channels in the LAURA device. In daily communication, subject EC is accustomed to much lower rates ( $3333/6 = 556$  pps per channel) than JH, SG, and HS (1250 pps per channel), and this may explain why her gap thresholds are not affected by a change in rate to 400 pps.

## VI. DISCUSSION

### A. Stimulus complexity

The foregoing experiments were carried out to examine which factors influence gap-detection performance, a frequently used measure of auditory temporal resolution. These factors were brought about by different complex patterns of stimulation. First, our study has shown that stimulus complexity does not necessarily affect gap detectability. Our four trained listeners performed equally well when the pre- and post-gap markers stimulated the same single or multiple channels. Hence, these measurements are consistent with the view that cochlear impairment does not have to affect the temporal resolution of the auditory system (Moore and Glasberg, 1988; Shannon, 1989).

### B. Electrode distance and neural interaction

Second, our data also illustrate that gap thresholds of some subjects deteriorate when the pre- and post-gap markers each stimulate different channels. Moreover, gap-detection thresholds increased with increasing distance between channels. This is in agreement with previous studies using both acoustical (e.g., Philips *et al.*, 1997, 1998) and electrical stimulation (Hanekom and Shannon, 1998), but the origin of this effect was a matter of debate. The results of Hanekom and Shannon (1998) suggest that the origin is in the amount of overlap in the neural populations stimulated by two channels, although the authors also argue that this is not the only factor that influences performance. Indeed, several factors suggest that the amount of overlap in neural populations may only mildly affect performance.

First, Chatterjee *et al.* (1998) demonstrate a similar effect for stimuli of unequal rate (or amplitude) presented to the same channels. Although the pre- and post-gap markers elicit different rate pitches, they probably do stimulate the same neural populations in the cochlea. As a result, the authors suggest that gap detection is performed at a central stage where changes in rate and amplitude have already been abstracted.

Second, our study has shown that gap thresholds of distant electrode pair configurations can improve dramatically after sufficient training, to the extent that the effect of channel separation even disappears for some subjects. The exact amount of training differs from subject to subject, but once the temporal cue is detected among other confounding cues (some subjects never succeed in doing this and some succeed only partly), the V-shaped flanks, indicating decreased performance with increasing channel distance, essentially disappear. These data suggest that gap thresholds in the across-channel conditions often do not reflect the true temporal resolution of the auditory system. Increased sensitivity may not result from increased neural interaction, but may rather

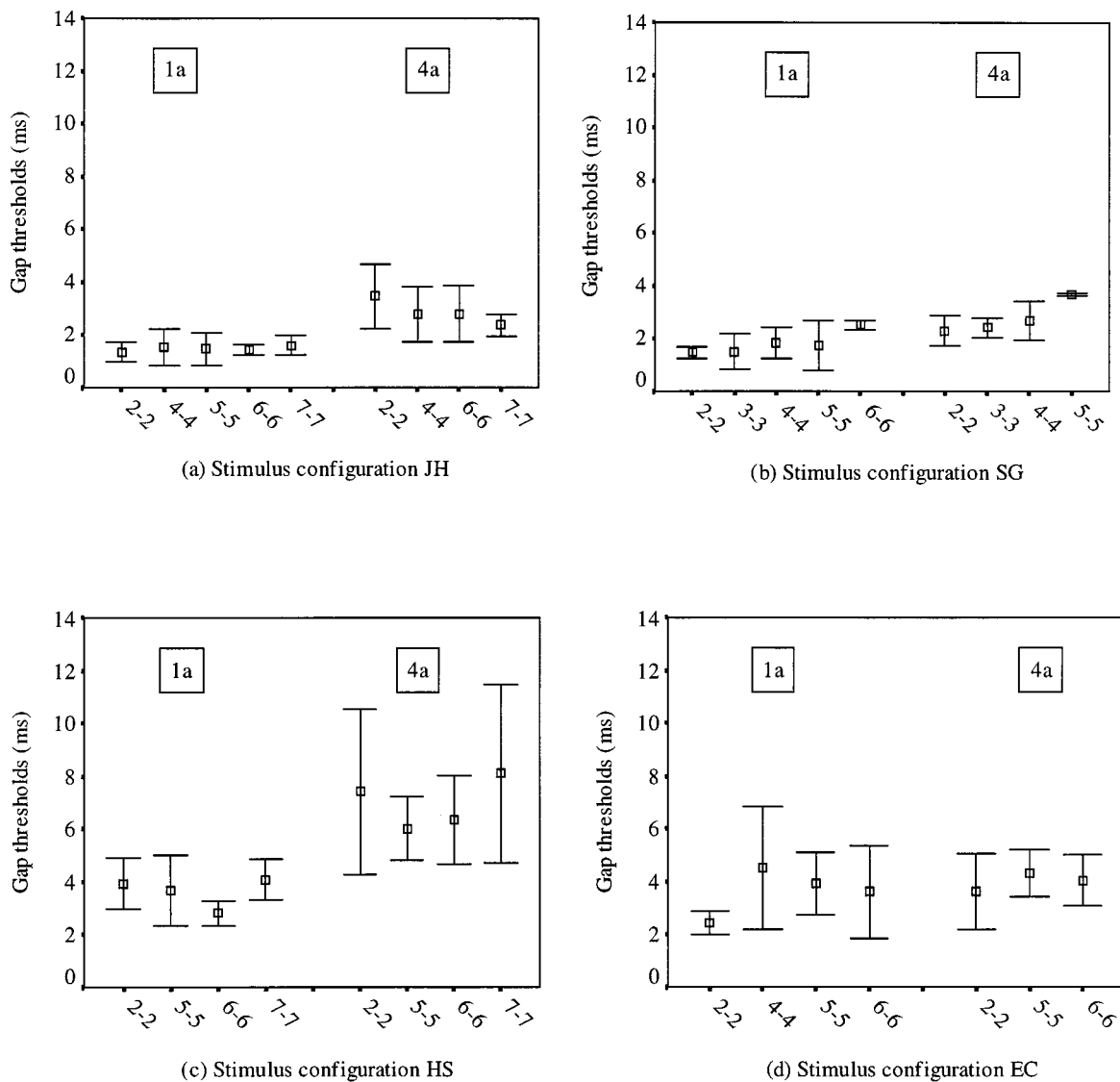


FIG. 8. Mean and standard deviations of the gap thresholds for 1250-pps (1a, exp. 1) and 400-pps (4a) stimuli. The abscissa indicates which channel is stimulated by the pre- and post-gap markers.

arise from the subject's ability to ignore the (confusing) perceptual difference between the pre- and post-gap marker.

Third, the masker had no clear effect on the data. If the masker had been effective, one would expect much larger gap thresholds as a result of reduced neural interaction, especially for nearby channels with the masker inserted in between. Indeed, gap-detection thresholds of subject HS do increase substantially in the masker condition, but probably because of increasing difficulty in listening selectively to the silent interval in the presence of other cognitively demanding cues, such as the masker.

Fourth, if gap detection could be used to infer the amount of overlap in neural populations stimulated by two channels, an effect of channel separation would also be expected for adjacent or distant channels stimulated within the same pre- or post-gap marker (e.g., ch.2-ch.3 vs ch.2-ch.6). However, none of our different multiple-channel configurations exhibited an effect of channel distance. Although there are no other electrical studies to compare our data with, several acoustical experiments have already shown that gap-

detection performance is independent of the frequency separation between different tones or noise bands in the same marker (Hall *et al.*, 1996; Formby *et al.*, 1998).

The relation between marker (a)symmetry and channel distance also deserves some more attention. Our study has shown that single- or multiple-channel symmetric configurations yielded the best gap thresholds (experiment 1), and that single-channel, asymmetric stimulus configurations yielded the worst gap thresholds (experiment 2). Gap detection, however, improved when a channel was added to either the pre- or post-gap marker (two-channel configuration, experiment 3). This is probably because the stimulus configuration now allowed discontinuity detection, albeit on one channel, instead of the more complex timing operation across different channels (Phillips *et al.*, 1997). In a pilot study, the relation between symmetry and channel distance was examined in more detail for electrical stimulation in two-channel stimuli. In this pilot test, both the pre- and post-gap marker stimulated two channels. However, the positions of the channels in the post-gap marker were either adjacent to the ones in the



pre-gap marker (e.g., pre-gap marker: ch.1 and ch.8, post-gap marker: ch.2 and ch.7), or distant from them (e.g., ch.1 and ch.8 vs ch.4 and ch.5). Gap thresholds of subjects JH and HS were very small when the pre- and post-gap markers stimulated the same (experiment 1) or adjacent channels, but performance deteriorated as the distance between the channels in the pre- and post-gap markers became larger (data not shown). This effect is similar to the one described for the single-channel stimulus configurations in experiment 2.

In conclusion, gap-detection thresholds in our study are short when different channels are stimulated quasi simultaneously within the same marker. However, for some cochlear implantees thresholds increase substantially when different channels are stimulated successively by the pre- and post-gap markers.

Judging by our subjects' observations, we are inclined to believe that performance is largely affected by the extent to which the subjects are capable of paying selective attention to the relevant temporal cue. This notion was already proposed by Phillips *et al.* (1997) and supported by Formby *et al.* (1998), both for acoustical data. The pre- and post-gap markers and the silent interval are probably grouped together and perceived as an auditory event, a process which requires a comparison of activity in different perceptual channels. This grouping is easy if the different perceptual channels occur both in the pre- and post-gap markers, but is difficult when they occur separately in time. It is possible that some subjects, like JH, do not perceive the stimulus as an entity, but are capable of attending analytically to the changing cue. Psychophysical correlates of selective auditory attention have been studied in so-called "streaming" experiments where some portions of a stimulus were found to have greater influence on performance than others (Bregman, 1990). However, although both gap detectability and stream segregation show an effect of frequency separation, the two processes may not necessarily be related to each other in a direct way (Neff *et al.*, 1982). More specific experiments are needed to understand how different "cross-spectral" phenomena influence auditory perception.

Due to methodological differences, it is difficult to compare the magnitude of our gap thresholds with those of Hanekom and Shannon (1998) for the across-channel stimulus conditions. In all our experiments, the duration of the marker was varied, to prevent overall duration from being a cue in detecting silent intervals (Formby and Forrest, 1991). If the duration of each marker is fixed, the stimulus containing the gap will always be longer than the one without the gap, and this may affect performance. However, the individual differences between the subjects will probably have more effect on the magnitude of the gap thresholds than the duration of the marker or the step size. Previously, it has been stated that gap detection is not a strong predictor of speech recognition (Shannon, 1989). However, this conclusion is based on symmetric, single-channel stimulus configurations, which yield similar thresholds across subjects (see also experiment 1). Although all our subjects were trained, performance differed substantially for the more complex patterns. While subject JH was capable of detecting short gaps from the beginning, irrespective of channel separation, mask-

ing (exp. 2) or asymmetry (exp. 3), subjects SG and HS needed training to ignore across-channel cues (although masking always negatively influenced gap detectability for HS). In addition, subject EC always performed relatively poorly in the across-channel task, despite extra training. For many it would be interesting to know whether these differences in performance are related to the subject's speech perception abilities, and whether differences in speech performance only occur as a result of differences in auditory acuity, or also as a result of the subject's (in)ability to combine information across different perceptual channels. Similarly, sensitivity to (changes in) rate may be related to the standard rate employed by a subject's speech processing strategy. This may explain why neither the Nucleus cochlear implant users (Busby and Clark, 1999) nor one of the LAURA implantees in this study showed an effect of rate, while the three other implantees clearly did. In the future these issues will be examined further in psychophysical experiments and speech tests, both with pre- and with post-lingually deafened implantees

## ACKNOWLEDGMENTS

We sincerely thank our subjects, who patiently and conscientiously detected gaps for many hours. We also thank Professor E. F. Offeciers of the University Dept. ORL of the St. Augustinus Hospital in Antwerp, who implanted these four subjects and provided access to them. Furthermore, we thank ir. L. Geurts (Lab. Exp. ORL, Katholieke Universiteit Leuven), for developing the software that drives the internal part of the LAURA device, and for writing the application software that enables stimulus presentation and response collection for different types of perceptual experiments. In addition, we would like to express our thanks to Dr. R. P. Carlyon (MRC, Cognition and Brain Sciences Unit, Cambridge, UK) for suggestions concerning methodological issues. This work was supported by the Fund for Scientific Research-Flanders (Belgium), and the Research Council of the Katholieke Universiteit Leuven.

- Bregman, A. S. (1990). *Auditory Scene Analyses* (MIT Press, Cambridge, MA).
- Busby, P. A., and Clark, G. M. (1999). "Gap detection by early-deafened cochlear implant subjects," *J. Acoust. Soc. Am.* **105**, 1841-1852.
- Chatterjee, M., Fu, Q.-J., and Shannon, R. V. (1998). "Within-channel gap detection using dissimilar markers in cochlear implant listeners," *J. Acoust. Soc. Am.* **103**, 2515-2519.
- Dobie, R. A., and Dillier, N. (1985). "Some aspects of temporal coding for single-channel electrical stimulation of the cochlea," *Hearing Res.* **18**, 41-55.
- Formby, C., and Forrest, T. G. (1991). "Detection of silent temporal gaps in sinusoidal markers," *J. Acoust. Soc. Am.* **89**, 830-837.
- Formby, C., Barker, C., Abbey, H., and Raney, J. J. (1993). "Detection of silent temporal gaps between narrow-band noise markers having second-formantlike properties of voiceless stop/vowel combinations," *J. Acoust. Soc. Am.* **93**, 1023-1027.
- Formby, C., Sherlock, L. P., and Forrest, T. G. (1996). "An asymmetric roex filter model for describing detection of silent gaps in sinusoidal markers," *Aud. Neurosci.* **3**, 1-20.
- Formby, C., Sherlock, L. P., and Li, S. (1998). "Temporal gap detection measured with multiple sinusoidal markers: Effects of marker number, frequency, and temporal position," *J. Acoust. Soc. Am.* **104**, 984-998.
- 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., and Hall, J. W. (1993). "Gap detection in a narrow band of noise in the presence of a flanking band of noise," *J. Acoust. Soc. Am.* **93**, 1645–1648.
- 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.
- Hanekom, J., and Shannon, R. V. (1998). "Gap detection as a measure of electrode interaction in cochlear implants," *J. Acoust. Soc. Am.* **104**, 2372–2384.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Moore, B. C. J., and Glasberg, B. R. (1988). "Gap detection with sinusoids and noise in normal, impaired and electrically stimulated ears," *J. Acoust. Soc. Am.* **83**, 1093–1101.
- Neff, D. L., Jesteadt, W., and Brown, E. L. (1982). "The relation between gap discrimination and auditory stream segregation," *Percept. Psychophys.* **31**(5), 493–501.
- Peeters, S., Offeciers, F. E., Joris, Ph., and Moeneclaey, L. (1993). "The Laura cochlear implant programmed with the continuous interleaved and phase-locked continuous interleaved strategies," *Adv. Oto-Rhino-Laryngol.* **48**, 261–268.
- Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (1997). "Detection of silent interval between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," *J. Acoust. Soc. Am.* **101**, 3694–3705.
- Phillips, D. P., Hall, S. E., Harrington, I. A., and Taylor, T. L. (1998). "'Central' auditory gap detection: A spatial case," *J. Acoust. Soc. Am.* **103**, 2064–2068.
- Plack, C. J., and Viemeister, N. F. (1992). "Intensity discrimination under backward masking," *J. Acoust. Soc. Am.* **92**, 3087–3101.
- Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (1995). "Intensity discrimination under forward and backward masking: Role of referential coding," *J. Acoust. Soc. Am.* **97**, 1141–1149.
- Preece, J. P., and Tyler, R. S. (1989). "Temporal-gap detection by cochlear prosthesis users," *J. Speech Hear. Res.* **32**, 849–856.
- Shannon, R. V. (1989). "Detection of gaps in sinusoids and pulse trains by patients with cochlear implant," *J. Acoust. Soc. Am.* **85**, 2587–2592.
- SPSS 7.5 (1997). SPSS, Inc., Chicago, IL.
- van Wieringen, A., and Wouters, J. (1999). "Natural vowel and consonant recognition by Laura cochlear implantees," *Ear Hear.* **20**, 89–103.
- Williams, K. N., and Perrott, D. R. (1972). "Temporal resolution of tonal pulses," *J. Acoust. Soc. Am.* **51**, 644–647.

# Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters

Steven van de Par

*IPO—Center for Research on User-System Interaction, NL-5600 MB Eindhoven, The Netherlands*

Armin Kohlrausch

*IPO—Center for Research on User-System Interaction, NL-5600 MB Eindhoven, The Netherlands*

*and Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands*

(Received 10 December 1997; revised 13 April 1999; accepted 22 June 1999)

Thresholds for sinusoidal signals masked by noise of various bandwidths were obtained for three binaural configurations:  $N_0S_0$  (both masker and signal interaurally in phase),  $N_0S_\pi$  (masker interaurally in phase and signal interaurally phase-reversed), and  $N_\pi S_0$  (masker interaurally phase-reversed and signal interaurally in phase). Signal frequencies of 125, 250, 500, 1000, 2000, and 4000 Hz were combined with masker bandwidths of 5, 10, 25, 50, 100, 250, 500, 1000, 2000, 4000, and 8000 Hz, with the restriction that masker bandwidths never exceeded twice the signal frequency. The overall noise power was kept constant at 70 dB SPL for all bandwidths. Results, expressed as signal-to-total-noise power ratios, show that  $N_0S_0$  thresholds generally decrease with increasing bandwidth, even for subcritical bandwidths. Only at frequencies of 2 and 4 kHz do thresholds appear to remain constant for bandwidths around the critical bandwidth.  $N_0S_\pi$  thresholds are generally less dependent on bandwidth up to two or three times the (monaural) critical bandwidth. Beyond this bandwidth, thresholds decrease with a similar slope as for the  $N_0S_0$  condition.  $N_\pi S_0$  conditions show about the same bandwidth dependence as  $N_0S_\pi$ , but thresholds in the former condition are generally higher. This threshold difference is largest at low frequencies and disappears above 2 kHz. An explanation for wider operational binaural critical bandwidth is given which assumes that binaural disparities are combined across frequency in an optimally weighted way. © 1999 Acoustical Society of America. [S0001-4966(99)02010-X]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Dc [DWG]

## INTRODUCTION

The difference in masked threshold that is obtained for a diotic  $N_0S_0$  condition (masker and signal interaurally in phase) and a dichotic  $N_0S_\pi$  condition (masker in phase, signal out of phase) has been studied extensively. This difference has been termed the binaural masking level difference (BMLD).

The understanding of the dependence of thresholds in a dichotic condition on the interaural parameters of the masker and signal is good, and can be modeled quite accurately with several binaural models (e.g., Durlach, 1972; Colburn, 1977). Domnitz and Colburn (1976) showed that in fact most binaural models can be classified in two groups which both give very similar predictions for a large set of binaural experiments involving diotic Gaussian-noise maskers. One group of models is based on the evaluation of interaural intensity differences and time delays (e.g., Jeffress *et al.*, 1956; Colburn and Durlach, 1978). The other group includes binaural cross-correlation models (e.g., Colburn, 1977; Lindemann, 1986; Stern and Shear, 1996). In terms of spectral properties, many stimulus parameters can be varied. The majority of BMLD studies are restricted to configurations where the masker and the signal are spectrally overlapping (e.g., Zurek and Durlach, 1987). Some studies, however, also investigated spectrally nonoverlapping maskers and signals and found smaller BMLDs than under similar conditions with spectral overlap (van der Heijden *et al.*, 1997).

Especially in the earlier studies on BMLDs, the masker usually was broadband and the dependence on interaural parameters and signal frequency was studied (e.g., Hirsh, 1948; Webster, 1951). A very general finding is that  $N_0S_\pi$  BMLDs are around 15 dB at low frequencies, but that BMLDs tend to decrease toward high frequencies. Above about 1.5 kHz, BMLDs of only a few decibels are found. For  $N_0S_m$  conditions (masker presented in phase, signal presented monaurally), BMLDs are generally 6 dB smaller at low frequencies than  $N_0S_\pi$  BMLDs, while at high frequencies no BMLDs are found. The loss of fine-structure coding has been assumed to be an important factor reducing BMLDs toward high frequencies (Zurek and Durlach, 1987; Bernstein and Trahiotis, 1996). Peripheral compression has been proposed to be an additional factor (van de Par and Kohlrausch, 1998; van de Par and Kohlrausch, 1999). Peripheral compression is assumed to reduce the size of interaural differences in level. This has been shown to be specifically detrimental to high-frequency binaural detection, since for such conditions only interaural differences within the envelopes of the stimuli are available for binaural processing. At low frequencies, binaural thresholds are not affected much by compression, since for such conditions additional interaural time delays in the fine structure are available which are not reduced by compression.

Another spectral variation which has been evaluated explicitly is the dependence of BMLDs on masker bandwidth (e.g., Bourbon and Jeffress, 1965; Sever and Small, 1979;

Zurek and Durlach, 1987). As far as we are aware, no binaural model has explicitly been evaluated for this dependence. However, the bandwidth dependence of the BMLD exhibits some interesting features. Generally, considerable increases in the BMLD are found when the masker bandwidth is decreased. At low frequencies using narrowband maskers, BMLDs of 25 dB are easily found, while for broadband maskers BMLDs of only 15 dB are reported (Zurek and Durlach, 1987; Kohlrausch *et al.*, 1995). Although there are no quantitative models that account for this bandwidth dependence, Zurek and Durlach (1987) suggested that the sluggishness of the binaural system may play an important role in this respect. At narrow bandwidths, the inherent rate of fluctuation in the masker is small. Consequently, the interaural differences in intensity and timing will vary slowly. Increasing the masker bandwidth will increase the rate of change in these interaural cues, and this should be detrimental to binaural detection.

Another feature of the bandwidth dependence of the BMLD is that in  $N_0S_\pi$  band-widening experiments, the estimated filter bandwidth seems to be a factor 2 or 3 larger than for  $N_0S_0$  stimuli (Bourbon and Jeffress, 1965; Sever and Small, 1979; Hall *et al.*, 1983; van der Heijden and Trahiotis, 1998).

Despite the large body of binaural data, the interaction between center frequency, masker bandwidth, and binaural condition has never been measured in the same subjects and laboratory conditions. Therefore, this study provides measurements of  $N_0S_0$ ,  $N_0S_\pi$ , and  $N_\pi S_0$  data for center frequencies from 125 Hz to 4 kHz and bandwidths varying from 5 Hz up to two times the center frequency of the masker, all measured with the same subjects and the same experimental setup.

On the one hand, this set of data can serve as a useful reference for modeling dependencies of binaural detection on center frequency, bandwidth, and interaural parameters. On the other hand, this coherent set of data covering a broad range of stimulus conditions will be used in the discussion to introduce and support a framework for understanding the differential bandwidth dependencies that are found in monaural and binaural signal detection experiments.

## I. METHOD

### A. Procedure

A three-interval forced-choice procedure with adaptive signal-level adjustment was used to determine masked thresholds. The three masker intervals of 400-ms duration were separated by pauses of 200 ms. A signal of 300-ms duration was added in the temporal center to one of these three intervals. The subject's task was to indicate which of the three intervals contained the signal. Feedback was provided to the subjects after each trial.

The signal level was adjusted according to a two-down one-up rule tracking the 70.7% correct response level (Levitt, 1971). The initial step size for adjusting the level was 8 dB. After each second reversal of the level track, the step size was halved until a step size of 1 dB was reached, after which the run was continued for another eight reversals.

From the level of these last eight reversals, the median was calculated and used as a threshold value. At least four threshold values were obtained and averaged for each parameter value and subject. Thresholds are expressed as signal-to-overall-noise power ratio.

Three subjects participated in this experiment, among them the authors. One other subject was paid for his services. All subjects had normal hearing.

### B. Stimuli

All stimuli were generated digitally and converted to analog signals with a two-channel, 16-bit D/A converter at a sampling rate of 32 kHz. The signals were presented to the subjects over Telephonics TDH-49p headphones with supra-aural cushions. For each masker bandwidth, the masker level was set to an overall sound-pressure level of 70 dB. Subjects were seated in a sound-insulated booth that was placed within a sound-insulated room in our laboratory.

For each interval, a new 400-ms masker sample was obtained by randomly selecting a segment from a 2000-ms bandpass-noise buffer. Before each adaptive run, a new noise buffer was created. The band-limited noise buffer was created in the frequency domain by generating a flat spectrum within the passband and randomizing the phases. After an inverse Fourier transform, the noise buffer of 2000 ms was obtained. The 300-ms signals were sinusoids with a frequency equal to the center frequency of the noise masker. In order to avoid spectral splatter, the signal and the maskers were gated with 50-ms raised-cosine ramps. Thresholds are expressed as signal-to-overall-noise level.

$N_0S_0$ ,  $N_0S_\pi$ , and  $N_\pi S_0$  thresholds were measured for maskers with a center frequency of 125, 250, 500, 1000, 2000, and 4000, and bandwidths of 5, 10, 25, 50, 100, 250, 500, 1000, 2000, 4000, and 8000 Hz, with the restriction that bandwidths never exceeded twice the center frequency of the masker.

## II. RESULTS

Although for some conditions there were considerable differences in the results of the three subjects, we present only results averaged across subjects. We remark that inter-subject variability was small for most conditions. However, at a center frequency of 4 kHz, for bandwidths up to 1000 Hz, a standard deviation across subjects of about 5 dB was found. Furthermore, at a center frequency of 125 Hz, a somewhat larger than average standard deviation across subjects of 3 dB was found for all bandwidths. A relatively large intersubject variability at high frequencies has also been reported by Koehnke *et al.* (1986).

Average  $N_0S_0$ ,  $N_0S_\pi$ , and  $N_\pi S_0$  thresholds are shown in Fig. 1 as a function of masker bandwidth. The six panels show thresholds for six center frequencies ranging from 125 Hz to 4 kHz. The isolated symbols which are shown at the top right of each panel show standard deviations of the data across subjects averaged across bandwidth; the threshold levels for these conditions have no meaning.

In all conditions,  $N_0S_\pi$  (squares) and  $N_\pi S_0$  (diamonds) thresholds are considerably lower than  $N_0S_0$  (circles) thresh-



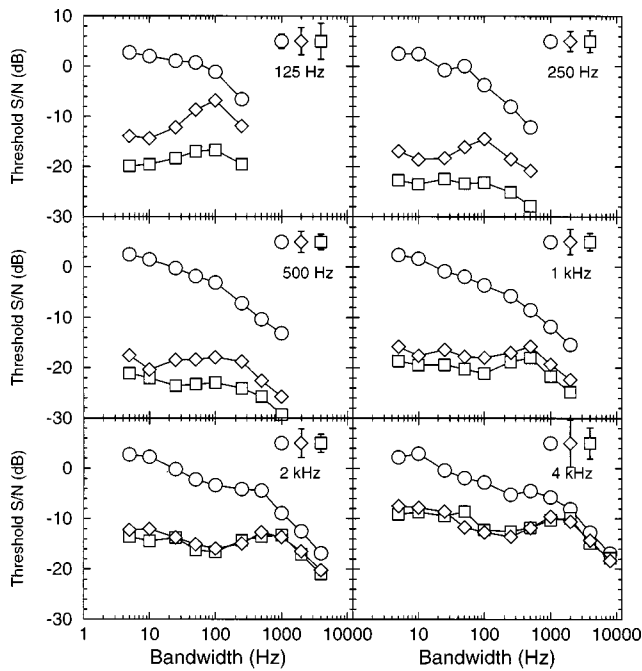


FIG. 1.  $N_0S_0$  (circles),  $N_0S_\pi$  (squares), and  $N_\pi S_0$  (diamonds) thresholds as a function of masker bandwidth expressed in signal-to-overall-noise power ratio. The six panels show data at various center frequencies. The isolated symbols show standard deviations across subjects, averaged across masker bandwidth.

olds. Only at 4 kHz for bandwidths beyond 1 kHz is there nearly no BMLD, in line with data in Hirsh and Burgeat (1958) and Zurek and Durlach (1987).  $N_0S_\pi$  and  $N_\pi S_0$  data are lowest at 250 and 500 Hz and show a general tendency to increase toward high frequencies. For  $N_\pi S_0$ , the thresholds at 125-Hz center frequency are relatively high, in some cases even higher than at 4 kHz (cf. van de Par and Kohlrausch, 1997). At the lowest frequencies,  $N_0S_\pi$  thresholds are 5 to 10 dB lower than  $N_\pi S_0$  thresholds. This difference decreases toward 1 kHz and at 2 and 4 kHz no difference is measured. Such results were also found by Hirsh and Burgeat (1958) in broadband masking noise.

Beyond a certain critical bandwidth all curves show a relatively steep decrease. The slope of the curves beyond these bandwidths is fairly similar. These critical bandwidths seem to be larger for the  $N_0S_\pi$  and  $N_\pi S_0$  conditions than for the  $N_0S_0$  condition. For narrow bandwidths and center frequencies of 500 Hz and more, the thresholds for the  $N_0S_\pi$  and  $N_\pi S_0$  condition are more constant than the  $N_0S_0$  thresholds, which consistently decrease with increasing bandwidth. Thus it is mainly due to the decrease in  $N_0S_0$  thresholds that BMLDs decline when bandwidths are increased up to the critical bandwidth.

In order to compare each of the conditions  $N_0S_0$ ,  $N_0S_\pi$ , and  $N_\pi S_0$  across center frequencies, data are replotted into a single figure for each of these three conditions. In Fig. 2,  $N_0S_0$  thresholds are shown as a function of masker bandwidth for the six center frequencies. Thresholds are very similar for small masker bandwidths, with the exception of 125-Hz thresholds, which are somewhat higher. In this range there is a general decrease of 1.5 dB octave, indicated by the solid auxiliary line. Beyond a certain bandwidth, thresholds

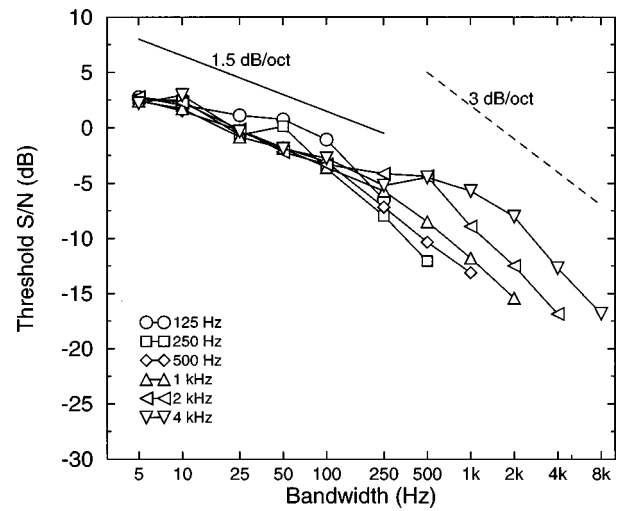


FIG. 2.  $N_0S_0$  thresholds for center frequencies of 125 (circles), 250 (squares), 500 Hz (diamonds), 1 (triangles up), 2 (triangles left), and 4 kHz (triangles down) as a function of masker bandwidth for a constant overall masker level of 70 dB SPL. Thresholds are expressed as signal-to-overall-noise power ratio. The solid and dashed auxiliary lines indicate 1.5 dB/octave and 3 dB/octave slopes, respectively.

decrease with a slope of 3 dB/octave (see dashed auxiliary line) or even 4.5 dB/octave for 2 and 4 kHz. For increasing center frequency, the transition between the two slopes of 1.5 dB/octave and 3 dB/octave occurs at larger bandwidths.

In Fig. 3, thresholds are shown for  $N_0S_\pi$  in the same format as Fig. 2. Clearly,  $N_0S_\pi$  thresholds are lowest for center frequencies in the range of 250 to 1000 Hz, while the 4-kHz thresholds are much higher than the rest.

For the smaller bandwidths, thresholds are rather independent of masker bandwidth. Only the 125-Hz data increase toward the 100-Hz bandwidth before they start to decrease with 3 dB/octave (see dashed auxiliary line). The 2- and 4-kHz data first decrease somewhat at the smallest bandwidths, then increase again before the 3-dB/octave decrease in thresholds starts. Beyond a certain point, all curves de-

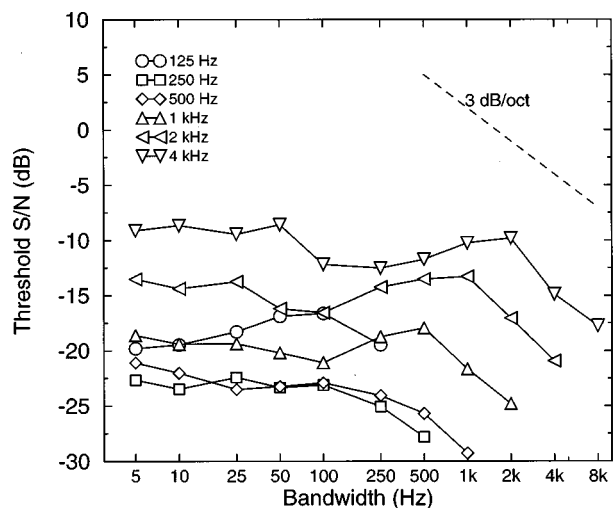


FIG. 3.  $N_0S_\pi$  thresholds for center frequencies of 125 (circles), 250 (squares), 500 Hz (diamonds), 1 (triangles up), 2 (triangles left), and 4 kHz (triangles down) as a function of masker bandwidth. Thresholds are expressed as signal-to-overall-noise power ratio.

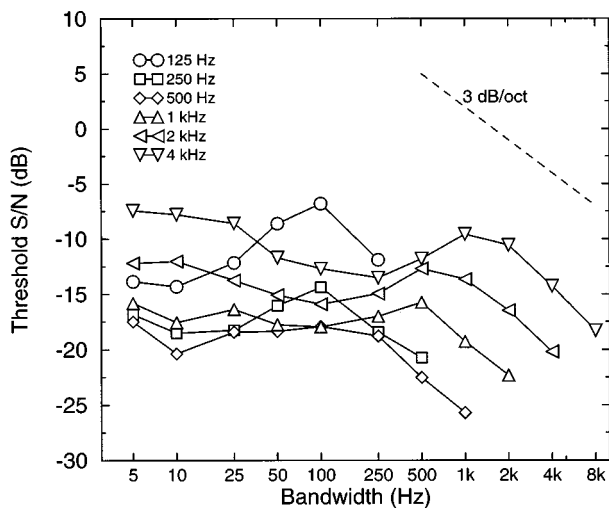


FIG. 4.  $N_{\pi}S_0$  thresholds for center frequencies of 125 (circles), 250 (squares), 500 Hz (diamonds), 1 (triangles up), 2 (triangles left), and 4 kHz (triangles down) as a function of masker bandwidth. Thresholds are expressed as signal-to-overall-noise power ratio.

crease with 3 dB/octave, in line with data of Zurek and Durlach (1987) measured at center frequencies of 250 Hz and 4 kHz.

In Fig. 4,  $N_{\pi}S_0$  data are shown. For this condition, a similar behavior is found as in Fig. 3. The lowest thresholds are found over a larger range of center frequencies (250 to 2000 Hz). Thresholds for a center frequency of 125 Hz for the  $N_{\pi}S_0$  condition are clearly higher than for the  $N_0S_{\pi}$  condition in Fig. 3, and increase strongly toward a bandwidth of 100 Hz. The  $N_{\pi}S_0$  thresholds at center frequencies of 125 and 500 Hz correspond well to data of Langford and Jeffress (1964) and Breebaart *et al.* (1998). Beyond a certain bandwidth, all thresholds decrease at a 3-dB/octave rate again.

In order to compare estimated critical bandwidths from  $N_0S_0$  results to those obtained in notched-noise experiments, the values of Fig. 2 are replotted in Fig. 5, with bandwidth

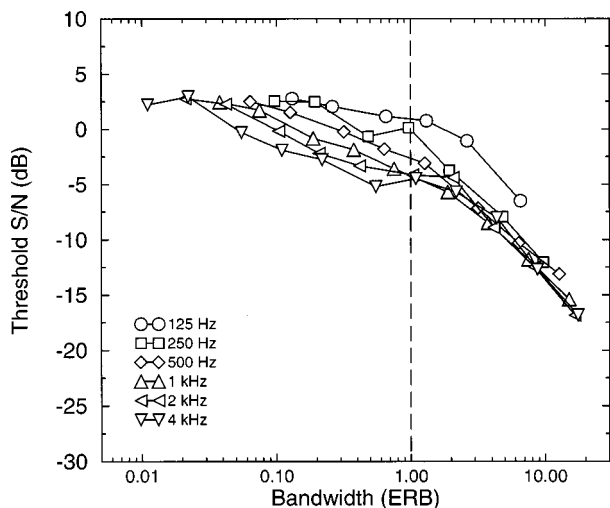


FIG. 5.  $N_0S_0$  thresholds for center frequencies of 125 (circles), 250 (squares), 500 Hz (diamonds), 1 (triangles up), 2 (triangles left), and 4 kHz (triangles down) as a function of masker bandwidth expressed in ERBs. The dashed line indicates a bandwidth of 1 ERB. Thresholds are expressed as signal-to-overall-noise power ratio.

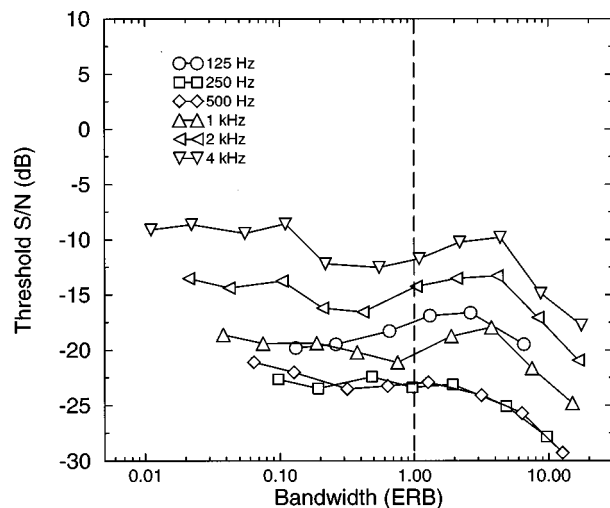


FIG. 6.  $N_0S_{\pi}$  thresholds for center frequencies of 125 (circles), 250 (squares), 500 Hz (diamonds), 1 (triangles up), 2 (triangles left), and 4 kHz (triangles down) as a function of masker bandwidth expressed in ERBs. Thresholds are expressed as signal-to-overall-noise power ratio.

expressed in units of equivalent rectangular bandwidths (ERBs) such as derived by Glasberg and Moore (1990).<sup>1</sup> Plotting the data in this way is expected to result in an alignment of all curves such that the points where the 1.5-dB/octave slope changes into a 3-dB/octave slope are all found at 1-ERB bandwidth. At medium center frequencies, the critical bandwidths of our experiment are reasonably in line with those obtained in notched-noise experiments (ERB=1, see dashed line). At high center frequencies, our experimental data reveal critical bandwidths of about 2 ERB. Furthermore, all but the 125-Hz curves coincide beyond 2 ERB.

The  $N_0S_{\pi}$  data are replotted in Fig. 6, with bandwidth expressed in units of ERBs. Unlike the  $N_0S_0$  curves in Fig. 5, the  $N_0S_{\pi}$  curves do not coincide beyond the critical bandwidth. Furthermore, the data for the  $N_0S_{\pi}$  condition show the 3-dB/octave decrease only when bandwidths are a factor 2 to 4 larger than the monaural critical bandwidths as measured by Glasberg and Moore (1990). High-frequency thresholds even seem to increase beyond 1 ERB. Such a large apparent binaural critical bandwidth had been reported earlier by, e.g., Bourbon and Jeffress (1965), Sever and Small (1979), Hall *et al.* (1983), and van der Heijden and Trahiotis (1998).

### III. DISCUSSION

#### A. Monaural signal detection: Dependence on masker bandwidth

The  $N_0S_0$  thresholds reveal a bandwidth dependence that is well in line with the view that the decision variable that is used in the detection process has the same statistics as the energy at the output of an auditory filter (Richards and Heller, 1991). For bandwidths which are smaller than a certain critical bandwidth, thresholds decrease with 1.5 dB per doubling of the bandwidth. This dependence probably results from the sample-to-sample variability of the masker energy (Bos and de Boer, 1966; Green and Swets, 1974), which, for any fixed sample duration, decreases with 1.5 dB/octave.

Figure 2 shows that for subcritical bandwidths nearly all curves fall on top of each other. This is in line with the assumption that the stimulus variability limits the detection process, because this variability is identical for all center frequencies.

For some conditions, the 1.5-dB/octave slope is not found. For example, for 125-Hz center frequency, the overall slope is somewhat less, while above 2 kHz, thresholds seem to change less with bandwidth just around the critical bandwidth. This effect seems to indicate that for these specific conditions, not the external stimulus variability but rather the internal noise is the dominant factor.

The fact that almost all  $N_0S_0$  curves in Fig. 5 coincide beyond 1 ERB is in line with a filter bank model where the average energy at the output of the filter centered around the signal has to increase with a fixed percentage for a signal to be detected. If the masking noise sufficiently exceeds the critical bandwidth, the total energy of the masker within an auditory filter is proportional to the ERB divided by the total masker bandwidth. Since we use a fixed detection criterion, this ratio will be the only parameter determining detection thresholds provided that bandwidths are larger than the critical bandwidth. In Fig. 5, the reciprocal value of this ratio is plotted along the  $x$  axis, and therefore beyond 1 ERB all curves are expected to coincide. The data for a center frequency of 125 Hz are somewhat puzzling, and we have no clear explanation why they deviate from the other data.

Matters are more complicated when the variability in the masker energy is taken into account. Since the critical bandwidth varies with center frequency, low signal-frequency conditions in our experiments are expected to have more variability at the output of the filters at detection threshold, which would lead to relatively high signal-to-noise ratios. Such a frequency dependence of the signal-to-noise ratio, however, is not reflected in the  $N_0S_0$  curves, most of which coincide beyond the critical bandwidth. The fact that this changing signal-to-noise ratio is not observed in Fig. 5 may mean two things. Either noise variability does not influence detection in broadband noise, or the ERBs reported in the literature are also affected by variability in stimulus energy in some way.

## B. An explanation for the wider operational binaural critical bandwidth

The  $N_0S_\pi$  data were replotted in Fig. 6 in a similar way as in Fig. 5. This figure reveals two important aspects of binaural detection. Unlike the  $N_0S_0$  curves in Fig. 5, the  $N_0S_\pi$  curves do not coincide beyond the critical bandwidth. Instead, the tendency is more that the various curves differ in terms of a more or less constant offset in threshold. This suggests that, after compensation for the differences in critical bandwidth, what is differing across center frequency in terms of the detection process is the sensitivity to interaural differences.

The operational binaural critical bandwidths that are revealed by the data for the  $N_0S_\pi$  condition are a factor 2 to 4 larger than the monaural critical bandwidths as measured by Glasberg and Moore (1990). In addition, for high frequencies

there even seems to be an increase in thresholds beyond 1 ERB.

An explanation for the wider operational binaural critical bandwidth has been presented by Hall *et al.* (1983). They assume that in wideband noise the binaural system cannot concentrate only on the central filter but is also influenced by the adjacent filters. For broadband noise, the adjacent filters contain hardly any interaural differences, and Hall *et al.* assume that incorporation of this information is detrimental for detecting the  $S_\pi$  signal. When narrowing the bandwidth down to the monaural critical bandwidth, keeping the spectral level of the noise constant, the number of auditory filters involved in the detection process is assumed to be reduced. In this process, a larger percentage of the filters will contain interaural differences, which leads to a better detectability of the signal. Thus thresholds will start to decrease before the critical bandwidth is reached. This view cannot explain why a large operational binaural critical bandwidth is only observed at high stimulus levels.

Recently, van der Heijden and Trahiotis (1998) studied detection thresholds for an interaurally out-of-phase tone in noise maskers with interaural correlations from  $-1$  to  $1$  ( $N_\rho S_\pi$  condition) as a function of masker bandwidth. For nearly all interaural correlations, except for correlations very close to  $1$ , the pattern of thresholds revealed an underlying filter bandwidth well in line with monaural estimates of critical bandwidth. Such a pattern was observed even for interaural noise correlations where there were sizable MLDs indicating that binaural processing was effective. Only for interaural noise correlations of  $1$ , or close to  $1$ , did the patterns of thresholds indicate a large operational critical bandwidth.

These results led van der Heijden and Trahiotis (1998) to conclude that the large operational critical bandwidth did not result from a reduced frequency selectivity in binaural processing; otherwise, large bandwidths should have been found for all those interaural noise correlations that led to a BMLD. Rather, the specific details of the data suggest that the large operational binaural critical bandwidths are related to a dependence of the internal noise on masker bandwidth yielding more internal noise with increasing masker bandwidth. Note that Metz *et al.* (1968) suggested a similar assumption with respect to internal noise in the context of an EC model.

Along the lines set out in van der Heijden and Trahiotis (1998), we therefore suggest the following explanation for the “wider critical band” seen in band-widening experiments. We assume that the binaural system either attends to only one or to more than one filter channel, depending on which option results in the best detection given the specific stimulus properties. Such a scheme is in fact the binaural equivalent of a model proposed by Florentine and Buus (1981) for monaural level discrimination. In this model, the total sensitivity,  $d'$ , is the root-mean-square (RMS) of the  $d'$ 's at the outputs of all individual auditory filters. This RMS combination of  $d'$ 's is valid provided that internal noise is the limiting factor for detection and that the internal noises in all filters are uncorrelated. We assume that these assumptions also apply to binaural detection.

For binaural signal detection in narrow-band noise, both the central filter and adjacent filters in principal provide useful binaural information. Combining the binaural information of these filters by taking the weighted sum of the decision variables at the output of all separate filters will improve the detection process with respect to a situation where only one filter is taken into account. This is true provided that the internal noises within each frequency channel are independent. In such a case, the effective amount of internal noise will be reduced with respect to the case where only one filter can be used for detection. Langhans and Kohlrausch (1992) showed that the binaural system is indeed able to integrate binaural information across multiple filters by using multi-component signals in a broadband masker, suggesting that indeed internal noise among frequency channels is independent.

For broadband maskers, where only the central filter provides useful information, attending to more than this filter would reduce sensitivity. Thus in this case only one filter would be used, instead of various filters such as with narrow-band noise, and therefore the effective internal noise would be smaller in a narrow band as compared to a broadband masker. As mentioned earlier, a similar conclusion about internal noise was reached by van der Heijden and Trahiotis (1998) on the basis of experiments with  $N_\rho S_\pi$  signals with various values of  $\rho$  and various masker bandwidths. Thus this difference in internal noise for broadband and narrow-band maskers can explain part of the smaller BMLDs that are observed for broadband maskers. In addition, the 1.5-dB/octave reduction for the  $N_0 S_0$  condition is also an important factor reducing the BMLD with increasing bandwidth.

For the  $N_0 S_\pi$  band-widening experiments presented in this study, the large apparent binaural critical bandwidth can be explained as follows. For masker bandwidths which are narrower than the monaural critical bandwidth, the interaural differences across the whole excitation pattern can be integrated to reduce internal noise in the binaural detection process. Increasing bandwidths to slightly beyond the monaural critical bandwidth will introduce two effects. First of all, the masker energy in the central filter will be reduced due to the peripheral filtering (the reader is reminded that the *total* masker level is kept constant). On the other hand, signals in the off-frequency filters will start to exhibit less interaural differences. Thus the binaural system will start to attend more to the central filter, which will lead to an effective increment in internal noise which in turn will tend to increase thresholds. The filtering effect and the increased internal noise have opposite effects on thresholds (expressed as signal-to-overall-noise ratio) and thus thresholds remain rather constant with increasing bandwidth. With even wider bandwidths, the off-frequency filters will eventually cease to contain interaural differences and only the central filter will be used for detection. At this point, further increments of the masker bandwidth will lead to a 3-dB/octave reduction in thresholds. Thus due to these opposite effects, the binaural critical bandwidth appears to be larger than the monaural critical bandwidth.

When we consider the study by Hall *et al.* (1983), their band-limiting experiment showed a large operational binau-

ral critical bandwidth at medium and high masker levels, in line with our view. In addition, Hall *et al.* (1983) found that operational binaural critical bandwidths are smaller at low masker levels. This can be shown to be also in line with our proposal. For low levels, the excitation pattern is fairly limited in width, and therefore off-frequency filters will not contribute much to the detection process and a single-filter binaural model provides an adequate description in this case, resulting in a small operational binaural critical bandwidth.

A similar line of reasoning applies to the notched-noise experiments presented in Hall *et al.* (1983). In these experiments, the binaural processor has to rely on the on-frequency filter, since all other filters contain no interaural differences. Thus for such an experiment, a single-filter binaural model is adequate and an operational binaural critical bandwidth should be measured that is similar to the monaural critical bandwidth.

In fact, the same argument that is used for notched-noise experiments holds for all conditions where a tonal signal is used together with a broadband masker as, e.g., in Kohlrausch (1988). In this study the auditory filter bandwidth was determined by measuring the detectability of in-phase and out-of-phase tonal signals of variable frequency. These signals were presented within a broadband masker which had an interaural correlation of 1 below a certain transition frequency and  $-1$  above the transition frequency. In this case, independent of the specific condition, there will always be only one filter where there is useful information for either the binaural or monaural processor, and integration across frequency channels cannot create apparently larger critical bandwidths.

### C. Implications for monaural detection

At this point it is worthwhile to consider the implications of this scheme for the monaural domain. As mentioned before, the combination of binaural information across several filters effectively leads to a reduction in internal noise. However, when the monaural detection process is considered for a narrow-band  $N_0 S_0$  condition, it seems that the sample-to-sample variability in the masker energy is the limiting factor for signal detection, and this variability is strongly correlated across auditory filters. Therefore, for a narrow-band monaural condition, combining information across frequency will not lead to any improvement in signal detection.

Increasing the masker bandwidth in an  $N_0 S_0$  condition will lead to a decrease in the sample-to-sample variability of the masker energy. Eventually, it is expected that the variability is smaller than the internal noise of the monaural processor. If this point is reached before the monaural critical bandwidth is exceeded, we expect in such a case that similar to the binaural detector, the apparent monaural critical bandwidth will be wider than, for example, the critical bandwidth measured in a notched-noise experiment. We know that the monaural system is able to detect level differences of about 1 dB. In order to increase the level of a noise stimulus by 1 dB by adding a tone, the tone would have to be added with a signal-to-noise ratio of  $-5.9$  dB. The  $N_0 S_0$  thresholds for 2 and 4 kHz for bandwidths of 250 and 500 Hz shown in Fig. 5 are not too far from this signal-to-noise ratio. For these



conditions it is not unlikely that internal noise starts to play a role, and that critical bandwidths are expected to exceed 1 ERB. Indeed this effect can be seen in the  $N_0S_0$  data for 2- and 4-kHz center frequencies.

#### D. Dependencies on center frequency and interaural parameters

For narrow bandwidths, the  $N_0S_\pi$  thresholds show the general trend of increasing toward high frequencies. A total increment of about 13 dB is found between 500 Hz and 4 kHz. Generally, this increment has been attributed to the loss of phase locking in the auditory nerve above about 1.5 kHz (e.g., Durlach, 1964; Zurek and Durlach, 1987; Bernstein and Trahiotis, 1996). Due to this loss of phase locking, only interaural intensity differences (IIDs) and no interaural time delays in the fine structure (ITDs) are available for the binaural processor at high frequencies. At low frequencies both binaural cues can be used, leading to a better binaural detection. However, within the framework of cross-correlation models, only a 3-dB difference would be expected between narrow-band low- and high-frequency  $N_0S_\pi$  thresholds if equal sensitivity is assumed at low and high frequencies (van de Par and Kohlrausch, 1995).

Recently, results of several experiments have suggested that the compressive response of the basilar membrane to auditory stimulation is an additional important factor (van de Par and Kohlrausch, 1998; van de Par and Kohlrausch, 1999). Such a compressive response will reduce the size of IIDs both at low and high frequencies. However, at low frequencies this will have a relatively small effect on thresholds since for low frequencies, fine-structure ITDs can also be used, while at high frequencies only interaural differences within the envelope can be used. This combined effect of the loss of fine-structure coding at high frequencies and peripheral compression can explain the much larger than 3-dB difference between low frequency and high frequency thresholds.

As an additional factor for the lower sensitivity at high frequencies, the sluggishness of the binaural system has been mentioned (Zurek and Durlach, 1987). When broadband-noise maskers are used, the increase in auditory-filter bandwidth toward high frequencies results in an increase in the rate of fluctuation of interaural cues, which is suggested to be detrimental to binaural detection. This would suggest that, provided that the auditory filter bandwidth is large enough to allow sufficiently rapid fluctuation rates, there would be a certain value for the masker bandwidth (smaller than the monaural critical bandwidth) beyond which binaural thresholds should start to increase. However, the  $N_0S_\pi$  data in Fig. 3 do not show a systematic increase in thresholds above a certain fixed subcritical bandwidth, as would be expected when binaural sluggishness plays a role.

Comparing  $N_0S_\pi$  with  $N_\pi S_0$  thresholds, we find that the difference in thresholds increases toward low frequencies. For high frequencies where only interaural differences in the envelope can be used, one would expect that  $N_0S_\pi$  thresholds and  $N_\pi S_0$  thresholds would be identical.

However, besides this effect that is due to the loss of fine-structure coding, there is an additional tendency to see

larger differences between  $N_0S_\pi$  and  $N_\pi S_0$  thresholds for decreasing center frequencies below 1 kHz. Especially at 125 Hz, there is a large difference. In fact,  $N_\pi S_0$  thresholds for some of the bandwidths exceed similar thresholds at high frequencies. It seems that, in general, the interaurally phase-reversed character of the noise affects detection more toward lower frequencies. Probably the large internal delay that corresponds to the interaurally phase-reversed masker for the lowest center frequencies is a determining factor.

In conclusion, we found apparently wider critical bandwidths for the  $N_0S_\pi$  and  $N_\pi S_0$  conditions as compared to the monaural  $N_0S_0$  condition at all frequencies. It seems that a scheme that combines binaural information across frequency in an optimal way provides an explanation for the wider operational binaural critical bandwidth that is observed in band-widening experiments. In addition, it is in line with several other studies that have measured binaural critical bandwidth with different experimental techniques, revealing no differences from monaural estimates.

#### ACKNOWLEDGMENTS

The authors thank Professor Adrian Houtsma, the associate editor, D. W. Grantham, and two anonymous reviewers for their very helpful comments on this manuscript.

<sup>1</sup>Bandwidths in units of ERBs are obtained by dividing bandwidths expressed in Hz by the equivalent rectangular bandwidth of the filter that has the same center frequency as the masker. The formula that was used to calculate the equivalent rectangular bandwidth is taken from Glasberg and Moore (1990):  $ERB = 24.7(1 + 0.0047f_0)$ , where  $f_0$  is the center frequency of the filter and the masking noise.

- Bernstein, L. R., and Trahiotis, C. (1996). "The normalized correlation: Accounting for binaural detection across center frequency," *J. Acoust. Soc. Am.* **100**, 3774–3784.
- Bos, C. E., and de Boer, E. (1966). "Masking and discrimination," *J. Acoust. Soc. Am.* **39**, 708–715.
- Bourbon, W. T., and Jeffress, L. A. (1965). "Effect of bandwidth of masking noise on the detection of homophasic and antiphase tonal signals," *J. Acoust. Soc. Am.* **37**, 1180–1181.
- Breebaart, J., van de Par, S., and Kohlrausch, A. (1998). "Binaural signal detection with phase-shifted and time-delayed noise maskers," *J. Acoust. Soc. Am.* **103**, 2079–2083.
- Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," *J. Acoust. Soc. Am.* **61**, 525–533.
- Colburn, H. S., and Durlach, N. I. (1978). "Models of binaural interaction," in *Hand book of Perception: Hearing*, edited by E. Carterette and M. Friedman (Academic, New York).
- Domnitz, R. H., and Colburn, H. S. (1976). "Analysis of binaural detection models for dependence of interaural parameters," *J. Acoust. Soc. Am.* **59**, 598–601.
- Durlach, N. I. (1964). "Note on binaural masking-level differences at high frequencies," *J. Acoust. Soc. Am.* **36**, 576–581.
- Durlach, N. I. (1972). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. II.
- Florentine, M., and Buus, S. (1981). "An excitation-pattern model for intensity discrimination," *J. Acoust. Soc. Am.* **70**, 1646–1654.
- Green, D. M., and Swets, J. A. (1974). *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966); reprinted by Krieger, New York.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.* **47**, 103–138.
- Hall, J. W., Tyler, R. S., and Fernandez, M. A. (1983). "Monaural and binaural auditory frequency resolution measured using bandlimiting noise and notched-noise masking," *J. Acoust. Soc. Am.* **73**, 894–898.

- van der Heijden, M. L., and Trahiotis, C. (1998). "Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution," *J. Acoust. Soc. Am.* **103**, 1609–1614.
- van der Heijden, M. L., Trahiotis, C., Kohlrausch, A., and van de Par, S. (1997). "Binaural detection with spectrally non-overlapping signals and maskers: Evidence for masking by aural distortion products," *J. Acoust. Soc. Am.* **102**, 2966–2972.
- Hirsh, I. J. (1948). "The influence of interaural phase on interaural summation and inhibition," *J. Acoust. Soc. Am.* **20**, 536–544.
- Hirsh, I. J., and Burgeat, M. (1958). "Binaural effects in remote masking," *J. Acoust. Soc. Am.* **30**, 827–832.
- Jeffress, L. A., Blodgett, H. C., Sandel, T. T., and Wood III, C. L. (1956). "Masking of tonal signals," *J. Acoust. Soc. Am.* **28**, 416–426.
- Koehnke, J., Durlach, N. I., and Colburn, H. S. (1986). "Performance in several binaural-interaction experiments," *J. Acoust. Soc. Am.* **79**, 1558–1562.
- Kohlrausch, A. (1988). "Auditory filter shape derived from binaural masking experiments," *J. Acoust. Soc. Am.* **84**, 573–583.
- Kohlrausch, A. van de Par, S.L.J.D.E., and Houtsma, A. J. M. (1995). "A new approach to study binaural interaction at high frequencies," in *Advances in Hearing Research, Proceedings of the 10th International Symposium on Hearing*, edited by G. A. Manley, G. M. Klump, C. Köppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore), pp. 343–353.
- Langford, T. L., and Jeffress, L. A. (1964). "Effect of noise cross correlation on binaural signal detection," *J. Acoust. Soc. Am.* **36**, 1455–1458.
- Langhans, A., and Kohlrausch, A. (1992). "Spectral integration of broadband signals in diotic and dichotic masking experiments," *J. Acoust. Soc. Am.* **91**, 317–326.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Lindemann, W. (1986). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.* **80**, 1608–1622.
- Metz, P. J., von Bismarck, G., and Durlach, N. I. (1968). "Further results on binaural unmasking and the EC model. II. Noise bandwidth and interaural phase," *J. Acoust. Soc. Am.* **43**, 1085–1091.
- van de Par, S., and Kohlrausch, A. (1995). "Analytical expressions for the envelope correlation of certain narrowband stimuli," *J. Acoust. Soc. Am.* **98**, 3157–3169.
- van de Par, S., and Kohlrausch, A. (1997). "A new approach to comparing binaural masking level differences at low and high frequencies," *J. Acoust. Soc. Am.* **101**, 1671–1680.
- van de Par, S., and Kohlrausch, A. (1998). "Diotic and dichotic detection using multiplied-noise maskers," *J. Acoust. Soc. Am.* **103**, 2100–2110.
- van de Par, S., and Kohlrausch, A. (1999). "The influence of basilar-membrane compression on binaural detection," *J. Acoust. Soc. Am.* (submitted).
- Richards, V. M., and Heller, L. M. (1991). "The detection of a tone added to a narrow band of noise: The energy model revisited," *Q. J. Exp. Psychol.* **43a**, 481–501.
- Sever, Jr., J. C., and Small, Jr., A. M. (1979). "Binaural critical masking bandwidths," *J. Acoust. Soc. Am.* **66**, 1343–1350.
- Stern, R. M., and Shear, G. D. (1996). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," *J. Acoust. Soc. Am.* **100**, 2278–2288.
- Webster, F. A. (1951). "The influence of interaural phase on masked thresholds I. The role of interaural time-deviation," *J. Acoust. Soc. Am.* **23**, 452–462.
- Zurek, P. M., and Durlach, N. I. (1987). "Masker-bandwidth dependence in homophasic and antiphase tone detection," *J. Acoust. Soc. Am.* **81**, 459–464.

# Azimuthal tuning of human perceptual channels for sound location

Susan E. Boehnke and Dennis P. Phillips

*Hearing Research Laboratory, Department of Psychology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada*

(Received 29 March 1999; revised 14 June 1999; accepted 25 June 1999)

Human sound localization is acute for frontal locations, but relatively poor in the lateral hemifields. Previous studies in man have not, however, provided evidence on the tuning of the perceptual channels for auditory space that subserve this pattern of acuity. The spatial tuning of perceptual channels used in human azimuthal sound localization was determined using a between-channel auditory temporal gap detection paradigm. In this paradigm, gap thresholds are low when the markers bounding the silent period (gap) activate the same perceptual channel but are elevated when the two markers activate different channels. To determine the tuning of spatial channels, gap thresholds were obtained in an anechoic room with white noise markers coming from each combination of 12 leading marker locations and 18 trailing marker locations throughout the full 360° of azimuth in the horizontal plane through the interaural axis. Gap thresholds remained low (2–4 ms) for all combinations of leading and trailing markers between 30° and 150° in both lateral hemifields. When the leading marker was located deep in one hemifield, and the trailing marker was in the opposite hemifield, gap thresholds rose to 8–16 ms. For leading marker locations at 30° from the midline, gap thresholds were low for all trailing marker locations in the ipsilateral hemifield and locations near the midline in the contralateral hemifield, and were elevated (6–8 ms) only near the contralateral pole. Finally, for leading marker locations at 0° or 180°, gap thresholds were low for any trailing location within 30° of the midline at the front or back, and thresholds were elevated for trailing locations at the lateral poles. These data are accountable in terms of two broadly tuned perceptual channels, occupying the left and right auditory hemifields, respectively, each extending 30° across the midline. These channels have widths and locations similar to the spatial receptive fields previously described for central auditory neurons in animals. The data suggest a model of spatial acuity based on the rates of activation of two spatially overlapping channels, rather than the selective activation of members of a large population of finely tuned channels. © 1999 Acoustical Society of America. [S0001-4966(99)03210-5]

PACS numbers: 43.66.Qp, 43.66.Mk, 43.66.Lk [DWG]

## INTRODUCTION

One of the challenges in sensory neuroscience is to link the properties of neuronal receptive fields with the properties of the perceptual dimensions they underlie (after Teller, 1984). The following report describes the architecture of the perceptual mechanisms used for the human spatial localization of sound, and reveals those mechanisms to have properties comparable to those seen in the central auditory system of other mammals.

Behavioral lesion studies have demonstrated that the primary auditory sensory cortex (AI) of each cerebral hemisphere is required for the accurate localization of sounds in the contralateral hemifield of auditory space (Jenkins and Masterton, 1982; Jenkins and Merzenich, 1984; Kavanagh and Kelly, 1987; Heffner, 1997). Localization experiments with macaques that had unilateral or bilateral lesions to the auditory cortex demonstrated that each cortex contains a representation of contralateral auditory space which also extends 20 deg or so into the auditory hemifield ipsilateral to the ablation (Heffner, 1997). Neurophysiological recordings of single neurons in AI of animals have revealed relatively broad azimuthal tuning in most cells, especially at suprathreshold levels of stimulation. Of the neurons with spatially

restricted receptive fields, most of those receptive fields occupy much of the contralateral auditory hemifield, and have boundaries near the midsagittal plane (Middlebrooks and Pettigrew, 1981; Phillips and Brugge, 1985; Imig *et al.*, 1990; Rajan *et al.*, 1990; Clarey *et al.*, 1995; Brugge *et al.*, 1996), a finding that extends to nonprimary auditory cortex (Middlebrooks *et al.*, 1998) and to the tectum (e.g., Semple *et al.*, 1983; Wise and Irvine, 1985).

The stimulus information available for determining the azimuthal location of a sound source is the set of disparities in the timing and amplitude of the sounds as they arrive at the two ears (Middlebrooks and Green, 1991; Wightman and Kistler, 1992; Middlebrooks, 1992). The azimuthal extents of spatial receptive fields are probably based on neural sensitivity to these interaural disparities. However, despite the knowledge that human listeners have greatest azimuthal sound localization acuity for sources near the midsagittal plane (Mills, 1972; Middlebrooks and Green, 1991; Recanzone *et al.*, 1998), there has been no demonstration of the spatial tuning of the perceptual channels that mediate human localization performance.

We have exploited a recent innovation in human auditory gap detection paradigms to address this issue. Gap de-

tection refers to the ability to detect a brief silent interval in an otherwise homogeneous stream of sound (Eddins and Green, 1995). If the sounds that mark the beginning and end of the silent period are the same (i.e., if they activate the same perceptual channel), then the minimum detectable gap can be as short as a few ms (Plomp, 1964; Penner, 1977; Eddins and Green, 1995; Phillips *et al.*, 1997, 1998; Formby *et al.*, 1998; Chatterjee *et al.*, 1998; Taylor *et al.*, 1999). If, however, the gap's markers are different (i.e., if the markers activate different perceptual channels), then the shortest detectable gap (gap threshold) can be lengthened to 10–50 ms. This acuity difference likely reflects the difference between that of a discontinuity detection mechanism in the within-channel case, and a mechanism for the relative timing of activity in different channels in the between-channel case (Phillips *et al.*, 1997; Formby *et al.*, 1998; Taylor *et al.*, 1999).

In the spatial domain, it is known that gap thresholds are low (2–3 ms) when white noise markers come from the same location (directly opposite a given ear), but are highly elevated (10–30 ms) when the two white noise markers come from opposite hemifields, indicating that the latter markers were activating different central perceptual channels for spatial location (Phillips *et al.*, 1998). In the present study, it was hypothesized that by systematically varying the location of the leading and trailing markers, we could determine the azimuthal tuning of these perceptual channels. That is, low gap thresholds would indicate that the two marker locations were activating the same perceptual channel and asymptotically high gap thresholds would indicate the two marker locations were activating different perceptual channels (see Phillips *et al.*, 1997, 1998; Formby *et al.*, 1998).

## I. METHOD

### A. Subjects

Data presented are from one highly trained and two relatively untrained female listeners (one of them S.E.B.) who provided data for locations throughout the 360° of azimuth. An additional four listeners (two untrained males, one trained male, one trained female), whose data are not presented in what follows, were tested throughout the frontal 180° only (–90° to +90°). The data from these subjects were completely consistent with those of the three more intensely studied listeners. The listeners ranged in age from 20 to 32 years with a mean age of 26 years. The three trained listeners had significant experience in auditory psychophysical studies and had participated in previous studies of gap detection by Phillips *et al.* (1997, 1998). The untrained listeners had little or no experience in auditory perception studies. However, all were musically trained and the two males were professional musicians. All of the listeners had free-field noise detection thresholds within laboratory norms. The purpose of the experiment was known to all participants and all except the author were paid for their participation.

### B. Gap stimuli and apparatus

Stimulus presentation and data acquisition were controlled by a custom-programmed Macintosh IIfx computer.

All stimuli were wide-band (20 kHz) noise bursts, generated by a white-noise source (Coulbourn Instruments White Noise Generator–S81-02). All rise–fall times used to shape the stimuli were linear and of 0.5-ms duration. The gap durations are defined as the length of the silent period excluding the rise–fall times. The gated stimuli were passed through a passive attenuator (Hewlett Packard 350D attenuator set), a power amplifier (Realistic SA-155 integrated stereo amplifier), and were transduced by two Optimus speakers (XTS 36 40-1994) with a frequency response of 0.14–20.0 kHz. Each stimulus consisted of a leading marker that was 10 ms in duration and a trailing marker that was 300 ms in duration. This leading marker duration was used because it is short enough to enable visualization of a robust between-channel effect (Phillips *et al.*, 1998), yet is long enough to result in a manageable task for participants. The two speakers were suspended at approximately ear level by inflexible metal rods from a 70-in. diameter circular track mounted on the ceiling of an Eckel sound-attenuating room (95-in. wide, 83-in. deep, and 80-in. high internal measurements). The ceiling and walls of this booth had been lined with 1-in. thick acoustic foam. This foam had 0.5-in. deep (egg carton) ripples to improve sound absorption. A vertical yellow line (1-in. width) was placed through zero azimuth to act as a visual fixation point in an attempt to control for gross movements of the head.

## C. Procedure

### 1. Sensation level determination

Each participant was seated below the center of the circular track from which the speakers were hung, and was therefore in the direct field of the sound. Noise detection thresholds were obtained by the method of limits using 300-ms noise bursts from a speaker located directly ahead at 0° azimuth. This was repeated for both speakers used in the experiment and there was no difference between speakers in the thresholds obtained through them. For each subject, all experimental trials were presented at 35-dB sensation level. Although the detection threshold is somewhat lower (about 3 dB) when the speaker location is in the lateral positions (90° azimuth), a previous study in which small loudness differences between markers were compensated for suggested that the compensation was inconsequential to the patterns of data obtained (Phillips *et al.*, 1997). Thus stimulus sensation level was always based on thresholds obtained at the 0° azimuth location.

### 2. The gap detection task

Each stimulus trial consisted of two sequences of noise bursts, usually separated by an interval of 300 ms (see below). (One untrained subject used a 500-ms interval.) One of the two sequences of noise—the *signal*—contained a detectable silent period (gap) between the leading and trailing marker. The other sequence was the *standard*, which included a 1.0-ms gap between the leading and trailing marker. This 1-ms gap was below the gap-detection threshold for each listener and was present to control for gating transients that were present in the signal sequence. In both sequences,



the leading marker was 10 ms and the trailing marker was 300 ms in duration, and the two sequences came from the same locations in any given trial. In conditions in which the leading and trailing marker were located at the same eccentricity, a single speaker was used.

In the trials for each gap threshold determination, the locations of the leading and trailing markers were held constant. On each trial, the order of the standard and signal presentation was randomized by the computer and the task of the subject was to indicate by button press whether the signal stimulus came first or second. No feedback was provided. Gap thresholds for every stimulus condition were obtained in blocks of two-interval, two-alternative forced-choice trials. The inter-trial interval was typically 300–500 ms plus a 500-ms warning light. The listener tracked the gap duration across the block of trials in a two-down/one-up adaptive paradigm (after Levitt, 1971). The starting duration of the gap in the signal stimulus was long, around 20–30 ms for within-channel conditions and 30–60 ms for between-channel conditions. After every two successive correct responses, the length of the gap was decreased by a factor of 1.2, and after every incorrect response, the length of the gap was increased by the same factor. This two-down/one-up procedure was continued until there had been 14 reversals in the direction of change of the gap duration. The gap threshold for that block of trials was taken as the mean gap duration across the last 10 trials that produced reversals in the adaptive step. Listeners typically took 40–80 trials to reach this criterion.

There were 12 locations of the leading marker ( $0^\circ$ ,  $\pm 30^\circ$ ,  $\pm 60^\circ$ ,  $\pm 90^\circ$ ,  $\pm 120^\circ$ ,  $\pm 150^\circ$ , and  $180^\circ$ ), and for each leading marker condition gap thresholds were obtained with trailing markers at each of 18 locations ( $0^\circ$ ,  $\pm 15^\circ$ ,  $\pm 30^\circ$ ,  $\pm 45^\circ$ ,  $\pm 60^\circ$ ,  $\pm 90^\circ$ ,  $\pm 120^\circ$ ,  $\pm 150^\circ$ ,  $\pm 165^\circ$ , and  $180^\circ$ ). Trailing marker locations were more closely spaced near the midline because it was expected that gap thresholds would be more sensitive to changes in trailing marker near the midline. Gap-detection thresholds were obtained for each condition for each subject (fully repeated measures). The conditions with frontally located leading markers ( $0^\circ$  to  $\pm 90^\circ$ ) were completed first and studied in random order. The rear leading-marker conditions ( $\pm 120^\circ$ ,  $\pm 150^\circ$ ,  $180^\circ$ ) were completed later, also in random order. The randomization procedure across all conditions was repeated at least three times and the subjects' final gap threshold for each condition was the mean of the three last repetitions of each condition. The order of trailing conditions within a leading condition began at a random location and progressed, in serial order, around the azimuth. The direction of progression was usually alternated across repetitions, and no effect of trailing marker order was observed. Full randomization of conditions was not done because of the importance of stimulus certainty in obtaining the best possible performance. While there were no systematic order effects, there was a general practice effect that was distributed randomly across conditions and will be discussed later.

## II. RESULTS

### A. Dependence of gap threshold on leading and trailing marker azimuth

Functions relating gap threshold to trailing marker location were obtained for each leading marker condition. There was no laterality effect between any leading marker conditions ( $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$ ,  $150^\circ$ ) at the same spatial location in opposite hemifields. The left leading function for each location was flipped to coincide with the right and no pair interacted statistically ( $p > 0.13$  for each comparison) indicating the left and right functions for any given leading marker location did not differ. The data for the  $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$  and  $150^\circ$  leading-marker conditions were therefore collapsed across the hemifield and all further analyses were completed using these collapsed data.

The effect of trailing marker on gap thresholds was different for different leading marker conditions. These differences between curves were demonstrated statistically by an interaction between leading and trailing markers in a  $7 \times 18$  fully repeated measures analysis of variance (ANOVA) [ $F(102,204) = 6.95$ ,  $p < 0.0002$ ] on the collapsed data and the  $0^\circ$  and  $180^\circ$  leading marker conditions. For this and all subsequent tests, the H-F correction (Huynh and Feldt, 1976) for violations of compound symmetry in repeated measures designs was used, unless otherwise specified.

Figures 1–3 present the gap threshold mean data for the three listeners. Each figure presents the data in two forms. In Figs. 1(A), 2(A), and 3(A), gap threshold has been plotted as a function of trailing marker location, separately for the leading marker locations indicated. In Figs. 1(B), 2(B), and 3(B), gap threshold is illustrated in polar plots, in which gap thresholds (in ms) are plotted along the radius, as a function of trailing marker location, separately for the specified leading marker locations. The three groupings of these functions are based on their visual and statistical similarity. The curves for  $60^\circ$ ,  $90^\circ$ , and  $120^\circ$  leading conditions (Fig. 1) did not interact when tested separately [ $F(34,68) < 1$ , ns], indicating that they did not differ significantly. The two midline-leading conditions,  $0^\circ$  and  $180^\circ$ , also did not interact [ $F(17,34) = 1.09$ , ns] indicating they did not differ either in their dependence of gap threshold on trailing marker location (Fig. 2). Finally, the locations  $30^\circ$  from the midline,  $30^\circ$  and  $150^\circ$ , did not interact [ $F(17,34) = 3.31$ ,  $p = 0.15$ ] and thus did not differ (Fig. 3). In this final test, the Greenhouse and Geisser (1959) correction had to be used due to an overly elevated epsilon value. The interaction observed in the omnibus ANOVA was between these three sets of groups. Error bars shown in Figs. 1–3 are the standard error of the three repetitions for each condition calculated separately for each subject and averaged. This measure avoided the inclusion of between-subject magnitude differences which were inconsequential to the pattern of results. In practice, the three listeners had similar absolute within-channel gap thresholds (often in the range from 2.1 to 4.0 ms), but their between-channel thresholds had ceilings that differed, being in the ranges from 8 to 9 ms, 9 to 10 ms, and 12 to 16 ms, respectively.

In Fig. 1(A), data are shown for stimulus conditions with leading markers at  $60^\circ$ ,  $90^\circ$  and  $120^\circ$ . It is apparent that both

### Azimuthal Functions for Leading Markers Deep in the Lateral Hemifields

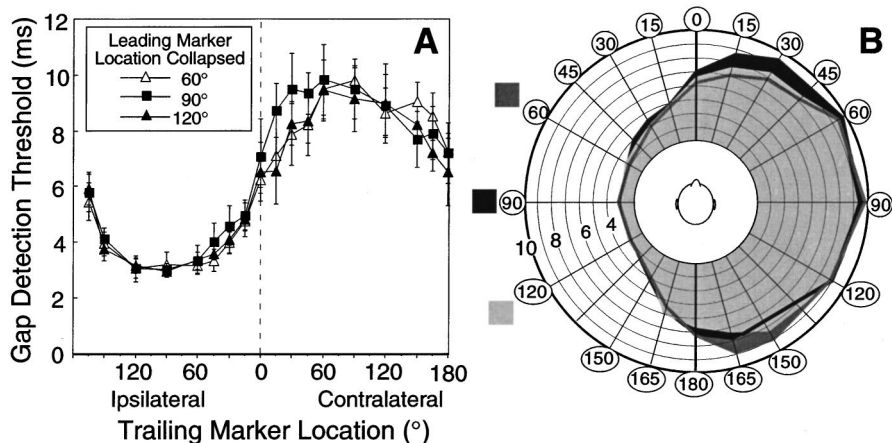


FIG. 1. Dependence of gap threshold on trailing marker location, for each of three leading marker locations deep in the lateral hemifields. (A) Data are presented as a conventional  $x$ - $y$  plot, with the parameter being the location of the leading marker. Values plotted are the means of three threshold measurements averaged across each of the three listeners, i.e., a mean of nine thresholds. Error bars are the standard error of the three repetitions for each condition calculated separately for each subject and averaged. (B) Same data presented as a polar plot, in which gap threshold (in ms) plotted along the radius, and separate plots for data obtained with the three different leading markers. Gradations along the radius are steps of 2.0 ms. Shading in the polar plot identifies the plots for the data for each leading marker location (shown as shaded rectangles to left of plot).

the absolute gap thresholds, and the pattern of dependence of gap threshold on the location of the trailing marker, were virtually identical. Gap thresholds were low (“within-channel”) when the markers were in the same hemifield, and became elevated when the trailing marker was within about 30° of the midline, reaching asymptotically high values when the trailing marker was 30° or more into the contralateral hemifield. The coincident spatial boundaries associated with low gap thresholds, in the face of widely separated leading markers (60°, 90°, 120°), indicates that these three leading markers are within a single perceptual channel, and that this channel has borders extending 30° into the contralateral hemifield.

Figure 2(A) and (B) shows the data obtained when the leading markers were in the midsagittal plane (not collapsed). For both the 0° and 180° leading markers, gap thresholds were at minima when the trailing marker was within 30° of the midline, irrespective of whether the markers were in the same frontal or caudal hemifield. Gap thresholds were elevated for trailing markers in the lateral hemifields, but the absolute maximum values of the thresholds

were approximately 30% lower than those seen when leading markers were located deep in the lateral hemifields [Fig. 1(A), (B)].

Finally, Fig. 3(A) and (B) presents collapsed data obtained when the leading markers were positioned at 30° and 150°, i.e., 30 degrees from the midsagittal plane. For these conditions, gap thresholds were at minima for all trailing marker locations in the ipsilateral hemifield, and were elevated to maxima only for locations more than 30° into the contralateral hemifield. Note that this asymptotic maximum in the contralateral hemifield is lower than that seen in Fig. 1 for more lateral leading marker locations.

### B. A two-channel model for auditory space perception

The similarity of the data for the 60°, 90°, and 120° leading conditions (Fig. 1) is important. First, it indicates that it was not simply the eccentricity difference *per se* in

### Azimuthal Functions for Leading Markers on the Mid-Sagittal Plane

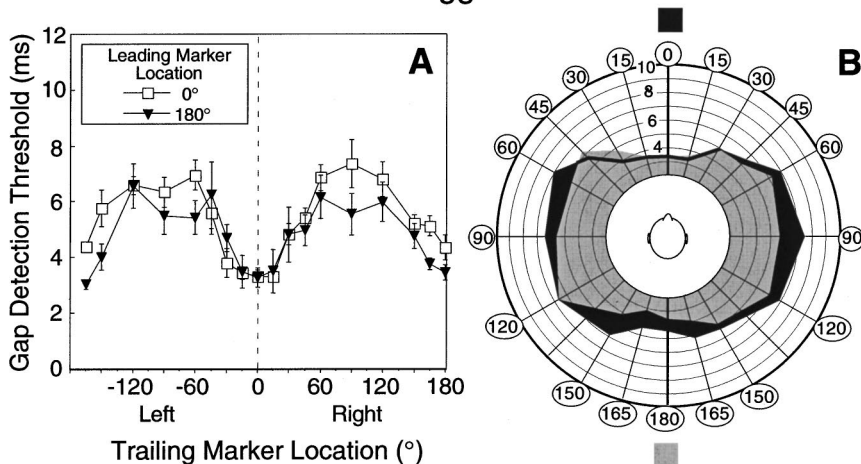


FIG. 2. Dependence of gap threshold on trailing marker location, for each of two leading marker locations in the midsagittal plane. Other details as for Fig. 1.

## Azimuthal Functions for Leading Markers 30° from the Mid-Sagittal plane

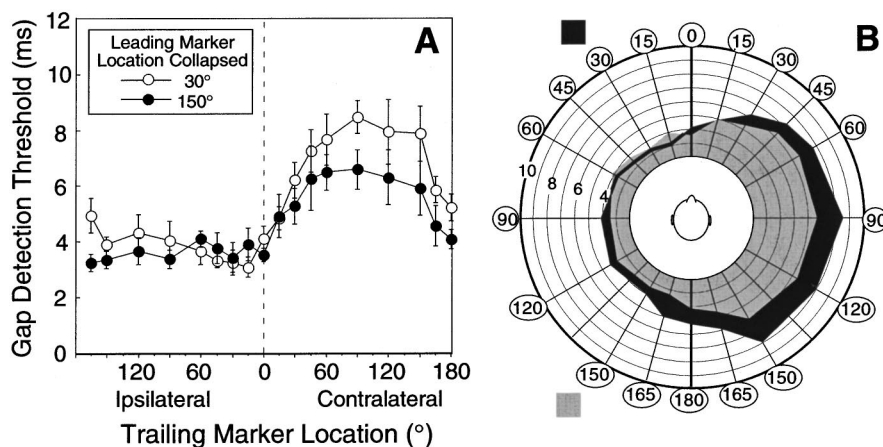


FIG. 3. Dependence of gap threshold on trailing marker location for each of two leading marker locations 30 degrees from the midsagittal plane. Other details as for Fig. 1.

marker locations which resulted in elevated gap thresholds. Second, as mentioned above, the fact that the spatial boundaries of the “within-channel” gap thresholds did not vary over a 60° variation of leading marker locations suggests that those boundaries delimit a single spatial channel for each of the left and right auditory hemifields. Because there was a spatial region associated with intermediate gap thresholds (those trailing marker locations  $\leq 30^\circ$  from the midline), the borders of such hemifield channels must extend about  $30^\circ$  into the contralateral hemifield, i.e., the two channels overlap through the midline region of space, and the region of overlap extends to the azimuth where asymptotic gap thresholds were reached.

Figure 4(A) depicts the response profiles of the two hypothetical channels, i.e., it shows a plot of the extent to which each channel would be activated by a source located anywhere in the frontal sound field (from  $90^\circ$  left to  $90^\circ$  right). The profiles, based on the data in Fig. 1, would be mirror symmetric for the caudal hemifield. Note that activation rates are high, and undifferentiated, for sources deep in the lateral hemifields. Activation rates decline in the region spanning  $\pm 30^\circ$  from the midline. Now, we have argued previously that gap thresholds in the between-channel paradigm depend on the extent to which the leading and trailing markers activate the same perceptual channels (Phillips *et al.*, 1997; Taylor *et al.*, 1999; see also Formby *et al.*, 1998). For the stimuli used in the present study, a ratio of the activation of each channel can be obtained for each combination of azimuths used.

We calculated the proportion of “representational overlap,” i.e., the extent to which the two channels were co-activated, for every stimulus pair used, and then plotted our direct measurements of gap threshold as a function of that overlap. To illustrate the percent overlap calculation, consider the case of a leading marker at  $30^\circ$  Left, and a trailing marker at  $15^\circ$  Right. From Fig. 4(A), the left channel would be activated to 90% and 25% of maximum by the leading and trailing markers, respectively (for a 25% overlap), and the right channel would be activated to 10% and 75% of maximum (10% overlap) by the same two markers. The total representational overlap was thus 35% (i.e., 25% plus 10%).

The data are shown in Fig. 4(B). We emphasize that the only data contributing to the profiles of the hypothetical channels [Fig. 4(A)] were those of Fig. 1. In contrast, the gap thresholds plotted in Fig. 4(B) include those obtained in all the stimulus conditions. It is apparent that there is a nearly linear relationship between representational overlap and gap

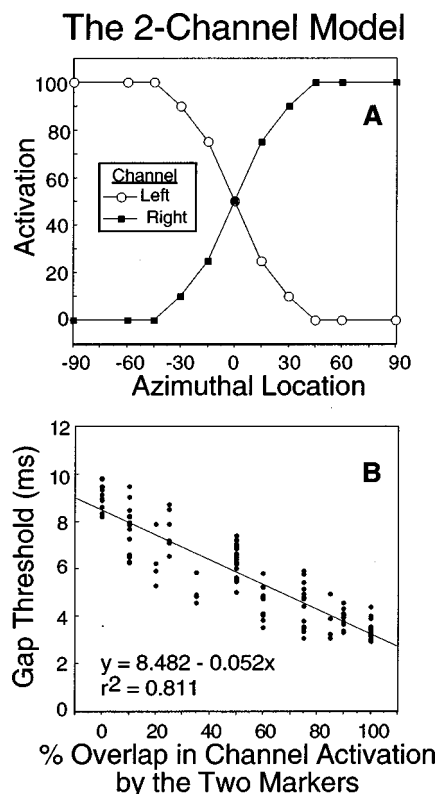


FIG. 4. (A) Model of a two-channel mechanism of sound localization. A left and a right hemifield channel respond at a relatively undifferentiated rate for a broad range of azimuths deep in their own hemifields; their activation rates decline for sources near the midline, to near zero for sources 30 deg into the opposite hemifield. Model is loosely based on the data in Fig. 1. (B) Measured gap thresholds (from Figs. 1–3) plotted as a function of the “representational overlap” of the markers, i.e., as a function of the proportion of channel activation common to both markers. Note that the relationship is linear, and that it accounts for more than 80% of the data variance.



threshold. The simple two-channel model accounted for 81% of the variance in all the data.

### III. DISCUSSION

Our working hypothesis is that when gap thresholds are at minima, the leading and trailing markers are activating the same perceptual channel (or stimulus representation), and when gap thresholds are elevated, the two markers are activating channels that are only partially overlapping. Asymptotically high gap thresholds are seen when there is minimal overlap in the neural representations of the stimuli (Phillips *et al.*, 1997; Formby *et al.*, 1998; Taylor *et al.*, 1999). The results with leading markers deep in the lateral hemifields indicate that there is probably one spatial channel serving each lateral hemifield, and that their borders extend about 30° into the contralateral side [Fig. 1(A), (B); Fig. 4(A)]. The argument is specifically for one spatial channel per hemifield, because a 60° variation in the location of the leading marker (60°, 90°, 120°) resulted in no difference in the spatial boundaries of the low gap threshold region. The overlapping borders of the left and right hemifield channels are inferred from the left–right, mirror-symmetric trailing marker locations associated with rising gap thresholds (which started at locations 30° ipsilateral, and extended 30° into the contralateral hemifield). Beyond that, gap thresholds were asymptotically high, indicating minimal representational overlap of the markers, and thus stimulation of separate spatial perceptual channels.

Patterns of gap thresholds for leading markers in the other locations are explicable in terms of the relative activations of the two hemifield channels. For a leading marker located in the midsagittal plane (Fig. 2), there will always be some representational overlap of the two markers, irrespective of where the trailing marker is located. The activation of the two channels will be most equivalent for trailing markers very close to the midline; the representational overlap will be poorest, but still existent, for trailing marker locations deep in either hemifield. It is for this reason that the functions in Fig. 2 have minima for near-midline locations of the trailing marker, and sub-asymptotic maxima for trailing markers at the lateral poles. When the leading marker is located at the extreme border of one hemifield channel (Fig. 3), trailing marker locations anywhere within the ipsilateral channel will elicit minimal gap thresholds; and sub-maximal gap thresholds are to be expected when the trailing marker is deep within the contralateral hemifield. Finally, the data are not explicable simply by the angular separation of the markers, because the pattern of gap threshold dependence on trailing marker location varies with the leading marker locus. Thus a fixed separation between the two markers (e.g., 60°) resulted in low or high gap thresholds, depending on whether the two markers were within a single channel or in different ones.

The low gap thresholds values seen for “within-channel” conditions are consistent with those observed in previous studies (Phillips *et al.*, 1997, 1998; Formby *et al.*, 1998; Taylor *et al.*, 1999). The “between-channel” conditions, where gap thresholds are elevated to asymptotic levels, i.e., trailing locations deep in the hemifield contralateral to 60°, 90°, or 120° leading markers, are in the low end of the

range of gap thresholds seen in studies of between-channel gap detection (Phillips *et al.*, 1998). This is most likely due to experience, since the listeners became highly trained through the course of the study. This practice effect is evident in the small overall improvement in performance seen for the rear leading marker conditions, data for which were obtained after all frontal locations had been completed.

A description of human sound localization mechanisms based on two broadly tuned, hemifield channels is compatible with behavior-lesion evidence in animals which indicates that each side of the auditory forebrain contains a representation of the contralateral auditory hemifield (Jenkins and Masterton, 1982; Jenkins and Merzenich, 1984; Kavanagh and Kelly, 1987; Heffner, 1997), and there is some evidence in the neurological literature that this contralaterality of spatial representation may extend to humans beings (Sanchez-Longo and Forster, 1958; Heilman and Valenstein, 1972; Zatorre *et al.*, 1995; Phillips, 1995). It gains further support from animal neurophysiological studies in cats (Middlebrooks and Pettigrew, 1981; Middlebrooks *et al.*, 1998; Imig *et al.*, 1990; Rajan *et al.*, 1990; Clarey *et al.*, 1995; Brugge *et al.*, 1996) and primates (Ahissar *et al.*, 1992; Benson *et al.*, 1981) in which it has been shown that, of the cortical neurons with spatially restricted receptive fields, most have those receptive fields centered in the contralateral auditory hemifield, their medial borders close to the midsagittal plane, and often extending into the ipsilateral hemifield. The hemifield tuning seen in the present study is also strikingly reminiscent of that of the population-vector model of sound location coding recently described by Eggemont and Mossop (1998).

To be sure, a minority of central neurons have spatial receptive fields centered on the frontal midline (Middlebrooks and Pettigrew, 1981; Wise and Irvine, 1985; Imig *et al.*, 1990; Rajan *et al.*, 1990; Clarey *et al.*, 1995), or have binaural input patterns that would underlie such spatial selectivity (Kitzes *et al.*, 1980; Phillips and Irvine, 1981). However, the present data appear to be explicable without the need to postulate the existence of a midline perceptual channel. In this regard, one independent study (Guard *et al.*, 1998) suggested that primate sound localization acuity could in principle be based on the relative activities of two (left and right) populations of neurons with hemifield spatial tuning, again without requiring the postulation of a midline channel.

A further question about the perceptual channels concerns the location in the brain of their neurophysiological substrates. The processing of stimulus information for sound localization is initiated in the auditory brainstem, but the fact that lesions restricted to the auditory cortex (e.g., Heffner and Heffner, 1990) and, in particular, lesions restricted to the primary field (Jenkins and Merzenich, 1984), are devastating to localization performance, suggests a cortical site for the neural mechanism(s) tapped in the present study. Now, since most of the cortical cells with restricted spatial receptive fields have those fields located in the contralateral auditory hemifield, it is possible that the neural underpinnings of the left and right perceptual channels reside in the right and left cerebral hemispheres, respectively. However, in each primary auditory cortex, there is also a small population of



neurons with receptive fields in ipsilateral auditory space, or which have binaural inputs that would predict ipsilateral receptive fields (Rajan *et al.*, 1990; Imig *et al.*, 1990; Phillips and Irvine, 1981). This means that it is possible that each of the two perceptual channels has substrate in both brain hemispheres.

The present data offer a link between the azimuthal tuning of central neurons on the one hand, and behavioral azimuthal acuity on the other. It does so by describing the azimuthal tuning of the perceptual channels used in sound localization. By the present account, the high behavioral spatial acuity seen for near-midline sources (Mills, 1972; Middlebrooks and Green, 1991; Recanzone *et al.*, 1998) is based on the unique rates of activity in one or both of two broadly tuned spatial channels for each stimulus location within that eccentricity range. In principle, this is not unlike the case of color vision in which three broadly tuned receptor systems provide a basis for the discrimination of thousands or even millions of colors (see Goldstein, 1989). This kind of system is in marked contrast to that previously described for the auditory nervous systems of barn owls (Knudsen and Konishi, 1978); in that species, superior localization acuity in a restricted region of auditory space is due to the presence of large numbers of neurons with very small, focal receptive fields in that region.

In contrast, measures of behavioral sound localization acuity in humans have consistently shown poor localization acuity for sources deep in the lateral hemifields. The case is worst for tonal stimuli, which constitute a limiting case for the poverty of information available to the listener (see Mills, 1972; Middlebrooks and Green, 1991). For noise stimuli, minimum audible angles at small eccentricities are in the range from 1.3°–1.8°, but closer to 9°–10° for eccentricities of 90° (Heffner and Heffner, 1988). Variability of localization judgements follow the same pattern (Makous and Middlebrooks, 1990; Recanzone *et al.*, 1998). This is explicable by the present model in terms of the ambiguity associated with similar activation of a single perceptual channel for locations across at least a 60° span of azimuth (60°–120°; Fig. 1). That is, sources widely separated but deep in the lateral hemifields activate the same perceptual channels in a relatively undifferentiated fashion, giving rise to relatively undifferentiated spatial percepts and poor localization acuity.

The present hypothesis would also predict good localization acuity for spatial locations near the rear midline. In practice, this is not always observed (Oldfield and Parker, 1984; Makous and Middlebrooks, 1990; Middlebrooks, 1992; Carlile *et al.*, 1997). However, typically these studies have required a different behavior to make frontal localization judgements than rear judgments, i.e., subjects were required to turn around to make their response to rear location, thus changing their position from that in which the stimulus was encoded. One previous study using a different methodology showed superior localization acuity for rear locations compared with lateral locations (Preibisch-Effenberger, 1966). The two-channel model also predicts the front–back confusions in localization that are sometimes seen in human localization studies (around 6% of localization judgements: Makous and Middlebrooks, 1990); those front and back locations

result in equal ratios of channel activation (e.g., 15° and 165°). The present account has no clear explanation of why front–back reversals are not more common than they are, or why some level of localization ability is possible deep in the lateral hemifields at all (see above). The listeners in our study reported that they readily formed veridical percepts of front and rear locations during the gap detection task. This is perhaps surprising since our two-channel model itself provides no basis for distinguishing front and back. Presumably, therefore, performance in the present gap-detection paradigm reflects processing at a level prior to the incorporation of the spectral cues that would otherwise resolve the front–back confusions and enable some level of azimuthal discriminations deep in the lateral hemifields.

The present findings also provide insight into other auditory spatial phenomena. First, cocktail party scenarios, in which auditory targets are easier to isolate if they are in different auditory hemifields than auditory distractors (Cherry, 1953; Yost *et al.*, 1996), are understandable because the perceptual or attentive systems are provided with the opportunity to exploit the spatial cue if the stimuli are in different hemifields, and thus are activating different spatial perceptual channels. Second, spatial masking phenomena are likely most marked when the signal and masker(s) are in the same auditory hemifield, because there is no spatial channel difference to aid the perceptual processor in the isolation of the stimuli (Kidd *et al.*, 1998). Further implications of this work extend into the clinical realm. For example, in the school classroom, those children who have perceptual or attentive difficulty with speech discrimination in noise may be able to have that disadvantage partially offset by a seating arrangement that places the instructor and the competing noise source in different auditory hemifields. The importance of the present study is in its description of a perceptual architecture that mediates between the spatial receptive field properties of central neurons and the more high-level perceptual or cognitive phenomena.

## ACKNOWLEDGMENTS

This research was supported grants from the Natural Sciences and Engineering Research Council of Canada to D.P.P. Special thanks are due to Susan Hall for technical assistance, and to Susan Hall, Raymond Klein, Glenn Schellenberg and David Shore for helpful discussions through the course of this work. Thanks also to Rickye Heffner and to two anonymous reviewers for helpful comments on a previous version of this article.

Ahissar, M., Ahissar, E., Bergman, H., and Vaadia, E. (1992). "Encoding of sound source locations and movement: Activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex." *J. Neurophysiol.* **67**, 203–215.

Benson, D. A., Hienz, R. D., and Goldstein, Jr., M. H. (1981). "Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: Spatial tuning and behavioral dependency." *Brain Res.* **219**, 249–267.

Brugge, J. F., Reale, R. A., and Hind, J. E. (1996). "The structure of spatial receptive fields of neurons in primary auditory cortex of the cat." *J. Neurosci.* **16**, 4420–4437.

- Carlisle, S., Leong, P., and Hyams, S. (1997). "The nature and distribution of errors in sound localization by human listeners," *Hearing Res.* **114**, 179–196.
- Chatterjee, M., Fu, Q-L., and Shannon, R. V. (1998). "Within-channel gap detection using dissimilar markers in cochlear implant listeners," *J. Acoust. Soc. Am.* **103**, 2515–2519.
- Cherry, E. C. (1953). "Some experiments upon the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **25**, 975–979.
- Clarey, J. C., Barone, P., Irons, W. A., Samson, F. R., and Imig, T. J. (1995). "Comparison of noise and tone azimuth tuning of neurons in cat primary auditory cortex and medial geniculate body," *J. Neurophysiol.* **74**, 961–980.
- Eddins, D. A., and Green, D. M. (1995). "Temporal integration and temporal resolution," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego, CA), pp. 207–242.
- Eggermont, J. J., and Mossop, J. E. (1998). "Azimuth coding in primary auditory cortex of the cat. I. Spike synchrony versus spike count representations," *J. Neurophysiol.* **80**, 2133–2150.
- Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (1998). "Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection," *J. Acoust. Soc. Am.* **103**, 3554–3560.
- Goldstein, E. B. (1989). *Sensation and Perception*, 3rd ed. (Wadsworth, Belmont, CA).
- Greenhouse, S. W., and Geisser, S. (1959). "On methods in the analysis of profile data," *Psychometrica* **24**, 95–112.
- Guard, D. C., Phan, M. L., Geiger, P. G., and Recanzone, G. H. (1998). "A population activity code of acoustic space in the primary auditory cortex (AI) of the macaque monkey," *Soc. Neurosci. Abstr.* **24**, 1402 (Abstract # 553-8).
- Heilman, K. M., and Valenstein, E. (1972). "Auditory neglect in man," *Arch. Neurol.* **26**, 32–35.
- Heffner, H. E. (1997). "The role of the macaque auditory cortex in sound localization," *Acta Otolaryngol. Suppl.* **532**, 22–27.
- Heffner, H. E., and Heffner, R. S. (1990). "Effect of bilateral auditory cortex lesions on sound localization in Japanese macaques," *J. Neurophysiol.* **64**, 915–931.
- Heffner, R. S., and Heffner, H. E. (1988). "Sound localization acuity in the cat: Effect of azimuth, signal duration, and test procedure," *Hearing Res.* **36**, 221–232.
- Huynh, H., and Feldt, L. S. (1976). "Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split plot designs," *J. Educat. Statist.* **1**, 964–973.
- Imig, T. J., Irons, W. A., and Samson, F. R. (1990). "Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex," *J. Neurophysiol.* **63**, 1448–1466.
- Jenkins, W. M., and Masterton, R. B. (1982). "Sound localization: Effects of unilateral lesions in the central auditory system," *J. Neurophysiol.* **47**, 987–1016.
- Jenkins, W. M., and Merzenich, M. M. (1984). "Role of cat primary auditory cortex for sound localization behavior," *J. Neurophysiol.* **52**, 819–847.
- Kavanagh, G. L., and Kelly, J. B. (1987). "Contribution of auditory cortex to sound localization by the ferret (*Mustela putorius*)," *J. Neurophysiol.* **57**, 1746–1766.
- Kidd, Jr., G., Mason, C. R., Rothla, T. L., and Deliwala, P. S. (1998). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **104**, 422–431.
- Kitzes, L. M., Wrege, K. S., and Cassady, J. M. (1980). "Patterns of response of cortical cells to binaural stimulation," *J. Comp. Neurol.* **192**, 455–472.
- Knudsen, E. I., and Konishi, M. (1978). "Space and frequency are represented separately in auditory midbrain of the owl," *J. Neurophysiol.* **41**, 870–884.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Makous, J. C., and Middlebrooks, J. C. (1990). "Two-dimensional sound localization by human listeners," *J. Acoust. Soc. Am.* **87**, 2188–2200.
- Middlebrooks, J. C. (1992). "Narrow-band sound localization related to external ear acoustics," *J. Acoust. Soc. Am.* **92**, 2607–2624.
- Middlebrooks, J. C., and Green, D. M. (1991). "Sound localization by human listeners," *Annu. Rev. Psychol.* **42**, 135–159.
- Middlebrooks, J. C., and Pettigrew, J. D. (1981). "Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location," *J. Neurosci.* **1**, 107–120.
- Middlebrooks, J. C., Xu, L., Eddins, A. C., and Green, D. M. (1998). "Codes for sound-source location in nontopographic auditory cortex," *J. Neurophysiol.* **80**, 863–881.
- Mills, A. W. (1972). "Auditory localization," in *Foundations of Modern Auditory Theory, Volume 2*, edited by J. V. Tobias (Academic, New York), pp. 303–348.
- Oldfield, S. R., and Parker, S. P. A. (1984). "Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions," *Percept.* **13**, 581–600.
- Penner, M. J. (1977). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," *J. Acoust. Soc. Am.* **61**, 552–557.
- Phillips D. P. (1995). "Central auditory processing: A view from auditory neuroscience," *Am. J. Otolaryngol.* **16**, 338–352.
- Phillips, D. P., and Brugge, J. F. (1985). "Progress in neurophysiology of sound localization," *Annu. Rev. Psychol.* **36**, 245–274.
- Phillips, D. P., Hall, S. E., Harrington, I. A., and Taylor, T. L. (1998). "'Central' auditory gap detection: A spatial case," *J. Acoust. Soc. Am.* **103**, 2064–2068.
- Phillips, D. P., and Irvine, D. R. F. (1981). "Responses of single neurons in physiologically defined area AI of cat cerebral cortex: Sensitivity to interaural intensity differences," *Hearing Res.* **4**, 299–307.
- Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (1997). "Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," *J. Acoust. Soc. Am.* **101**, 3694–3705.
- Plomp, R. (1964). "Rate of decay of auditory sensation," *J. Acoust. Soc. Am.* **36**, 277–282.
- Preibisch-Effenberger, R. (1966). "On the techniques of directional audiometry: Tests of the faculty of sound localization using an electronic delay line or by measurements in a free sound field," *Arch. Klin. Expt. Ohren-Nasen-Kehlkopf.* **187**, 588–592.
- Rajan, R., Aitkin, L. M., Irvine, D. R. F., and McKay, J. (1990). "Azimuthal sensitivity of neurons in primary auditory cortex of cats. I. Types of sensitivity and effects of variations in stimulus parameters," *J. Neurophysiol.* **64**, 872–887.
- Recanzone, G. H., Makhamra, S. D. D. R., and Guard, D. C. (1998). "Comparison of relative and absolute sound localization ability in humans," *J. Acoust. Soc. Am.* **103**, 1085–1097.
- Sanchez-Longo, L. P., and Forster, F. M. (1958). "Clinical significance of impairment of sound localization," *Neurology* **8**, 119–125.
- Semple, M. N., Aitkin, L. M., Calford, M. B., Pettigrew, J. D., and Phillips, D. P. (1983). "Spatial receptive fields in the cat inferior colliculus," *Hearing Res.* **10**, 203–215.
- Taylor, T. L., Hall, S. E., Boehnke, S. E., and Phillips, D. P. (1999). "Additivity of perceptual channel-crossing effects in auditory gap detection," *J. Acoust. Soc. Am.* **105**, 563–566.
- Teller, D. Y. (1984). "Linking propositions," *Vision Res.* **24**, 1233–1246.
- 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.
- Wise, L. Z., and Irvine, D. R. F. (1985). "Topographic organization of interaural intensity difference sensitivity in the deep layers of cat superior colliculus: Implications for auditory spatial representation," *J. Neurophysiol.* **54**, 185–211.
- Yost, W. A., Dye, R. H., and Sheft, S. (1996). "A simulated 'cocktail party' with up to three sound sources," *Percept. Psychophys.* **58**, 1026–1036.
- Zatorre, R. J., Ptito, A., and Villemure, J. G. (1995). "Preserved auditory spatial localization following cerebral hemispherectomy," *Brain* **118**, 879–889.

# Auditory localization of nearby sources. II. Localization of a broadband source

Douglas S. Brungart,<sup>a)</sup> Nathaniel I. Durlach, and William M. Rabinowitz<sup>b)</sup>  
*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge,  
Massachusetts 02139*

(Received 28 October 1998; revised 29 March 1999; accepted 3 July 1999)

Although many researchers have examined auditory localization for relatively distant sound sources, little is known about the spatial perception of nearby sources. In the region within 1 m of a listener's head, defined as the "proximal region," the interaural level difference increases dramatically as the source approaches the head, while the interaural time delay is roughly independent of distance. An experiment has been performed to evaluate proximal-region localization performance. An auditory point source was moved to a random position within 1 m of the subject's head, and the subject responded by pointing to the perceived location of the sound with an electromagnetic position sensor. The overall angular error ( $17^\circ$ ) was roughly comparable to previously measured results in distal-region experiments. Azimuth error increased slightly as the sound source approached the head, but elevation performance was essentially independent of source distance. Distance localization performance was generally better than has been reported in distal-region experiments and was strongly dependent on azimuth, with the stimulus-response correlation ranging from 0.85 to the side of the head to less than 0.4 in the median plane. The results suggest that the enlarged binaural difference cues found in the head-related transfer function (HRTF) for nearby sources are important to auditory distance perception in the proximal region. © 1999 Acoustical Society of America. [S0001-4966(99)04310-6]

PACS numbers: 43.66.Qp, 43.66.Pn [DWG]

## INTRODUCTION

Although human sound localization has been studied extensively in the past century, little is known about the spatial perception of nearby sources. The majority of experiments examining directional sound localization have been conducted at distances greater than 1 meter. In this region, the overall amplitude of the sound reaching the ears varies with distance, but the binaural and spectral cues that are used for directional localization are roughly independent of distance. At distances less than 1 m, however, there are important distance-dependent changes in the binaural and spectral characteristics of the sound reaching the ears. It is possible that these systematic changes allow listeners to make accurate judgments about source distance for nearby sources. Since nearly all of the perceptually relevant distance-dependent changes in auditory localization cues occur at distances less than 1 m, we will define this region as the "proximal region," and the region at distances greater than 1 m as the "distal region."<sup>1</sup>

This study examines localization accuracy in the proximal region in azimuth, elevation, and distance, and attempts to relate the findings to the proximal-region head-related transfer function.

## I. BACKGROUND

The basic mechanisms of directional sound localization are well documented. In the horizontal plane, interaural dif-

ference cues have long been recognized as the dominant localization cues. Lord Rayleigh, in his famous "duplex theory" (1907), observed that interaural time differences (ITDs) and interaural level differences (ILDs) provide salient information about the lateral position of a sound source. According to the duplex theory, ITDs dominate low-frequency sound localization, while ILDs dominate high-frequency sound localization. The ITD and ILD are important localization cues, but they cannot distinguish between sources located in the so-called "cones-of-confusion," where the interaural difference cues are constant, without exploratory head motions (Wallach, 1939; Perrett and Noble, 1997). Additional information is provided by the complex geometry of the pinnae, which filter the sound reaching the ear with a directionally dependent transfer function at high frequencies (above approximately 4 kHz). When some *a priori* information about the spectrum of the source is available, pinna filtering allows listeners to resolve front-back confusions (Musicant and Butler, 1984; Oldfield and Parker, 1986) and can provide substantial information about the azimuth of a sound source when binaural cues are completely eliminated by unilateral deafness (Slattery and Middlebrooks, 1994). Perhaps most importantly, pinnae cues allow listeners to judge the elevation of the sound sources (Roffler and Butler, 1968). All of the localization cues believed to be relevant to directional localization are included in the head-related transfer function (HRTF), which is the transfer function from a sound source to the eardrums of the listener. The HRTF includes the effects of diffraction by the head, neck, and torso, as well as the spectral shaping by the pinna.

The mechanisms that allow listeners to determine the

<sup>a)</sup>Currently at Human Effectiveness Directorate, Air Force Research Laboratory. Electronic mail: dbrungart@falcon.al.wpafb.af.mil

<sup>b)</sup>Currently at Bose Corporation.



distance of a sound source are less understood than those that allow directional localization. The most salient auditory distance cue under most conditions is the amplitude cue: the pressure of a spherically radiating sound wave is inversely proportional to the distance from the source. Spectral cues also play a role. Atmospheric absorption effectively low-pass filters sounds that propagate great distances, and low-frequency sounds propagate more effectively than high-frequency sounds around obstacles in a room. Both of these effects tend to cause more distant sound sources to appear low-pass filtered relative to closer sound sources, and may provide a spectral distance cue (Little, Mershon, and Cox, 1992). Amplitude and spectral-based distance cues are sufficient for judging changes in the relative distance of a sound source, but can only be used to make absolute distance judgments when the listener has *a priori* knowledge about the characteristics of the source. The ratio of direct to reverberant energy has been proposed as a possible absolute distance cue for localization in rooms (Mershon and King, 1975; Lounsbury and Butler, 1979; Butler, Levy, and Neff, 1980), and distance judgments in a reverberant environment are mildly correlated with source distance (Mershon and Bowers, 1979). Under free-field conditions with an unfamiliar source, distance perception is extremely inaccurate, and several researchers have reported that distance judgments in these conditions are effectively uncorrelated with the actual source position (Coleman, 1963; Mershon and Bowers, 1979; Holt and Thurlow, 1969; Gardner, 1969). A comprehensive review of distal-region localization is provided in Middlebrooks and Green (1991).

One aspect of auditory localization that has received almost no attention is the localization of sources close to the head. As early as 1911, Stewart recognized that interaural level differences increase significantly when a source approaches within a few centimeters of the head, while the interaural time delay is roughly independent of distance (Stewart, 1911a, 1911b). Stewart modeled the head as a rigid sphere with ears at diametrically opposed locations on its surface and used theoretical predictions of the sound pressure on the surface of a sphere to predict the ILD and ITD as a function of source distance and direction. Hartley and Fry (1921) manually tabulated these values at a variety of locations, and Coleman (1963) cited the increased ILDs for nearby sources as a potential auditory distance cue in the proximal region. Brungart and Rabinowitz (1996) have published a formula for evaluating proximal-region HRTFs using a sphere model and re-examined the possible use of proximal-region ILDs as a distance cue, and Duda and Martens (1998) have measured the range dependence of the HRTF for a model of the head based on a bowling ball. Each of these studies found that interaural level differences increase dramatically when the source is near the head, while interaural time delays increase only slightly for nearby sources.

In the past year, proximal-region HRTFs have been measured with a manikin head and a compact, nondirectional acoustic point source (Brungart and Rabinowitz, 1999). Many of the features of the measured HRTFs were similar to those predicted by the rigid-sphere models of Hartley and

Fry and of Brungart and Rabinowitz. The important aspects of the measured HRTFs can be summarized as follows:

- (1) The interaural level difference increases dramatically as the source approaches the head when the source is outside the median plane. This increase occurs even at low frequencies where head shadowing is negligible in the distal region. At 500 Hz, for example, the ILD increases from 4 to 20 dB as a source at 90° decreases in distance from 1 m to within a few centimeters of the head.
- (2) The interaural time delay is roughly independent of distance in the proximal region. Although the time delay can increase by as much as 100  $\mu$ s as the source approaches the head, this increase occurs only near the interaural axis, where the ITD is large and sensitivity to changes in the ITD is low.
- (3) The magnitude of the HRTF is relatively greater at low frequencies than at high frequencies when the source is near the head. This effective low-pass filtering of proximal-region sources results from a combination of diffraction at the ipsilateral side of the head and increased head shadowing at the contralateral ear.
- (4) The high-frequency features of the HRTF that are dependent on elevation are relatively insensitive to source distance. The features of the HRTF that changed significantly with elevation were not strongly dependent on source distance.
- (5) As the sound source approaches the head, the acoustic parallax effect shifts some of the high-frequency features of the HRTF at the ipsilateral ear laterally in azimuth. For example, the high-frequency patterns in the HRTF seen at 10° azimuth for a distant source might be most similar to the high-frequency patterns in the HRTF at 45° azimuth for a closer source. This parallax is a direct result of the geometric relationship between the locations of the source, the center of the head, and the ear: the 6–8 cm displacement between the center of the head and the location of the ear has little impact on the location of the source relative to the ear at 1 m, but when the source is only 12–15 cm from the center of the head, the angle of the source relative to the ear can differ from the angle of the source relative to the center of the head by 45° or more. The HRTFs measured on the KEMAR manikin generally exhibit this shift at high frequencies.

These results indicate the existence of unique physical acoustic cues in the proximal region that should be relevant to proximal-region localization. Yet, despite the recognition that localization cues are substantially different in the proximal and distal regions, no studies in the literature have systematically measured proximal-region localization performance. The experiments described here examine auditory localization in the proximal region with a broadband source. In particular, they focus on how localization accuracy changes as a function of azimuth, elevation, and distance in the proximal region. The next section discusses the experimental setup. Directional and distance localization are discussed separately in the following two sections. The last two sections compare the perceptual results to the physical local-



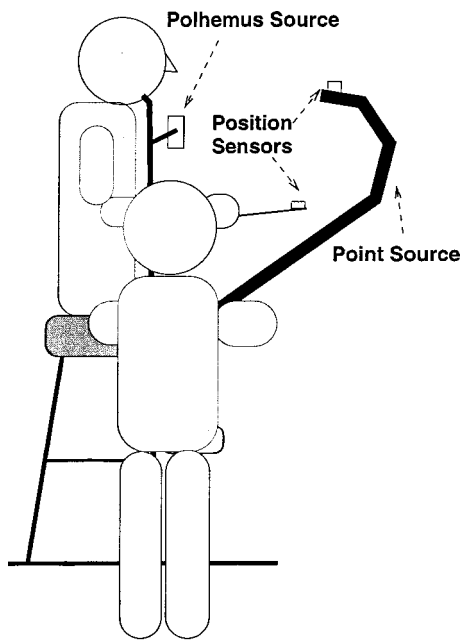


FIG. 1. Experimental setup. The experimenter stood approximately 1.5 m to the right of the listener in an anechoic chamber and manually placed the source at a random location in the subject's right hemisphere. Once the stimulus was produced, the source was moved away and the listener responded by moving a response sensor to the perceived location of the sound. See the text for details.

ization cues found in the HRTFs discussed in the first paper in the series (Brungart and Rabinowitz, 1999), and summarize the overall conclusions of this experiment.

## II. METHODS

### A. Subjects

Four right-handed male subjects, ages 20–25, participated in the experiment. Three of the subjects were paid volunteers, and the fourth was the first author. All reported normal hearing in both ears. Although three of the subjects had participated in psychoacoustic studies before, only the first author had participated in localization experiments.

### B. Apparatus

A simple diagram of the setup for the experiment is provided in Fig. 1. The experiments were conducted in MIT's anechoic chamber. The subjects were seated on a wooden stool located in the center of the chamber, and the stool was supported on the wire-frame floor by a foam-covered plywood platform. The subjects were provided with a chin rest which allowed them to immobilize their heads in a comfortable rest position during the experiments.

An experimenter, who stood approximately 1.5 m to the right of the subject, manually placed the sound source during each trial. The sound source consisted of an Electro-Voice DH1506 compression horn driver connected to 4 m of tubing with an internal diameter of 1.2 cm. The end of the tube was enclosed in a curved rigid wand constructed of PVC pipe. The curved shape allowed the stationary experimenter to place the source (the end of the tube) at any location in the right hemisphere of the subject with the opening of the

source pointing toward the subject's head. The small diameter of the tubing provided a relatively nondirectional sound source at the opening of the tube: the measured 3-dB beam-width of the source was approximately  $120^\circ$  at 15 kHz.

A Polhemus Navigation 3-Space Tracker position-sensing system measured the stimulus and response location during each trial. The electromagnetic source of the tracking system was mounted on the chin rest approximately 15 cm below and 15 cm to the left of the subject. Although the chin rest was not completely rigid, this arrangement fixed the relative positions of the Polhemus source and the subject's head, and therefore the coordinate system of the experiment was stable relative to the subject's head. One of the position sensors was mounted on the end of a 30-cm wooden rod, which the subject used to make responses. The second position sensor was mounted on the end of the experimenter's wand nearest the opening of the tube. Since it was impossible to place the sensor directly at the opening of the tube without interfering with the sound field, the orientation of the sensor and the offset between the sensor and the tube opening were used to calculate the location of the sound source on each trial. The Polhemus system is accurate within 0.25 cm in the X, Y, and Z coordinates up to approximately 1 m. In order to measure the effect of the correction on the accuracy of the location recording system, the response sensor was placed directly at the tube opening and the location of each sensor was measured by the Polhemus system. These two measurements of location differed by 2–3 cm, which can be considered an upper bound on the vector error of the system.

The control computer was a 386-based PC equipped with a 16-bit stereo sound card (Digital Audio Labs CARDD). One channel of the sound card was connected to a small ear-piece headphone worn by the experimenter. This channel was used to provide information to the experimenter during each trial. The other channel was connected to a power amplifier (Crown D-75), which was connected to the driver of the sound source. The Polhemus head tracker was connected to the PC through the RS-232 serial port, and a response switch was connected through the parallel port. The control computer automated all data recording and stimulus generation tasks in the experiment, and provided timing information to the subject and operator through its internal speaker.

### C. Stimulus

The stimuli were sequences of five rectangularly gated 150-ms pulses of noise, separated by 30-ms intervals of silence. The noise waveforms were constructed from white Gaussian noise that was filtered by a finite impulse response (FIR) filter to flatten the irregular frequency response of the point source. In addition, the noise was bandlimited to the frequency range 200 Hz–15 kHz (120 dB/decade roll-off out of band) and low-pass filtered with a 6-dB/octave roll-off above 200 Hz. This roll-off was used to maximize the non-distorted output level of the point source. Five different noise waveforms were stored on the control computer, and one

waveform was randomly chosen prior to each trial. This waveform was scaled in amplitude and then repeated five times to generate the stimuli for each trial.

The source was randomly located in the right hemisphere of the subject. Prior to each trial, the control computer read three random numbers, each ranging from 1 to 6, to the experimenter through the earphone connected to the second channel of the sound card. The experimenter used these three numbers to choose the approximate sound source location in azimuth (from near 0° for a 1 to near 180° for a 6), elevation (from near +90° for a 1 to near -90° for a 6), and distance (from 10–15 cm for a 1 to 1 m for a 6). Although the exact placement of the source varied across experimenters and some source locations were inaccessible due to interference by the subject's body or the chin-rest apparatus, this source placement system generated a reasonably broad distribution of source locations throughout the right hemisphere.

Once the source was placed, the control computer recorded the location of the source through the Polhemus tracker, and crudely normalized the amplitude of the stimulus signal to eliminate amplitude-based distance cues. The normalization was based on the distance of the source from the left and right ears of the subject. The correction normalized the amplitude so that the maximum output would occur at a distance of 1 meter. The scaling factor for this correction was

$$\frac{1}{\frac{50}{\text{Distance to left ear (cm)}} + \frac{50}{\text{Distance to right ear (cm)}}}.$$

The distance to the right ear dominates the scaling factor when the source is near the ear, but the scaling factor also considers the contribution of the left ear to perceived loudness when the source is in the median plane or is relatively distant. In addition to correction for distance, the source amplitude was randomized an additional 15 dB (from 0 to 15 dB in 1-dB steps). The amplitude scaling was accomplished by multiplying the noise waveform file by a scaling factor prior to playback. The maximum amplitude of the stimulus was approximately 59 dBA SPL (as measured by a B&K 4131 microphone) at 1 m, so with randomization and correction the effective stimulus amplitude ranged from 44–59 dBA.

#### D. Procedure

The experiment was divided into blocks of 100 trials, with each block taking approximately 20 min. At the beginning of each block of trials, the subject placed his head in a comfortable position in the chin rest and the locations of three reference points were recorded using the response sensor: the opening of the left ear canal, the opening of the right ear canal, and the tip of the nose. These locations were used to correct for stimulus distance and to define a vertical spherical coordinate system based on the subject's head, with its origin at the midpoint of the left and right ears, its horizontal plane defined by the locations of the left and right ears and the nose, and its median plane perpendicular to the interaural axis and passing as close as possible to the location

of the nose (Brungart, Rabinowitz, and Durlach, 1999). In this coordinate system, azimuth is the angle around the vertical axis, with 0° directly in front of the head, positive values in the left hemisphere, and negative values in the right hemisphere. Elevation is the angle above (positive values) or below (negative values) the horizontal plane. Note that, in this coordinate system, a one-degree change in azimuth corresponds to a shorter distance on the surface of a sphere at high and low elevations than in the horizontal plane.

Each trial was initiated when the control computer read the three source coordinates to the operator through the earpiece headphone. A beep then instructed the subject to close his eyes while the operator moved the source to the appropriate location. Once the source was positioned, the operator pressed a response switch and the control computer initiated the stimulus. First, the location of the source was read to allow for amplitude correction, then the stimulus was scaled and played through the sound source, and finally the source position was read again to verify that no movement had occurred during the stimulus presentation. If the source was stationary during stimulus presentation, the operator moved the source to a rest position and pressed the response switch again. The control computer then generated a second beep, prompting the subject to move the response sensor to the perceived location of the stimulus. The subjects were permitted to open their eyes during the response process, but usually chose not to do so. Once the subject had selected a response location, the operator once again pressed the response switch, and the control computer read the response location, generated three new coordinates for the next stimulus location, and beeped to tell the subject to close his eyes and prepare for the next stimulus. Each trial lasted approximately 12 s.

The response method used in the experiment, which we refer to as "direct location," was the method determined to be the least biased and most accurate among a number of three-dimensional proximal-region response methods considered in an earlier study (Brungart *et al.*, 1999). Using the direct-location method to identify the position of a visual target in the front hemisphere, the mean angular error was 4°. The subjects were also equally accurate at localizing sound sources in the front and rear hemispheres using direct location, indicating that precision does not fall off rapidly outside the visual field. The localization errors with a visual target using direct location were much smaller than those found when localizing sound sources, indicating that the response method probably contributed only a small fraction of the response errors in this experiment.

Although the subjects were asked to keep their eyes closed during the placement of the source, there were some extraneous cues (shadows visible through the closed eyelids, sounds generated by the experimenter, air movement during source placement, etc.) that may have allowed subjects to make judgments about the source location independently of the available audio information. In order to verify the insignificance of these cues, 100 trials were collected for each subject with the sound source disabled. The mean angular error in this condition was more than 50°, three times as large as when the sound source was enabled. The errors in

TABLE I. Mean angular errors. The mean errors and standard deviations for each of 27 stimulus regions were calculated separately for each of the four subjects, and then combined to generate the values in the table. The overall mean error at each distance is given in the last column, and the overall mean error at each azimuth is given in the last row. The standard errors are in parentheses. Trials where front-back confusions occurred have been excluded from these calculations.

Distance	Elevation		Azimuth			Mean
			Back < -120°	Side -120° to -60°	Front > -60°	
Close < 25 cm	High	>20°	27.0° (1.3°)	16.2° (0.5°)	20.6° (0.8°)	19.3°
	Mid	-20° to 20°	18.6° (0.5°)	16.3° (0.4°)	19.3° (0.9°)	
	Low	< -20°	20.4° (0.9°)	15.6° (0.6°)	19.6° (0.8°)	
Medium 25-50 cm	High	>20°	21.7° (0.8°)	15.5° (0.5°)	12.6° (0.6°)	15.8°
	Mid	-20° to 20°	17.9° (0.6°)	13.4° (0.4°)	14.6° (0.5°)	
	Low	< -20°	17.0° (0.6°)	12.3° (0.4°)	17.3° (0.5°)	
Far > 50 cm	High	>20°	22.9° (1.1°)	14.2° (0.7°)	12.6° (0.8°)	15.7°
	Mid	-20° to 20°	20.0° (0.8°)	13.3° (0.4°)	13.7° (0.6°)	
	Low	< -20°	14.8° (0.5°)	13.4° (0.5°)	16.1° (0.6°)	
Mean			20.0°	14.5°	16.3°	16.9°

azimuth, elevation, and distance were also much larger than in the audio experiment. Thus, although subjects received some information about source location from extraneous cues, this information was insignificant compared to the information provided by the intended auditory stimulus.

The data collection was divided into 2-h sessions, each consisting of four or five 100-trial blocks separated by short breaks. A total of 2000 trials per subject were collected over four or five 2-h sessions. Subjects participated in several training sessions prior to formal data collection in order to familiarize themselves with the experimental procedure. They were not, however, given feedback during these practice sessions.

### III. DIRECTIONAL LOCALIZATION RESULTS

#### A. Removal of front-back reversals

Front-back reversals are commonly reported in auditory localization. These reversals, which occur because the interaural level and time difference cues are approximately symmetric across the interaural axis of the head, cause listeners to perceive sounds at the mirror image of their true position across the frontal plane; a sound at 45° azimuth, for example, might be perceived at 135°. In this experiment, a relatively conservative definition was used to determine whether a reversal had occurred on a given trial: a reversal was declared only when the azimuth error was reduced at least 10° by reflecting the response across the frontal plane. According to this definition, front-back reversals occurred in approximately 10% of all trials in this experiment. In the analyses of directional localization performance, all trials where reversals occurred were omitted from the calculations. The distributions of front-back reversals across locations and across subjects are discussed later.

#### B. Angular error

The simplest measure of directional error is the angular error, which corresponds to the angle between the vector

from the center of the head to the source location and the vector from the center of the head to the response location. The angular error is a comprehensive measure of directional accuracy that incorporates both azimuth and elevation errors and includes the effects of systematic response biases and of response variability. The mean angular errors were calculated for stimuli in each of 27 different regions of space representing three azimuth ranges, three elevation ranges, and three distances. The results (Table I) show that overall directional accuracy varies by more than a factor of 2 with location of the source. The largest errors occurred at locations above and behind the subject, especially when the source was close (mean error 27°), and the smallest errors occurred at relatively distant locations in front of and to the side of the subject.

When averaged across all distances and all elevations, the angular error was smallest when the source was to the side, and greatest when the source was behind the subject (Table I, bottom row). The increase in error at locations behind the subjects may result, in part, from the awkwardness of moving the pointer behind the body. The error also increased substantially as the source approached within 25 cm of the head, especially for source in front of and above the subject (right column), and the error increased slightly with increasing elevation. Overall, averaged across all subjects and all locations, the mean angular error was 16.9°.

#### C. Azimuth error

The raw azimuth data give an indication of both the precision of azimuth localization and of any major response biases. In Figs. 2 and 3, azimuth data are shown for two subjects: CLL, who experienced an exceptionally large number of front-back reversals, and KMY, whose responses were typical of the other two subjects used in the study. In each panel of the figure, the second-order polynomial line best fitting the stimulus-response data has been plotted. This line approximates the systematic biases for that subject in the indicated source region.

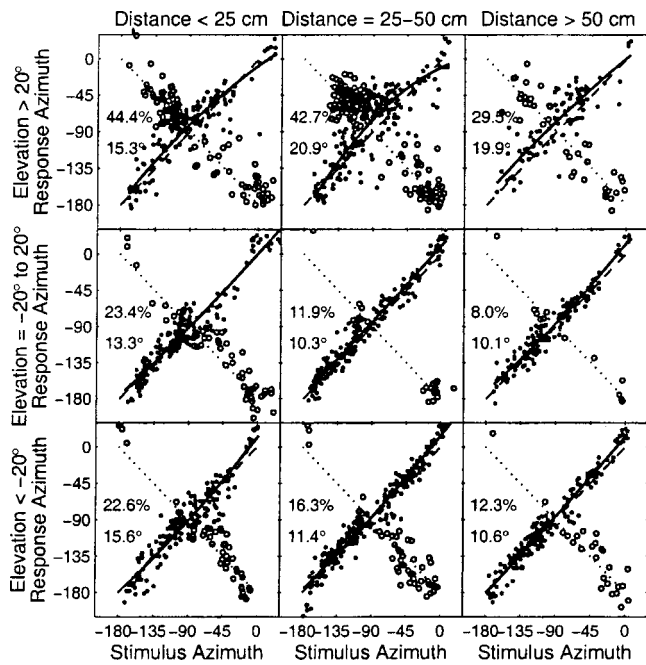


FIG. 2. Raw azimuth stimulus and response data for subject CLL. The data were sorted according to source distance into three regions: closer than 25 cm; from 25–50 cm; and farther than 50 cm. Similarly, the data were sorted by source elevation into three regions: above  $20^\circ$ ; between  $-20^\circ$  and  $20^\circ$ ; and below  $-20^\circ$ . Columns represent distance, increasing left to right, and rows represent elevation. The solid line is the second-order polynomial function of the stimulus location that best fits the response location. At the left of each panel are two numbers. The upper number is the percentage of front-back reversals (represented by open circles) in the region. The lower number is the BCRMS for all nonreversed responses in the panel (see the text). The dashed line represents “correct” responses, while the dotted line represents “perfect” reversals. Note that this subject exhibited an atypically large number of front-back reversals.

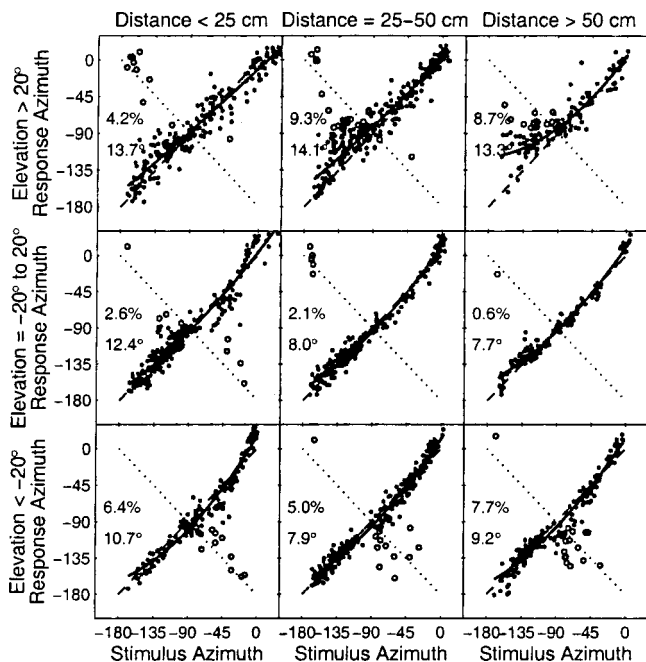


FIG. 3. Raw azimuth stimulus and response data for subject KMY. See Fig. 2 for details. This subject’s data are similar to those of DSB and DTD.

TABLE II. Mean BCRMS azimuth error at three distances (close= $<25$  cm, medium= $25\text{--}50$  cm, far= $>50$  cm) and at three elevations (high= $>20^\circ$ , medium= $-20^\circ$  to  $20^\circ$ , low= $>-20^\circ$ ). The data were calculated separately for each subject and averaged together to generate the figures in the table. Trials where front-back confusions occurred have been excluded from these calculations.

Elevation	Distance			Mean
	Close	Medium	Far	
High	14.2°	16.6°	14.8°	15.2°
Medium	12.2°	9.2°	9.5°	10.3°
Low	16.8°	10.4°	10.7°	12.6°
Mean	13.8°	12.1°	11.7°	12.6°

The biases change systematically with source location, and would confound a direct measure of response variability such as standard deviation. Therefore, a special measure of response variability that excludes the systematic response bias, called the bias-corrected root-mean-square (BCRMS) error, was calculated at nine elevation and distance locations for each subject. The BCRMS is simply rms error between the response location and the quadratic regression curve representing the best fit of the responses as a function of source azimuth. The BCRMS error is shown at the left of each panel in the figure, and the mean BCRMS azimuth errors averaged across the four subjects are shown in Table II. It should be noted that all four subjects performed quite similarly in terms of the BCRMS error, with the exception of subject CLL at high elevations.

The BCRMS error in azimuth was significantly larger at high elevations than at low or middle elevations for each of the four subjects (F-test,  $p < 0.002$ ). Note that this increase in error is at least in part a result of the increased sensitivity of azimuth at high elevations in the polar coordinate system. The same effect is not as pronounced at low elevations because more trials were collected at very high elevations ( $>45^\circ$ ) than at very low elevations ( $<-45^\circ$ ).

The BCRMS error was not strongly dependent on distance. The overall BCRMS error (averaged across all three elevations) was significantly larger at the closest source distances ( $<25$  cm) than at the greatest source distances ( $>50$  cm) for only two of the four subjects (KMY and CLL). Subject DSB actually exhibited significantly larger errors at intermediate distances (25–50 cm) than at close distances ( $<25$  cm) (F-test,  $p < 0.02$ ).

#### D. Distribution of front-back reversals

The top number on the left side of each panel in Figs. 2 and 3 indicates the percentage of trials where front-back reversals occurred in each location. Figure 4 summarizes the relationship between the percentage of front-back reversals and the source location. Four important observations can be made from the reversal data.

- (i) Subject CLL has far more reversals than any other subject, and dominates the mean reversal percentages across subjects. In certain locations, CLL reverses the majority of trials.



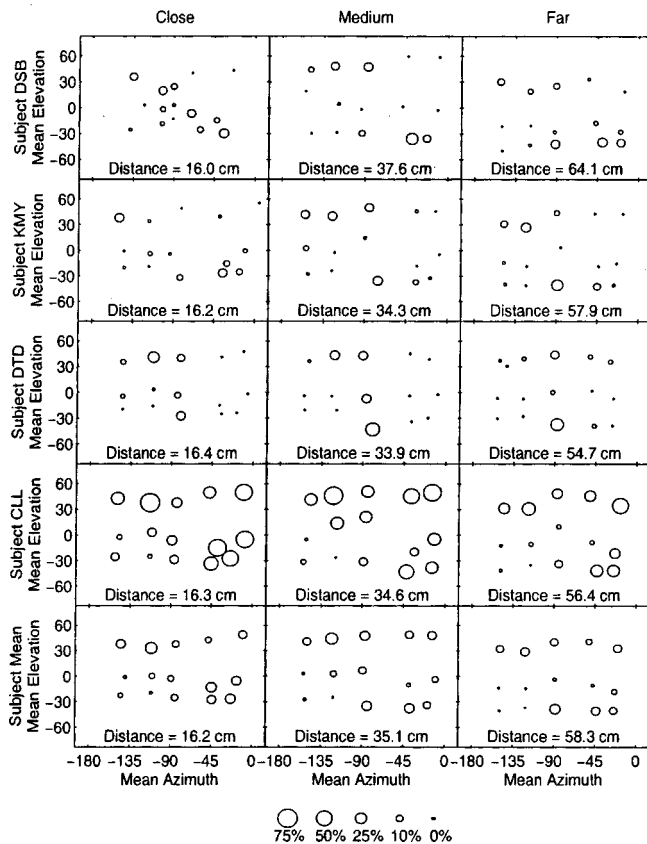


FIG. 4. Spatial distribution of front-back reversals. First, the data were sorted into three nonoverlapping distance bins, then each distance bin was sorted into eight overlapping azimuth bins, and finally each azimuth bin was sorted into three nonoverlapping elevation bins. The number of reversals in each bin is shown as a function of mean location for each individual subject and averaged across all four subjects, where a reversal is defined as any trial where the response was at least  $10^\circ$  closer to the mirror image of the source location across the frontal plane than to the actual source location (see the text). For clarity, only five bins are shown in azimuth. The three bins to the side and the rear are nonoverlapping, while the two bins near  $0^\circ$  are overlapping. The percentage reversals at each location are shown by the size of the circle, according to the code shown at the bottom of the figure. The distances at the bottom of each panel represent the mean distance of all the data points in that region.

- (ii) Only CLL shows a significant distance dependence in the percentage of front-back reversals. CLL reversed a significantly larger percentage of trials at close and medium distances ( $<50$  cm) than at far distances ( $>50$  cm) (one-tailed t-test,  $p < 0.005$ ).
- (iii) Relatively few reversals occur at middle elevations. One-tailed t-tests ( $\alpha = 0.005$ ) indicate that all four subjects reversed a significantly larger percentage of trials at high elevations ( $>20^\circ$ ) than at middle elevations, and that three subjects (DSB, KMY, and DTD) reversed a significantly larger percentage of trials at low elevations ( $<-20^\circ$ ) than at middle elevations.
- (iv) In the rear hemisphere, the vast majority of reversals occurs at high elevations. In contrast, almost all reversals in the front hemisphere for subjects DSB and KMY occur at low elevations. The variations in the placement of reversals across subjects is not surprising, since the subjects must essentially make an arbitrary

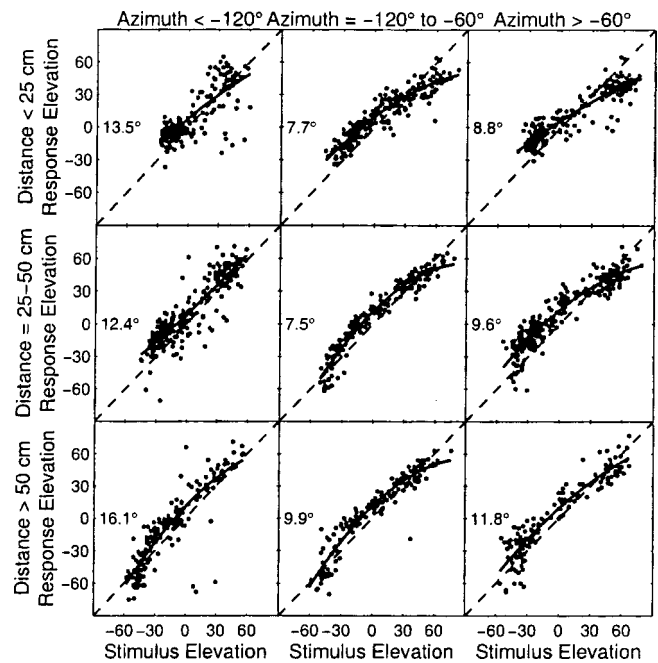


FIG. 5. The nine panels show raw elevation stimulus and response locations for subject KMY, a typical subject in the experiment. Azimuths were divided into regions less than  $-120^\circ$ , from  $-120^\circ$  to  $-60^\circ$ , and greater than  $-60^\circ$ . Distances were divided into regions less than 25 cm, from 25 to 50 cm, and greater than 50 cm. A dashed line represents correct responses. The BCRMS elevation error is shown at the left side of each panel, and the solid line represents the best second-order polynomial fit of the stimulus data to the response locations. Note that the stimuli range from approximately  $-45^\circ$  to  $80^\circ$  in elevation. The data are limited at low elevations because the subject's torso and the chin rest prevented placement at some source locations. For example, low elevations are particularly truncated at close distances behind the subject where the neck and back prevented placement at low elevations.

decision about the true location of the source whenever they are unsure about the actual hemisphere of the source.

### E. Elevation error

The raw elevation data for a typical subject are shown in Fig. 5 in a format similar to the one used for the raw azimuth data. These raw data are typical of the other subjects in the experiment. The overall BCRMS elevation errors, averaged across the four subjects, are provided in Table III. Several important observations can be made from these data:

TABLE III. Mean BCRMS elevation error at three distances (close =  $<25$  cm, medium =  $25-50$  cm, far =  $>50$  cm) and at three azimuths (rear =  $<-120^\circ$ , side =  $-120^\circ$  to  $-60^\circ$ , front =  $>-60^\circ$ ). The data were calculated separately for each subject and averaged together to produce the value shown in the table.

Distance	Azimuth			Mean
	Rear	Side	Front	
Close	14.6°	9.5°	11.2°	11.7°
Medium	13.0°	8.6°	10.0°	10.5°
Far	14.7°	9.3°	11.3°	11.8°
Mean	14.1°	9.1°	10.8°	11.3°

- (i) The elevation responses tended to show more dramatic biases than the azimuth data. In the data shown for KMY (Fig. 5), the quadratic line best fitting the data is typically concave down, in part because the subject tended to underestimate the elevation of high sources.
- (ii) The overall BCRMS error in elevation, which attempts to eliminate the effects of bias, was comparable to that for azimuth ( $11.3^\circ$  vs  $12.6^\circ$ ). Note that the azimuth figure is inflated somewhat by its increased sensitivity at high and low elevations in the polar coordinate system.
- (iii) Elevation localization performance was best to the side, and worst to the rear. An F-test on the BCRMS errors reveals that each of the four subjects was significantly more accurate in front than in back, and most accurate to the side ( $p < 0.01$  level).
- (iv) Elevation performance did not depend on distance in a consistent way. Two of the subjects (DSB and KMY) had significantly lower errors at distances less than 25 cm than at distances greater than 25 cm, and the other two subjects had significantly lower errors at distances greater than 25 cm than at distances closer than 25 cm (F-test,  $p < 0.005$ ).

## F. Response biases

To this point, the primary focus has been the variability of subject responses in the form of the bias-corrected unsigned error. Systematic directional biases are also of considerable interest. Figure 6 shows the response bias (mean uncorrected signed error) in azimuth and elevation as a function of source location. Note that the directional biases are generally invariant to source distance. Although the directional biases differ substantially from subject to subject, the general pattern of biases for each of the subjects is consistent across the three distance bins. Subject DSB, for example, has a bias up and toward the front for sources behind and above the head at all distances, while CLL is generally biased down and toward the front at high elevations. It appears that directional response biases are roughly independent of distance.

## G. Discussion

In order to put these results into context, it is useful to compare them to previous estimates of directional localization ability available in the literature. Although no data are available on proximal-region localization, our results at distances greater than 50 cm can be compared to previous data collected 1 m or farther from the subject. Two studies which have evaluated directional localization (position identification) are Wightman and Kistler (1989) and Makous and Middlebrooks (1990).

The overall angular errors measured in this study were substantially smaller than those measured by Wightman and Kistler ( $15.2^\circ$  at distances greater than 50 cm in this study, compared to  $21.1^\circ$  measured by Wightman and Kistler). This discrepancy most likely results from the increased stimulus

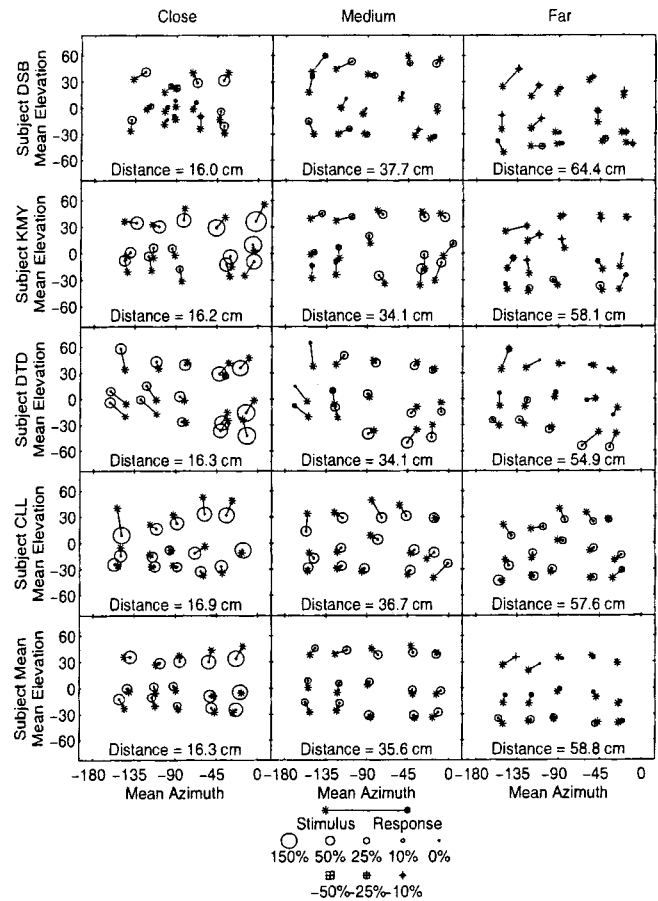


FIG. 6. Response bias size and direction as a function of source location. Details are similar to Fig. 4. The asterisk is the mean stimulus location, while the dot is the mean response location. The circles indicate that the subject overestimated distance (see the legend), while the crosses surrounded by squares indicate an underestimate of distance.

uncertainty in the Wightman and Kistler study (they used a scrambled stimulus spectrum rather than the fixed stimulus spectrum in this study).

In contrast, the standard deviations in azimuth and elevation measured in this study were substantially larger than those measured by Makous and Middlebrooks. For example, their results indicate a standard deviation of only  $1.9^\circ$  in azimuth and  $3.3^\circ$  in elevation for sources directly in front of the listener ( $0^\circ$  azimuth and  $-5^\circ$  in elevation), compared to standard deviations (BCRMS) of approximately  $6^\circ$  in azimuth and  $8^\circ$  in elevation for sources in front of the listener in this study.

All three studies indicate that directional localization judgments are least accurate when the source is located above and behind the head. Since all three studies report poorest performance in this region, using three different response methods, it is likely that there are some perceptual problems in localizing sound behind and above the head. However, the reasons for poor perception in this region are not obvious.

Finally, more front-back reversals occurred in this experiment than in the two earlier studies. If front-back confusions are counted whenever the stimulus and response locations are on opposite sides of the frontal plane, front-back confusions occurred in 13% of all trials in this experiment at

distances greater than 50 cm, compared to 6% of all trials in the other two experiments.

The comparison of the data at distances greater than 50 cm with previous data is useful for establishing a baseline for comparison with the results at closer distances. Of primary importance in this study, however, is the effect of an extremely close source on directional localization ability. The angular error increases significantly as distance decreases (Table I), but this increase is at least in part a result of larger response biases when the source is close, rather than larger response variability. Response variability, measured by the BCRMS errors in azimuth and elevation, did not vary consistently with distance. The azimuth error was significantly larger at close distances ( $<25$  cm) than at far distances ( $>50$  cm) for only two of the four subjects. The elevation bias-corrected error was significantly larger at close distances than at far distances for two subjects, but significantly larger at far distances than at close distances for the other two subjects. There is also some reason to believe that experimental error is slightly greater for very close sources than for more distant sources. At locations very close to the head, direction is very sensitive to small displacement errors. At 12 cm, for example, a 1-cm error in the subject response, or in the measurement of the stimulus and response locations, can cause a directional error of nearly  $5^\circ$ . When the increased error sensitivity of the response method for very near sources is weighed against the relatively minor decrease in performance at close distances, it appears that source distance has, at most, a marginal effect on directional accuracy in the proximal region.

Front-back reversals increased slightly at close distances for all four subjects, but only one subject (CLL) reversed a significantly larger percentage of trials at close distances than at far distances. CLL appeared to be a "poor localizer" in general, in that he experienced substantially more front-back confusions than the other subjects, even at the greatest distances tested. Although data are available from only one subject, it may be the case that the localization problems of poor localizers are exacerbated when the source is very near the head, but that normal localizers may be unaffected by sources very close to the head.

#### IV. DISTANCE LOCALIZATION

##### A. Results

One of the primary motivations for this experiment was an examination of the accuracy of auditory depth perception for nearby sources. The raw data for proximal-region distance perception for a typical subject are provided in Fig. 7. There are three striking features in these data:

- (i) The magnitudes of the distance errors tend to increase with distance. For this reason, the stimulus-response curves in distance have been plotted on a log-log graph rather than a linear graph.
- (ii) The distance errors are greater near the median plane [azimuths in back ( $<-120^\circ$ ) and in front ( $>-60^\circ$ )] than at more lateral locations.
- (iii) The distance errors are greater at high elevations ( $>20^\circ$ ) than at middle and low elevations.

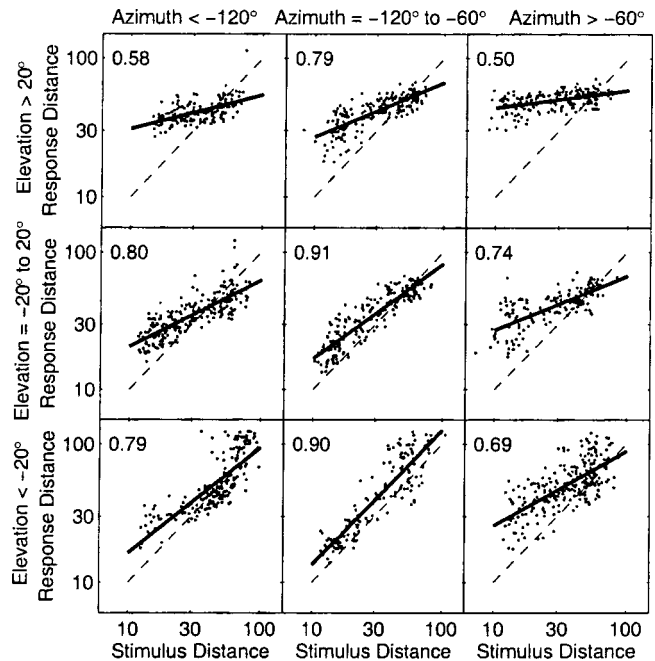


FIG. 7. Raw distance data for subject KMY. The columns represent different azimuths, and the rows represent different elevations. The data are plotted on a log-log scale representing the distance from the center of the head (in cm). Note that in a typical subject the tip of the nose is approximately 10 cm from the center of the head, and the ears are approximately 7 cm from the surface of the head. The correlation coefficient of the log stimulus distance and log response distance is shown at the top left of each panel. The dashed line indicates correct responses, and the solid line is the least-squared linear fit of the log-log data.

An overall summary of distance performance collapsed over all elevations is the RMS percentage distance error (Fig. 8). This measure indicates that the overall average error in distance is approximately 30%–40% across all azimuth locations.

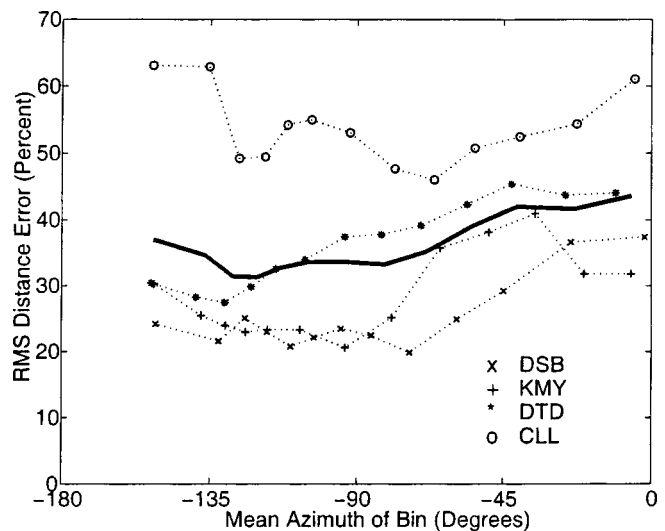


FIG. 8. rms percent distance error as a function of azimuth. Each subject's responses were sorted by azimuth into 13 overlapping bins containing 14% of the total number of trials. Then, the rms percentage distance error (uncorrected for bias) calculated in each bin was plotted as a function of the mean azimuth location in each of the 13 bins. The solid line shows the average of the percent rms errors calculated for each of the four subjects.

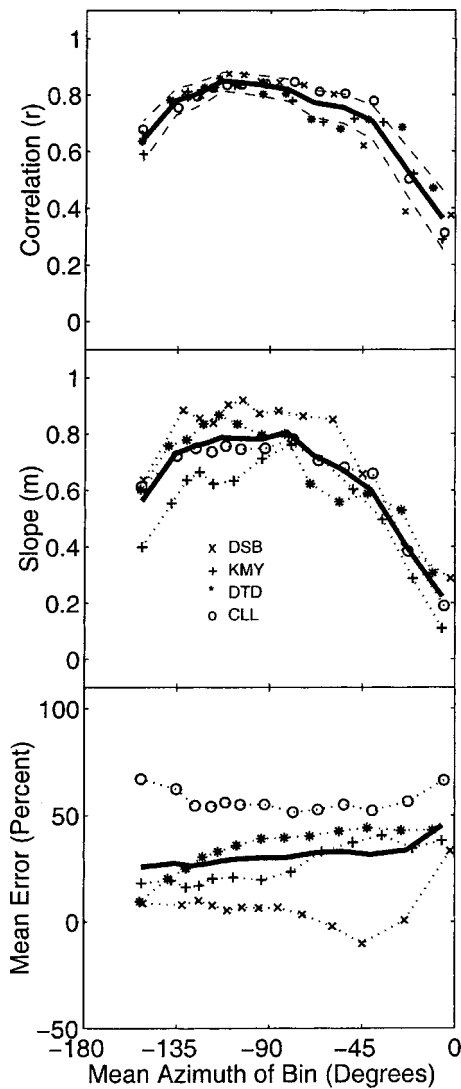


FIG. 9. Correlation, slope, and overall bias of the log distance responses as a function of azimuth. As in Fig. 8, the subject responses were sorted by azimuth into 13 overlapping bins, each containing 14% of the total number of trials, and the correlation coefficient (top panel), the slope of the linear regression line (middle panel), and the overall mean signed percentage error in distance (bottom panel) were calculated separately in each bin and plotted as a function of the mean azimuth of each bin. The solid lines represent the mean values across the four subjects. In the top panel, a Fisher transformation was used to average the correlation coefficients across the four subjects and to compute the 95% confidence interval (dashed lines).

The overall distance errors can be divided into three components: an overall distance-independent bias representing a general tendency to overestimate or underestimate distance (the mean percentage error in distance); a bias in “scaling” representing a tendency to compress or expand the responses in distance (represented by the slope of the least-squared linear fit of the data); and an error term related to the variability of the responses for a particular stimulus location (represented by the spread of responses around the least-squared fit line).

Figure 9 illustrates the bias and uncertainty components of the distance error separately. The top panel shows the correlation coefficient of the log stimulus and response distances as a function of source azimuth. The correlation coefficient is related to the bias-corrected variability in the re-

sponses and can be viewed as the degree to which the source distance can be determined from a linear function of the log response distance.<sup>2</sup> Note that the correlation coefficient is significantly greater for sources at lateral locations than for sources at medial locations, and that this pattern is consistent across each of the four subjects. Also note that the data are roughly symmetric in the front and rear hemispheres, but that there are no data points behind  $-150^\circ$  in azimuth.

The middle panel shows the slope of the line best fitting the log stimulus data to the log response data at each azimuth location. This slope is a measure of the scaling bias in the responses: a slope of less than 1 indicates that the responses varied over a narrower range of distances than the stimulus locations (a compression in the responses). Like the correlation coefficient, the slope is relatively high at lateral locations and relatively low at medial locations. This indicates that the subject’s responses are more sensitive to the true source location at lateral locations, which accounts for the increased correlation coefficient at these locations. Note that, even at lateral locations, the slope is less than 1 for all subjects, indicating a general tendency to compress distance responses.

The bottom panel shows the overall percentage error in distance, which indicates any distance-independent biases in the subject’s responses. This is the only performance measure that shows a substantial difference across the four subjects. In particular, subject CLL exhibits a strong tendency to overestimate distance (in excess of 50%). The other subjects also generally overestimated distance, but to a lesser extent. The overall pattern of this percentage error as a function of azimuth is quite similar to the rms percentage error in Fig. 8 for each of the four subjects, indicating that rms percentage error is dominated by overall bias and that it is only a weak indicator of response precision.

The degree to which the correlation coefficient and the slope of the linear regression line represent the characteristics of the subject responses can be illustrated by a comparison of the raw stimulus–response data in two azimuth bins, one near the median plane and one near the interaural axis. The results for subject DSB (Fig. 10), which are typical of those for the other subjects, confirm that the primary reason for the decrease in correlation in the median plane is that the slope of the stimulus–response line is much lower. In fact, almost all of the responses in the front bin are grouped around 60 cm, independent of the actual stimulus location. A similar pattern occurs in the data from all four subjects.

The correlation coefficient and the slope of the regression line also indicate that distance localization performance is substantially worse at high elevations (above  $20^\circ$ ) than at middle and low elevations. The correlation coefficient between the log stimulus and response distances, calculated separately for each subject in each of 13 azimuth bins and averaged using the Fisher transformation (Devore, 1991), decreases from 0.81 at locations below  $20^\circ$  in elevation to 0.65 at locations above  $20^\circ$ . Similarly, the average slope of the responses decreases from 0.76 at low and middle elevations to 0.42 at high elevations.



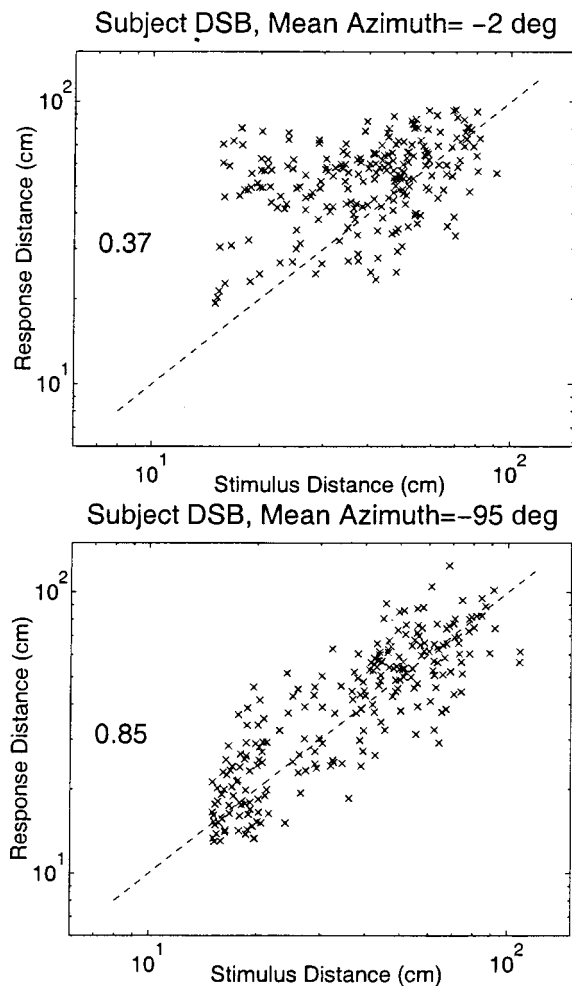


FIG. 10. Raw distance data for subject DSB in front and to the side. These plots show the data for subject DSB in the 1st and 7th azimuth bins used to calculate the results in Fig. 9. In front (top panel), the responses tend to be clustered around 60 cm independent of the source location, while to the side (bottom panel) the response distance varies systematically with stimulus distance.

## B. Discussion

It is difficult to compare distance localization in this experiment to previous results. Few studies have directly examined auditory distance perception as a function of direction, and even fewer have examined distance perception for nearby sources. We know of only two studies which have directly examined proximal-region distance perception. Ashmead, LeRoy, and Odom (1990) found that listeners were able to perceive relatively large ( $>16\%$ ) changes in the distance of a nearby sound source directly in front of them even when the amplitude of the source was manipulated to eliminate loudness-based distance cues. This indicates that some distance information (perhaps spectral) is available even in the median plane, where binaural distance cues are minimal.

Simpson and Stanton (1973) performed an experiment specifically designed to look for binaural distance cues for close sources. Subjects were asked to estimate the distance of a sound source placed directly in front of the listener at one of five locations ranging from 30 cm to 2.7 m. Some of the subjects used a fixed head position during the experiment, some were allowed to turn their heads, and some were re-

quired to move their heads. Simpson and Stanton found that head motion had no significant effect on distance perception. Since the results of the current experiment indicate that distance accuracy is substantially better for lateral sources than for medial sources, it is surprising that Simpson and Stanton's subjects were not able to judge distance more accurately when they were allowed to turn their heads away from the sound source. Amplitude and reverberation in the Simpson and Stanton study may account for the discrepancy. The amplitude of the source was fixed during their experiment, and their subjects were seated in the corner of a sound-treated listening booth, with their heads only 25 cm from either wall. Thus, it is likely that their subjects were able to use amplitude and reverberation cues to judge distance, and these cues may have dominated the binaural distance cues in their experiments.

In contrast to the Simpson and Stanton paper, two studies of distal-region localization have indicated that distance perception is better for sources along the interaural axis than for sources in the median plane when the amplitude of the source is randomized. Holt and Thurlow (1969) found that subjects could accurately determine the relative distances of the sound sources when they were lined up with the interaural axis (rank-order correlation of 0.93), but not when the sources were directly in front of the subject. Gardner (1969) informally reported a similar result. The relationship between azimuthal position and distance localization accuracy found in these earlier studies is in agreement with the results of this experiment, but we cannot explain why subjects were able to perform so well in the distal region where binaural distance cues are largely absent.

Other than the observations of Holt and Thurlow, and Gardner, no previous studies have indicated that distance perception is better for lateral sources than medial sources at close distances. Furthermore, the strong correlations found in this study (as large as 0.85 for sources near  $90^\circ$ ) indicate that distance perception is reasonably accurate in this region. Performance in this region appears to exceed that indicated in any previous distance study where overall level cues were unavailable, which is especially noteworthy considering the additional requirements of simultaneously determining source azimuth, elevation, and distance in this experiment.

## V. COMPARISON OF RESULTS TO PROXIMAL-REGION HRTF MEASUREMENTS

By comparing the results of this psychoacoustic experiment with previously measured HRTFs in the proximal region, we can gain insights into the mechanisms of proximal-region localization. The features of the proximal-region HRTFs, along with previous results from distal-region localization experiments, can explain the relatively weak distance dependence of directional localization, as well as the relatively accurate distance judgments for lateral sources.

Although three of the four subjects were slightly less accurate at azimuthal localization when the sound source was close to the head, the decrease in performance was relatively minor. Similar horizontal localization performance in the proximal and distal regions may indicate that low-frequency ITDs dominate azimuth judgments in the proximal region as

they have been shown to do in the distal region. Previous work by Wightman and Kistler (1992) has shown that ITDs tend to dominate azimuthal localization when the stimulus contains low-frequency energy. Note, however, that Wightman and Kistler's experiments manipulated the time delay in HRTFs measured in the distal region. The low-frequency time delay clearly dominates perception with the distal-region HRTFs, where the ILD was significant only at high frequencies. When the source is in the proximal region, however, the ILD can be large even at low frequencies, and the Wightman and Kistler data provide no direct evidence that the ITD dominates the influence of low-frequency ILD on azimuth perception. The absence of a strong lateral azimuth bias for nearby sources provides some indirect evidence that ITD dominance extends into the proximal region. In the proximal region, an increase in ILD could result either from a source moving closer to the head or from a source moving away from the median plane. If azimuth judgments were based on ILD, one might expect listeners to confuse the distance and direction of the source in the proximal region, resulting in a lateral bias for nearby sources. There is, however, no indication of such a bias in the data. The lack of lateral directional biases for nearby sources, coupled with comparable directional accuracy in the proximal and distal regions, indicates that proximal-region azimuth perception is most likely based on ITDs which are essentially independent of source distance.

The psychoacoustic results indicate that elevation perception does not depend on distance in a systematic way. Two subjects performed slightly better in elevation when the source was distant, and two performed better when the source was close. This is consistent with the observation (Brungart and Rabinowitz, 1999) that the high-frequency features of the HRTF which change systematically with elevation are relatively independent of distance.

The distance perception abilities of our subjects, and in particular their ability to make unbiased, accurate distance judgments about lateral sources and their inability to make distance judgments about medial sources, suggest that the variations in the ILD with angle and distance provide a useful binaural proximal-region distance cue. In the distal region, the ILD varies only with direction. In the proximal region, the ILD increases as the source approaches the head. The usefulness of this increase as a distance cue is related to the range over which the ILD varies in a particular direction. The range of possible ILDs is largest when the source is to the side, and decreases to zero in the median plane. This pattern mirrors the distance performance by the subjects, which was also best for lateral sources and worst in the median plane. In fact, the only major discrepancy between distance localization accuracy and the range of possible ILD values is that localization performance appears to plateau in the region from  $-45^\circ$  to  $-135^\circ$ , while the span of ILDs increases systematically up to  $-90^\circ$ . This could be explained by the well-known range effect in stimulus identification experiments, which causes sensitivity to changes in a stimulus to decrease when the range of possible values increases (Durlach and Braida, 1969; Koehnke and Durlach, 1989). The range effect, which is based on memory noise

rather than sensory noise, could explain the saturation in performance seen in this experiment. Note that ILD-based distance cues could also explain the decrease in performance at high elevations (and the more rapid decrease in performance away from  $-90^\circ$ ), since the ILD is smaller at high elevations than in the horizontal plane.

## VI. CONCLUSIONS

The general results of these experiments can be summarized as follows:

- (i) The angular error, which includes the effects of response bias and response variability, increases as the source approaches the head, particularly in front of and behind the listener.
- (ii) The bias-corrected rms azimuth error is greatest at high elevations and generally increases slightly at close distances.
- (iii) The bias-corrected rms elevation error is lowest for lateral sources and greatest behind the listener. It does not vary consistently with distance.
- (iv) Distance perception is most accurate for lateral sources and least accurate near the median plane. For lateral sources, distance judgments were highly correlated with actual source position ( $r > 0.85$ ), and were relatively unbiased; in the median plane, the correlations were low ( $r < 0.4$ ). The results generally indicate better distance perception in the proximal region than in previously reported studies involving sources of unknown strength in anechoic conditions.
- (v) The psychoacoustic results are consistent with previously measured HRTFs in the proximal region, which indicate that ILD varies with distance in the proximal region while ITDs are roughly independent of distance. In particular, the results support the hypothesis that ILDs are an important binaural distance cue in the proximal region.

It appears that directional localization is modestly degraded when sources are close to the head, but that distance perception may be substantially improved, at least for sources away from the median plane, by the availability of binaural distance cues in the proximal region. Additional experiments are necessary to fully understand the mechanisms of proximal-region localization. The next paper in this series will look at the effects of different stimuli (e.g., bandlimited, monaural, or fixed amplitude) on proximal-region localization.

## ACKNOWLEDGMENTS

The authors would like to thank Steve Colburn and Bill Peake for their assistance throughout these experiments. This work was supported in part by AFOSR Grant Nos. F49620-96-1-0202 and F49620-98-1-0108.

<sup>1</sup>Note that these regions have sometimes been referred to as the "near field" and the "far field." Since these terms have very specific meanings in physical acoustics, we have introduced new terminology to eliminate any ambiguities.

<sup>2</sup>The use of the correlation coefficient requires an approximately linear relationship between the two variables. An examination of the raw data indicates that this approximately linear relationship exists between the log of the stimulus distance and the log of the response distance for each of the four subjects.

- Ashmead, D., Davis, D. L., and Odom, R. D. (1990). "Perception of the relative distances of nearby sound sources," *Percept. Psychophys.* **47**, 326–331.
- Brungart, D., Rabinowitz, W., and Durlach, N. (1999). "Evaluation of response methods for near-field auditory localization experiments," *Percept. Psychophys.* (in press).
- Brungart, D., and Rabinowitz, W. (1996). "Auditory localization in the near-field," in *Proceedings of the Third International Conference on Auditory Display*, Santa Fe Institute.
- Brungart, D., and Rabinowitz, W. (1999). "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.* **106**, 1465–1479.
- Butler, R., Levy, E., and Neff, W. (1980). "Apparent distance of sounds recorded in echoic and anechoic chambers," *J. Exp. Psychol.* **6**, 745–750.
- Coleman, P. (1963). "An analysis of cues to auditory depth perception in free space," *Psychol. Bull.* **60**, 302–315.
- Devore, J. (1991). *Probability and Statistics for Engineering and the Sciences* (Brooks-Cole, Belmont, MA).
- Duda, R., and Martens, W. (1998). "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.* **104**, 3048–3058.
- Durlach, N., and Braida, L. (1969). "Intensity perception. i. preliminary theory of intensity perception," *J. Acoust. Soc. Am.* **46**, 372–383.
- Gardner, M. B. (1969). "Distance estimation of speech signals," *J. Acoust. Soc. Am.* **48**, 47–53.
- Hartley, R., and Fry, T. (1921). "The binaural location of pure tones," *Phys. Rev.* **18**, 431–442.
- Holt, R., and Thurlow, W. (1969). "Subject orientation and judgment of distance of a sound source," *J. Acoust. Soc. Am.* **46**, 1584–1585.
- Koehnke, J., and Durlach, N. (1989). "Range effects in the identification of lateral position," *J. Acoust. Soc. Am.* **86**, 1176–1178.
- Little, A., Mershon, D., and Cox, P. (1992). "Spherical content as a cue to perceived auditory distance," *Perception* **21**, 405–416.
- Lounsbury, B., and Butler, R. (1979). "Estimation of distances of recorded sounds presented through headphones," *Scand. Audiol.* **8**, 145–149.
- Makous and Middlebrooks (1990). "Two dimensional sound localization by human listeners," *J. Acoust. Soc. Am.* **85**, 2188–2200.
- Mershon, D., and Bowers, J. (1979). "Absolute and relative cues for the auditory perception of egocentric distance," *Perception* **8**, 311–322.
- Mershon, D., and King, L. (1975). "Intensity and reverberation as factors in the auditory perception of distance," *Percept. Psychophys.* **18**, 400–415.
- Middlebrooks, J., and Green, D. (1991). "Sound localization by human listeners," *Annu. Rev. Psychol.* **42**, 135–139.
- Musicant, A., and Butler, A. (1984). "The influence of pinnae-based spectral cues on sound localization," *J. Acoust. Soc. Am.* **75**, 1195–1200.
- Oldfield, S., and Parker, S. (1986). "Acuity of sound localization: a topography of auditory space. III. Monaural hearing conditions," *Perception* **15**, 67–81.
- Perrett, S., and Noble, W. (1997). "The effect of head rotations on vertical plane localization," *J. Acoust. Soc. Am.* **102**, 2325–2332.
- Rayleigh, L. (1907). "On our perception of sound direction," *Philos. Mag.* **13**, 214–232.
- Roffler, S., and Butler, R. (1968). "Factors that influence the localization of sound in the vertical plane," *J. Acoust. Soc. Am.* **43**, 1255–1259.
- Simpson, W., and Stanton, L. D. (1973). "Head movement does not facilitate perception of the distance of a sound," *Am. J. Psychol.* **86**, 151–159.
- Slattery, W., and Middlebrooks, J. (1994). "Monaural sound localization: acute versus chronic unilateral impairment," *Hearing Res.* **75**, 38–46.
- Stewart, G. (1911a). "The acoustic shadow of a rigid sphere with certain applications in architectural acoustics and audition," *Phys. Rev.* **33**, 467–479.
- Stewart, G. (1911b). "Phase relations in the acoustic shadow of a rigid sphere; phase difference at the ears," *Phys. Rev.* **34**, 252–258.
- Wallach, H. (1939). "On sound localization," *J. Acoust. Soc. Am.* **10**, 270–274.
- Wightman, F., and Kistler, D. (1989). "Headphone simulation of free-field listening. I. Stimulus synthesis," *J. Acoust. Soc. Am.* **85**, 858–877.
- Wightman, F., and Kistler, D. (1992). "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.* **91**, 1648–1660.

# Determination of optimal data placement for psychometric function estimation: A computer simulation

Chan F. Lam<sup>a)</sup>

*Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, South Carolina 29425*

Judy R. Dubno and John H. Mills

*Department of Otolaryngology and Communicative Sciences, Medical University of South Carolina, Charleston, South Carolina 29425*

(Received 27 July 1998; revised 25 February 1999; accepted 28 May 1999)

Psychometric functions are used to relate the responses of a subject to physical stimuli in a variety of psychophysical tasks. However, it is time consuming to obtain data to determine a psychometric function if many stimulus levels and many trials are required. A computer simulation was conducted to determine the minimum number of data points needed for such a determination. The computer simulation also determined the optimal placements of the stimuli and the number of trials per datum point for psychometric function determinations. Results indicate that a 2-point sampling method with 30–50 trials per point at optimal locations can produce a psychometric function with accurate spread and threshold estimates in a yes–no paradigm. However, the 4-point sampling method yields statistically smaller variances of the estimates. For the 2-alternative forced-choice paradigm, at least 120 trials per point are needed for the 2-point sampling method's estimated parameters to differ from the known parameter values by less than 5%. The simulation results suggest that 3-alternative or 4-alternative forced-choice is preferable to 2-alternative. Furthermore, when a criterion-free paradigm is not required, the yes–no paradigm is a better procedure than  $m$ -alternative forced-choice for obtaining the corresponding psychometric function because of smaller standard deviation of the estimates and smaller number of trials/point required. © 1999 Acoustical Society of America. [S0001-4966(99)02909-4]

PACS numbers: 43.66.Yw [JWH]

## INTRODUCTION

Psychometric functions have been used to relate the responses of a subject to physical stimuli in a variety of psychophysical tasks (Allen and Wightman, 1994; Dauman *et al.*, 1993; Dubno *et al.*, 1995; Klein, 1989; Martin *et al.*, 1996; McDevitt, 1985; O'Regan and Humbert, 1989; Paut *et al.*, 1992; Sand *et al.*, 1994; Wall *et al.*, 1996). Different methods for estimating psychometric functions have been used, including the constant stimuli method, the hybrid adaptive method of Hall (1981), and the adaptive up–down method (Leek *et al.*, 1992; Dai, 1995). In order to decrease the data collection time for a yes–no psychometric task, a 4-point sampling method was proposed by Lam *et al.* (1996, 1997).

Psychometric functions have also been estimated for various  $m$ -alternative forced-choice situations, including two, three, and sometimes four, alternatives. Would the 4-point sampling method also be applicable to the forced-choice situations? O'Regan and Humbert (1989) have found that there are significant biases in the threshold and slope estimates when small sample sizes are used in the forced-choice situations. By extending the theoretical variance calculation of the threshold and slope estimates for the yes–no paradigm developed by Berkson (1955) and Wetherill (1963) to the  $m$ -alternative forced-choice paradigms, O'Regan and Hum-

bert found that the optimal locations (i.e., minimum total variance of the two estimates) for two sampling points are not symmetric with respect to the threshold location for 2-, 3- and 4-alternative forced-choice paradigms. Using Monte Carlo simulations for 10 trials/point in a 10-point constant stimuli paradigm, they confirmed that biases exist in the threshold and slope estimates. Recently, Lam *et al.* (1996) have also found that 10 trials/point is a poor estimate for data points for the purpose of estimating threshold and slope parameters of a psychometric function in a yes–no paradigm. It would be of interest to determine the minimum number of trials needed per point for various forced-choice paradigms in comparison to the yes–no paradigm.

A concomitant question is the placement of the stimuli. Levitt (1971) stated that “a typical rule of thumb is for the experimenter to place the observations so as to roughly cover the range of  $X_{10}$  to  $X_{90}$ ,” where  $X_{10}$  and  $X_{90}$  are the stimulus levels corresponding, respectively, to 10% and 90% of the maximum response of the psychometric function. However, where exactly should the stimuli be located? Furthermore, the placements of the stimuli and the number of stimuli are probably inter-related.

It is the purpose of this paper to suggest answers to questions concerning the minimum number of data points, the optimal locations of stimuli, and the number of trials per datum point needed to estimate a psychometric function for both yes–no and forced-choice paradigms using computer simulations.

<sup>a)</sup>Electronic mail: lamcf@musc.edu



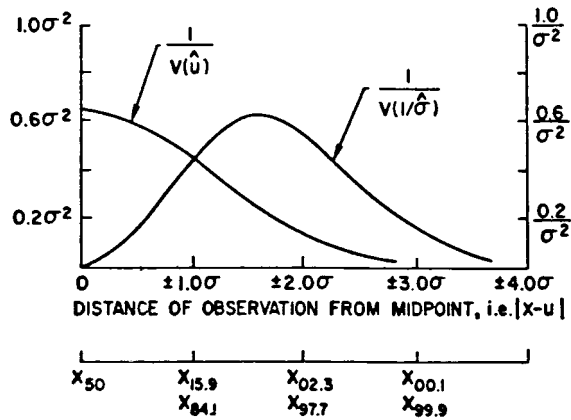


FIG. 1. The reciprocal of the error variance for the estimates of  $\mu$  and  $1/\sigma$ . The left-hand vertical axis is for the reciprocal of  $\text{Var}(\hat{\mu})$ , while the right-hand vertical axis is for the reciprocal of  $\text{Var}(1/\hat{\sigma})$ .  $X_{50}$  denotes the value of  $X$  for a 50% correct detection. (Adapted from Fig. 3 of Levitt, 1971.)

## I. METHOD

Using the cumulative normal function,  $N(\mu, \sigma^2)$ , Levitt (1971) determined the relationships between the reciprocal of the error variance of the estimate of the mean  $\mu$  and the estimate of the standard deviation  $\sigma$  and the distance between the observation and the true mean (i.e., midpoint). Such relationships were graphed in Fig. 3 of Levitt (1971), and an adaptation is shown here in Fig. 1. The two curves in Fig. 1 indicate that in order to obtain the smallest variance, hence highest confidence, in the estimated threshold parameter ( $\mu$ ), stimuli should be placed at  $\mu$ . On the other hand, if it is important to obtain a good estimate of the spread parameter ( $\sigma$ ), stimuli should be placed at  $1.57\sigma$  on both sides of  $\mu$ . When both  $\mu$  and  $\sigma$  must be estimated, as in the case of determining a psychometric function, a good compromise is to place the stimuli at a distance  $\sigma$  on both sides of  $\mu$  (i.e., at  $X_{15.9}$  and  $X_{84.1}$ ) (Levitt, 1971). It implicitly assumes that two data points are to be used to estimate a psychometric function. Is such an assumption optimal?

For purposes of convenience and speed of calculation, instead of the cumulative normal function, the logistic function as expressed in Eq. (1), a form also used by O'Regan and Humbert, will be used to describe the psychometric function

$$P(X) = \alpha + (\beta - \alpha) \frac{1}{1 + e^{-S(X-M)}}, \quad (1)$$

where  $P(X)$  is the proportion of correct responses at stimulus  $X$ ,  $\alpha$  is the false-alarm probability (of a positive response at a very low stimulus level) (Green, 1993) or equal to  $1/m$  in the  $m$ -alternative forced-choice paradigm,  $\beta$  is the maximum proportion of a correct response (and generally is equal to 1),  $M$  is the midpoint of the psychometric function or the threshold value of the stimulus, and  $S$  is the spread parameter that affects the slope of the function. The parameters  $M$  and  $S$  of Eq. (1) can then be estimated by the probit method, maximum likelihood method, or nonlinear least-square method.

It is difficult, if not impossible, to determine experimentally or analytically the minimum number of data points, the

optimal locations of the data placement, and the number of trials per datum point for estimating a psychometric function. Thus, a series of computer simulations was carried out in order to answer the aforementioned questions. In the computer simulations, it is assumed that the responses can range from 0% to 100% for a yes-no paradigm, and from  $100/m\%$  to 100% for an  $m$ -alternative forced-choice paradigm. Since Lam *et al.* (1996) have shown that a 4-point sampling method can result in psychometric functions similar to those from an 11-point constant stimuli method, it is not necessary to consider the case when the number of data points is greater than 4, at least for the yes-no paradigm. The smallest number of data points is 2 because, by assuming  $\beta=1$  and  $\alpha=0$  for the yes-no paradigm or  $\alpha=1/m$  in the case of  $m$ -alternative forced-choice, there are two parameters in the psychometric function of Eq. (1) that need to be estimated. The number of trials per datum point was varied from 10 to 50 (or higher in order to achieve parameter estimates within 5% of the known parameter values) in steps of 10. From Fig. 1, it seems that the  $(X_{16}, X_{84})$  data placement, which corresponds to  $(\mu \pm \sigma)$ , would result in good estimates of  $M$  and  $S$  for the 2-point sampling method. However, it is not known, *a priori*, if such data placements would be optimal. Instead, eight pairs of data placements in the range of  $(\mu - 2\sigma)$  to  $(\mu + 2\sigma)$  were used to generate data in the Monte Carlo simulations for the 2-point sampling method: [1]  $(X_2, X_{98})$  (or  $\mu \pm 2\sigma$ ); [2]  $(X_6, X_{94})$  (or  $\mu \pm 1.57\sigma$ ); [3]  $(X_{12}, X_{88})$  (or  $\mu \pm 1.17\sigma$ ); [4]  $(X_{16}, X_{84})$  (or  $\mu \pm \sigma$ ); [5]  $(X_{20}, X_{80})$  (or  $\mu \pm 0.91\sigma$ ); [6]  $(X_{22}, X_{78})$  (or  $\mu \pm 0.77\sigma$ ); [7]  $(X_{27}, X_{73})$  (or  $\mu \pm 0.61\sigma$ ); and [8]  $(X_{31}, X_{69})$  (or  $\mu \pm 0.5\sigma$ ).

For the 3-point sampling method, the  $X_{50}$  point was added to the sets of the 2-point sampling method, e.g., [1]  $(X_2, X_{50}, X_{98}, \dots)$ , etc.

Data placements for the 4-point sampling method are a combination of data placements from the 2-point sampling method, namely: [1]  $(X_2, X_6, X_{94}, X_{98})$ ; [2]  $(X_2, X_{16}, X_{84}, X_{98})$ ; [3]  $(X_2, X_{31}, X_{69}, X_{98})$ ; [4]  $(X_6, X_{16}, X_{84}, X_{94})$ ; [5]  $(X_6, X_{31}, X_{69}, X_{94})$ ; [6]  $(X_{12}, X_{27}, X_{73}, X_{88})$ ; and [7]  $(X_{12}, X_{31}, X_{69}, X_{88})$ .

In short, the eight pairs of data placements chosen for the 2-point sampling method, together with the  $X_{50}$  point, form all the data placements for the 2-point, 3-point, and 4-point sampling methods. In the computer simulations,  $M=0$  and  $S=1$  (as well as with the appropriate  $\alpha$ ) were used in Eq. (1) to calculate  $P(X_p)$  for the 17 different locations of  $X_p$ . A uniform (0, 1) random number was generated and compared to the corresponding calculated value of  $P(X_p)$ . If the random number was smaller than  $P(X_p)$ , the trial was counted as a correct response. The percent of correct responses for the appropriate number of trials per point (i.e., 10, 20, 30, 40, 50, or higher until the means of the estimates differ from the known parameter values by less than 5%) at each of the data placements was calculated. Such computer-simulated responses were then used to estimate  $M$  and  $S$  for either 2-point, 3-point, or 4-point data placements. This process was repeated 1000 times for each simulation condition. The mean and standard deviation of  $\hat{M}$  and  $\hat{S}$  of the 1000 repeats were then calculated and the means of the estimates compared to the known parameter values.

## A. Methods of parameter estimation

Maximum likelihood and probit methods have been used to estimate the  $M$  and  $S$  parameters. The maximum likelihood method is very time consuming. Instead, the probit method, a linearized logistic function of Eq. (1), as shown in Eq. (2) is preferred (O'Regan and Humbert, 1989)

$$\ln\left(\frac{1-P(X)}{P(X)-\alpha}\right) = aX + b, \quad (2)$$

where  $a = -S$ ,  $b = -S \cdot M$ . The linear least-squares method is then used to estimate  $a$  and  $b$  of the linearized function shown in Eq. (2).

Note that  $P(X)$ , either the computer-generated or the observed probability value at datum placement  $X$ , can be: (1) equal to 1; (2) equal to  $\alpha$ ; or (3) less than  $\alpha$ . Any of these situations causes problems because the logarithmic function (of the probit method) for such situations does not exist. Furthermore, as was pointed out by O'Regan and Humbert (1989), estimates determined by the probit method can be biased, especially when the sample size (number of trials/point) is small. To avoid undefined logarithmic values, the value of  $P(X)$  was modified slightly. If the computer generated  $P(X)$  was equal to 1,  $P(X)$  was set to 0.99999. When  $P(X)$  was equal to or less than  $\alpha$ ,  $P(X)$  was set to  $\alpha + 0.0001$ . The modified  $P(X)$  values were then used in Eq. (2) for linear least-squares estimation of  $a$  and  $b$ . The linear least-squares estimates of  $\hat{S} (= -\hat{a})$  and  $\hat{M} (= \hat{b}/\hat{a})$  were then used as initial estimates in a nonlinear least-squares algorithm. Of course, the original (i.e., unmodified)  $P(X)$  values were used in the nonlinear least-squares estimation process. This linear and then nonlinear least-squares combination overcomes the shortcoming of each of the estimation methods, namely; (1) the nonexistence of the logarithmic function for some data points in the linear least-squares method, and (2) the requirement of initial estimates of parameter values in a nonlinear estimation method.

The means and variances of each of the 1000 estimated parameters  $\hat{S}$  and  $\hat{M}$  were then calculated. The confidence interval of the estimated parameter,  $\hat{S}$ , for example, is

$$\hat{S} \pm 1.96 \times \text{STD}_{\hat{S}} / \sqrt{1000},$$

where  $\text{STD}_{\hat{S}}$  is the standard deviation of  $\hat{S}$ .

## B. Inexact data placements

Because the threshold and slope parameters of a psychophysical task are unknown, the exact placements of stimuli to obtain data for optimal psychometric function estimation would be difficult to determine *a priori*. Therefore, it is of interest to determine the effects of nonexact data placement on the parameter estimates. Instead of using data points symmetric with respect to the threshold (i.e.,  $(M \pm d)$ ) for 2-point sampling, or  $(M, M \pm d)$  for 3-point sampling, where  $d$  is some displacement from  $M$ , one datum point is chosen from one of the eight sets described in the Method section, while the other datum point is chosen from the next adjacent set. For example,  $X_2$  was chosen from [1] (i.e.,  $\mu - 2\sigma$ ) and  $X_{94}$  was chosen from [2] (i.e.,  $\mu + 1.57\sigma$ ), or  $X_{27}$  from [7]

(i.e.,  $\mu - 0.61\sigma$ ) and  $X_{69}$  from [8] (i.e.,  $\mu + 0.5\sigma$ ). The entire set of such inexact data placements for the 2-point sampling method is: [1] ( $X_6, X_{98}$ ), [2] ( $X_{12}, X_{94}$ ), [3] ( $X_{16}, X_{88}$ ), [4] ( $X_{20}, X_{84}$ ), [5] ( $X_{22}, X_{80}$ ), [6] ( $X_{27}, X_{78}$ ), [7] ( $X_{31}, X_{73}$ ), [8] ( $X_{27}, X_{69}$ ), [9] ( $X_{22}, X_{73}$ ), [10] ( $X_{20}, X_{78}$ ), [11] ( $X_{16}, X_{80}$ ), [12] ( $X_{12}, X_{84}$ ), [13] ( $X_6, X_{88}$ ), and [14] ( $X_2, X_{94}$ ). Similarly, the set of inexact data placements for the 3-point sampling method is: [1] ( $X_{12}, X_{69}, X_{88}$ ), [2] ( $X_{16}, X_{69}, X_{84}$ ), [3] ( $X_{20}, X_{69}, X_{80}$ ), [4] ( $X_{22}, X_{69}, X_{78}$ ), [5] ( $X_{23}, X_{69}, X_{73}$ ), [6] ( $X_{27}, X_{31}, X_{73}$ ), [7] ( $X_{22}, X_{31}, X_{78}$ ), [8] ( $X_{20}, X_{31}, X_{80}$ ), [9] ( $X_{16}, X_{31}, X_{84}$ ), and [10] ( $X_{12}, X_{31}, X_{88}$ ). The set of inexact data placements for the 4-point sampling method is: [1] ( $X_6, X_{12}, X_{94}, X_{98}$ ), [2] ( $X_{20}, X_{16}, X_{84}, X_{88}$ ), [3] ( $X_{12}, X_{17}, X_{78}, X_{80}$ ), [4] ( $X_{27}, X_{31}, X_{69}, X_{73}$ ), [5] ( $X_{20}, X_{22}, X_{73}, X_{78}$ ), [6] ( $X_{12}, X_{16}, X_{80}, X_{84}$ ), and [7] ( $X_2, X_6, X_{88}, X_{94}$ ).

## II. RESULTS

As stated previously, the purpose of this study was to determine the minimum number of data points, the optimum placement of the data points, and the minimum number of trials per datum point for estimating the threshold ( $M$ ) and slope ( $S$ ) parameters of a psychometric function.

The computer simulation results for the estimated threshold parameter,  $\hat{M}$ , for different data placements and different trials/point for the 2-point sampling method are shown in Fig. 2 for yes-no, 4-, 3-, and 2-alternative forced-choice paradigms. The results for the corresponding estimated slope parameter,  $\hat{S}$ , are shown in Fig. 3. In both Figs. 2 and 3 (and subsequent figures), the  $x$ -axis indicates the distance (in  $\sigma$  unit) from the threshold (i.e.,  $\mu \pm 2\sigma$ ,  $\mu \pm 1.57\sigma$ , etc.). In all figures, the open circle indicates the mean of the estimated parameters, while the error bars indicate  $\pm$  one standard deviation of the estimated parameter. The known parameter values for the computer simulations are indicated by horizontal lines at  $M=0$  or  $S=1$ . Only simulation results for  $\hat{S}$  are shown in Figs. 4 and 5, respectively, for the 3-point and 4-point sampling methods. The results for  $\hat{M}$  for 3-point and 4-point sampling methods are similar to those of 2-point sampling shown in Fig. 2 and will not be shown separately.

From Figs. 2-5, it is observed that the means of the estimated threshold parameters are close to the known values of  $M=0$  and  $S=1$  for some data placements (i.e., the means of the estimates differ from the known parameter values by less than 5%). In addition, as the number of trials/point increases from 20 to 50, the standard deviation of the estimates decreases. We will discuss separately the number of trials/point, the optimal data placements, and the number of data points needed for estimating the parameters of a psychometric function.

### A. Number of trials per datum point

For the yes-no paradigm (Figs. 2 and 3), 30-50 trials per datum point at appropriate placement of stimuli (more about optimal data placement later) yield threshold and slope parameter estimates whose means differ by less than

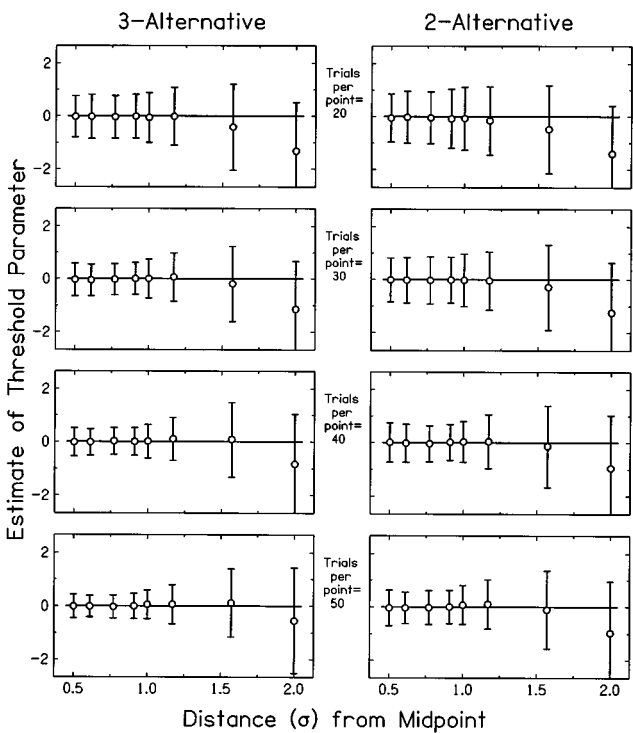
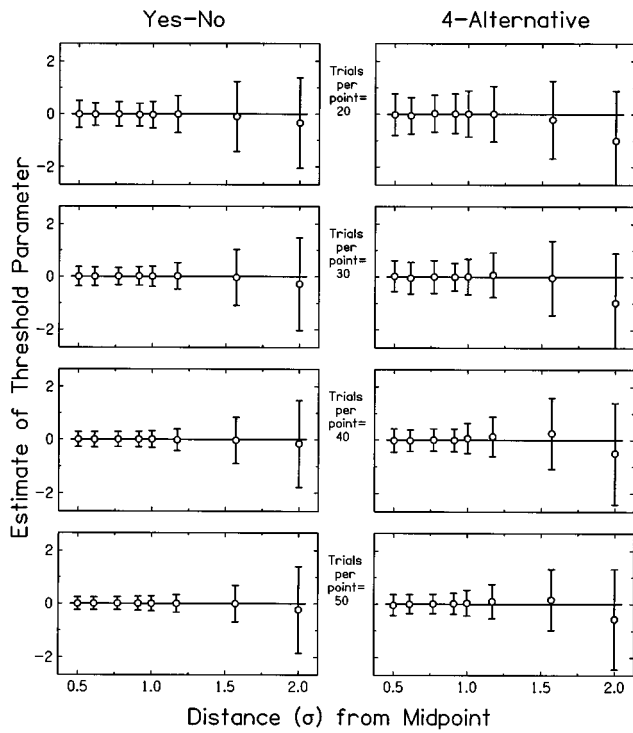


FIG. 2. The means (denoted by o) and means  $\pm$  one standard deviation (denoted by the error bars) of the estimated threshold parameter,  $M$ , of a logistic psychometric function for a 2-point sampling method for yes-no, 4-alternative, 3-alternative, and 2-alternative paradigms. The horizontal line indicates the known parameter value  $M=0$ . The  $x$ -axis indicates the distance (in  $\sigma$  unit) from the threshold.

5% from the known parameter values. The 30–50 trials/point simulation results agree with our previous experimental results (Lam *et al.*, 1996), even though 4 data points were used previously while 2 data points were used in the current Monte Carlo simulation shown in Figs. 2 and 3.

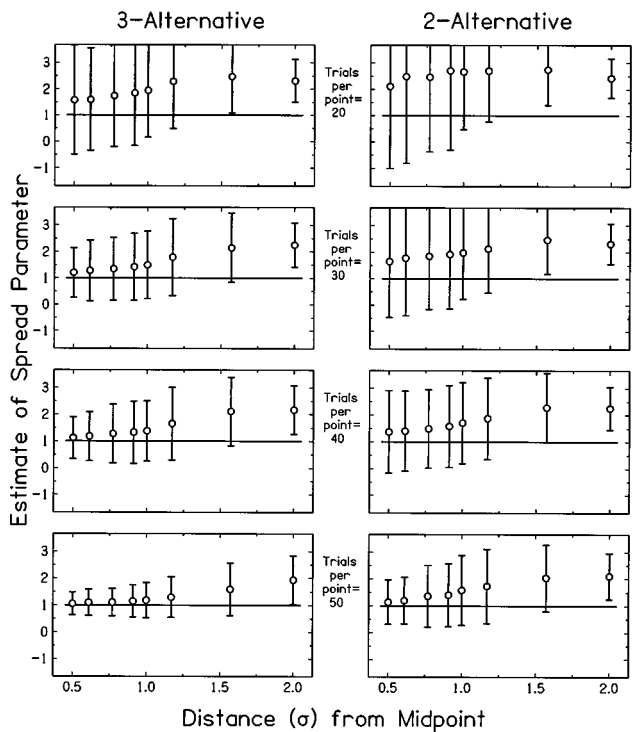
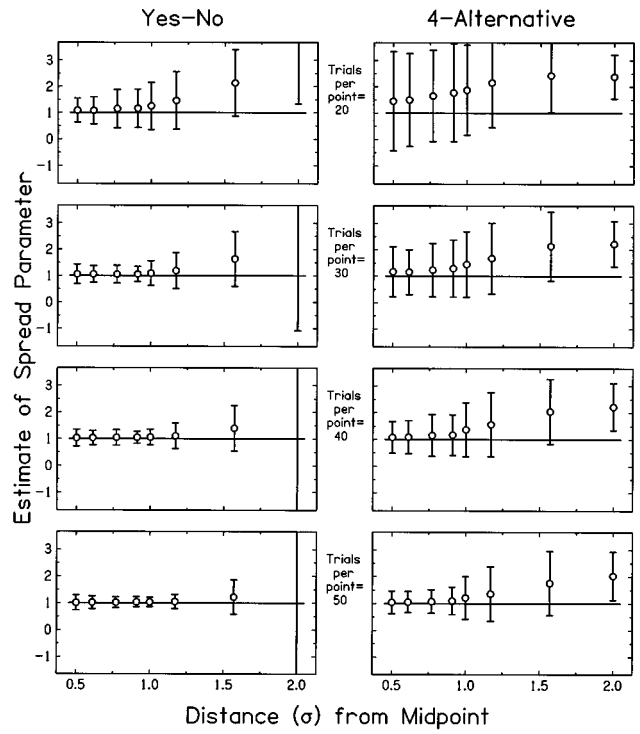


FIG. 3. The means (denoted by o) and means  $\pm$  one standard deviation (denoted by the error bars) of the estimated spread parameter,  $S$ , of a logistic psychometric function for a 2-point sampling method for yes-no, 4-alternative, 3-alternative, and 2-alternative paradigms. The horizontal line indicates the known parameter value  $S=1$ . The  $x$ -axis indicates the distance (in  $\sigma$  unit) from the threshold.

For the  $m$ -alternative forced-choice paradigms (Figs. 2 and 3), 40–50 trials per datum point are adequate for 4- and 3-alternative forced-choice paradigms using the 2-point sampling method. However, for the 2-alternative forced-choice paradigm, at least 120 trials per datum point (not shown in

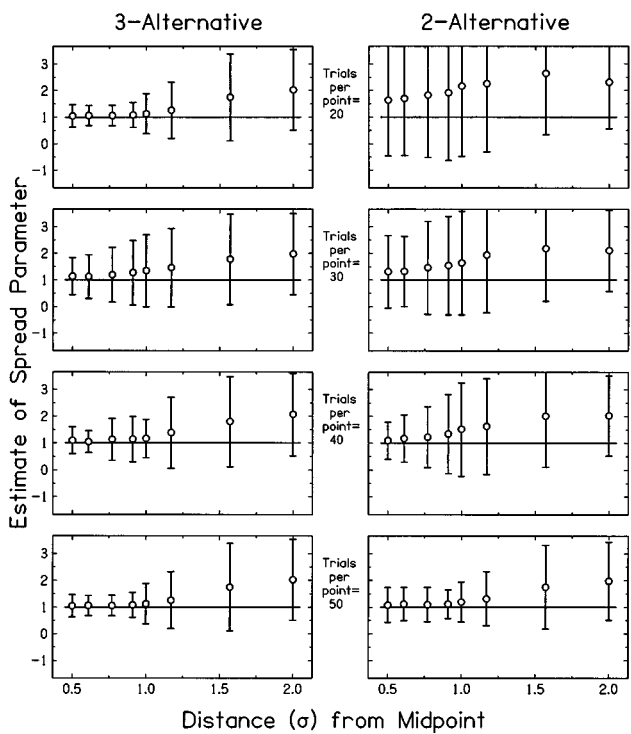
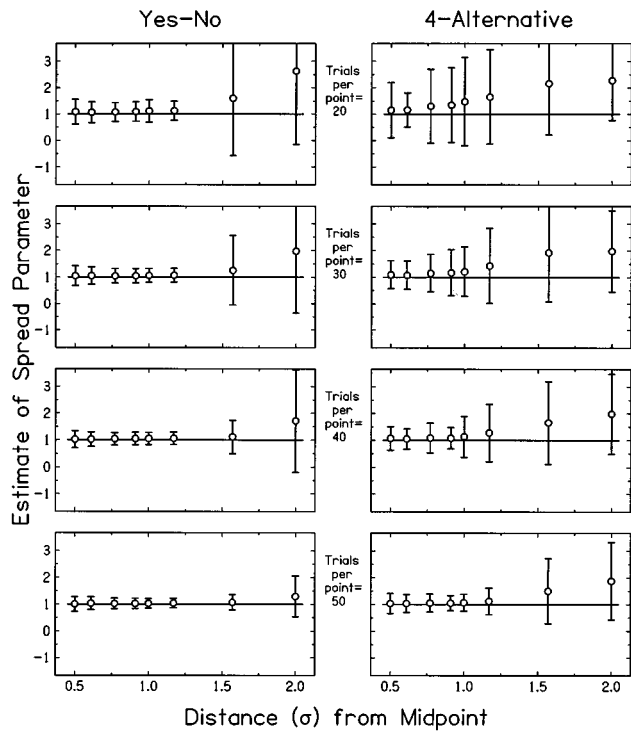


FIG. 4. The means (denoted by  $\circ$ ) and means  $\pm$  one standard deviation (denoted by the error bars) of the estimated spread parameter,  $S$ , of a logistic psychometric function for a 3-point sampling method for yes-no, 4-alternative, 3-alternative, and 2-alternative paradigms. The horizontal line indicates the known parameter value  $S=1$ . The x-axis indicates the distance (in  $\sigma$  unit) from the threshold.

Figs. 2 and 3) are needed even at the best data placement of  $(X_{31}, X_{69})$  for the 2-point sampling method to obtain good parameter estimates, and at least 80 trials per datum point are needed for 3- and 4-point sampling methods (not shown in Figs. 2 and 3).

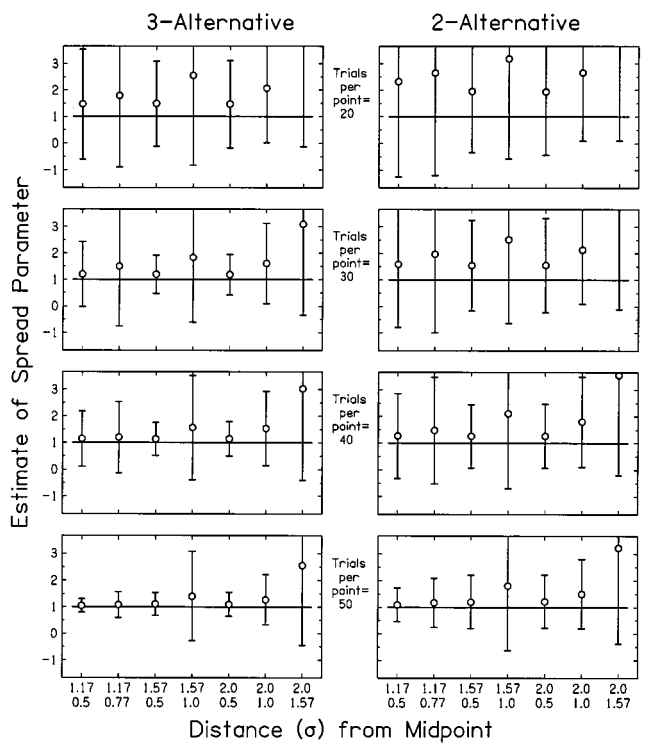
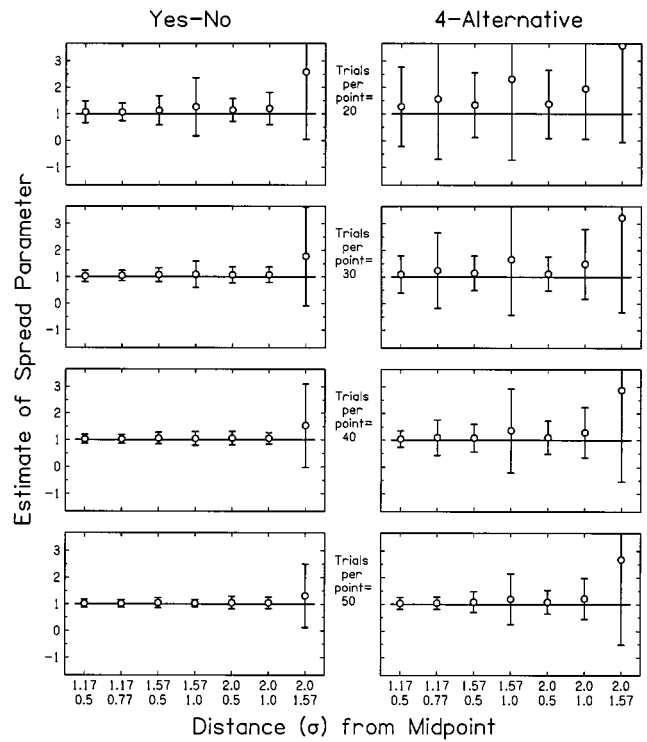


FIG. 5. The means (denoted by  $\circ$ ) and means  $\pm$  one standard deviation (denoted by the error bars) of the estimated spread parameter,  $S$ , of a logistic psychometric function for a 4-point sampling method for yes-no, 4-alternative, 3-alternative, and 2-alternative paradigms. The horizontal line indicates the known parameter value  $S=1$ . The two numbers in the x-axis for each datum point denote the two  $\sigma$  distances. For example,  $(1.17, 0.5)$  denotes  $(\mu \pm 1.17\sigma, \mu \pm 0.5\sigma)$  or  $(X_{88}, X_{69}, X_{31}, X_{12})$ .

## B. The optimal placements of data (stimuli)

From the computer simulation results shown in Figs. 2–5, it is obvious that some data placements yield better parameter estimates than others, regardless of the number of



trials/point, or whether it is a yes–no paradigm or an  $m$ -alternative forced-choice paradigm. With inappropriate placement of stimuli, the estimated  $M$  and  $S$  are very biased, especially for a small number of trials/point and for the  $m$ -alternative forced-choice paradigms. Biases in parameter estimates have been reported in the literature (Rose *et al.*, 1970; O'Regan and Humber, 1989; Schlauch and Rose, 1990). However, the results of this study indicate that with proper data placement, biases in parameter estimates can be minimized. For the yes–no paradigm (Figs. 2 to 5), the optimal data placements are those symmetric with respect to the threshold, with one point in the range of ( $X_{69}$  to  $X_{84}$ ) (or  $\mu + 0.5\sigma$  to  $\mu + \sigma$ ) and another point in the range of ( $X_{16}$  to  $X_{31}$ ) (or  $\mu - 0.5\sigma$  to  $\mu - \sigma$ ), either with (i.e., 3-point sampling) or without (i.e., 2- and 4-point sampling) a point at the threshold. They are ( $\mu \pm \sigma$ ), ( $\mu \pm 0.91\sigma$ ), ( $\mu \pm 0.77\sigma$ ), ( $\mu \pm 0.61\sigma$ ), and ( $\mu \pm 0.5\sigma$ ) of the 2-point and 3-point sampling methods described in the Method section. For 4-point sampling methods, the optimal data sets are ( $\mu \pm 2\sigma$ ,  $\mu \pm 1.57\sigma$ ), ( $\mu \pm 1.57\sigma$ ,  $\mu \pm 0.5\sigma$ ), ( $\mu \pm 1.17\sigma$ ,  $\mu \pm 0.61\sigma$ ), and ( $\mu \pm 1.17\sigma$ ,  $\mu \pm 0.5\sigma$ ). Note that Lam *et al.* (1996, 1997) have used ( $X_{12}, X_{31}, X_{69}, X_{88}$ ) or ( $\mu \pm 1.17\sigma$ ,  $\mu \pm 0.5\sigma$ ) in estimating psychometric functions for various psychophysical tasks.

It is interesting to point out that, for 2-point placement and yes–no paradigms, the stimulus placement at ( $\mu \pm \sigma$ ) (Fig. 3, with 50 points/trial) has the smallest standard deviation of the estimate of the spread parameter, as was suggested by Levitt (1971) and Fig. 1. The result remains about the same for 3-point placement (i.e.,  $\mu$ ,  $\mu \pm \sigma$ ) (Fig. 4). However, for 4-point sampling, the stimulus placements resulting in the smallest standard deviation of the estimate of the spread parameter are at ( $\mu \pm \sigma$ ,  $\mu \pm 1.57\sigma$ ) (Fig. 5).

For  $m$ -alternative forced-choice paradigms, the range for optimal data placement is considerably narrower. Only one data placement with 50 trials/point would result in reasonable estimates of  $M$  and  $S$  for 3- and 4-alternative forced-choice paradigms using 2-, 3-, and 4-point sampling methods, and they are, respectively, [ $X_{31}, X_{69}$ ] (i.e., ( $\mu \pm 0.5\sigma$ )) [ $X_{31}, X_{50}, X_{69}$ ] (i.e., ( $\mu$ ,  $\mu \pm 0.5\sigma$ )), and [ $X_{12}, X_{31}, X_{69}, X_{88}$ ] (i.e., ( $\mu \pm 1.17\sigma$ ,  $\mu \pm 0.5\sigma$ )). As stated previously, however, even at the best data placement (i.e., [ $X_{31}, X_{69}$ ]), at least 120 trials/point are needed in the case of a 2-alternative forced-choice paradigm to obtain parameter estimates of  $S$  whose means differ by less than 5% from the known values.

### C. Number of data points needed per psychometric function

It is well-known that the conventional constant stimuli method of estimating the parameters of a psychometric function is very time consuming. Thus, a 4-point sampling method was proposed by Lam *et al.* (1996, 1997). These current simulation studies were initiated partly to determine whether good parameter estimates can be obtained with an even smaller number of data points. The results shown in Figs. 2–5 indicate that with sufficient number of trials per datum point, say 50, 2 data points are sufficient for the yes–no paradigm. The Student's  $t$ -test indicates that there is no statistically significant difference (at  $\alpha=0.05$ ) between

the best slope estimate of  $\hat{S}=1.010\pm 0.276$  using 2-point sampling method at [ $X_{31}, X_{69}$ ] and that using 3-point sampling method of  $\hat{S}=1.005\pm 0.276$  at [ $X_{31}, X_{50}, X_{69}$ ], where the  $\pm$  value denotes one standard deviation of the parameter estimate. Also, there is no statistically significant difference between the best 2-point estimated  $\hat{S}$  at [ $X_{31}, X_{69}$ ] which is  $\hat{S}=1.005\pm 0.276$  and the best 4-point estimated  $\hat{S}=1.019\pm 0.151$  at [ $X_{12}, X_{31}, X_{69}, X_{84}$ ]. It is interesting to note that the standard deviations of the best estimate of  $\hat{S}$  are about the same between the 2-point sampling method and the 3-point sampling method. However, based on the  $F$  distribution, the standard deviation from the best estimated  $\hat{S}$  using the 4-point sampling method (i.e., [ $X_{12}, X_{31}, X_{69}, X_{84}$ ]) is statistically significantly smaller than that using the 2-point sampling method (i.e., [ $X_{31}, X_{69}$ ]).

For the 4-alternative forced-choice paradigm, there is no statistically significant difference between the means of the best slope estimates using 2-point sampling ( $\hat{S}=1.046\pm 0.416$ ) at [ $X_{31}, X_{69}$ ], using 3-point sampling ( $\hat{S}=1.038\pm 0.376$ ) at [ $X_{31}, X_{50}, X_{69}$ ], and using 4-point sampling ( $\hat{S}=1.040\pm 0.225$ ) at [ $X_{12}, X_{31}, X_{69}, X_{84}$ ]. However, the variance of the best slope estimate using the 2-point sampling method is statistically different from that using the 3-point sampling method. The 4-point sampling method results in the statistically smallest of the three variances from 2-, 3- and 4-point sampling methods.

A similar conclusion applies to the 3-alternative forced-choice paradigm. The means of the best slope estimates are  $\hat{S}=1.039\pm 0.441$ ,  $\hat{S}=1.046\pm 0.420$ , and  $\hat{S}=1.049\pm 0.251$ , respectively, for 2-point, 3-point, and 4-point sampling methods. There is no statistically significant difference among the means of these slope estimates. However, the variance of the best slope estimate from the 4-point sampling method is statistically smaller than that from the 2-point sampling method. On the other hand, there is no statistically significant difference in the variances of the estimates between the 2- and 3-point sampling methods.

### D. Effect of inexact data placement

For the yes–no paradigm, it is not necessary to place the stimuli exactly at the specified locations. In other words, as long as the stimuli are in the ranges of ( $\mu \pm \sigma$ ) to ( $\mu \pm 0.5\sigma$ ), one can obtain good estimates of  $M$  and  $S$ .

For  $m$ -alternative forced-choice paradigms, there is only one inexact data placement that yields a good estimate for 4-alternative and 3-alternative forced-choices with 50 trials/point, which is in the neighborhood of ( $\mu \pm 0.5\sigma$ ) and ( $\mu$ ,  $\mu \pm 0.5\sigma$ ) for 3-point and 4-point sampling methods, respectively. For the 2-alternative forced-choice paradigm, 120 trials/point are needed and only in the neighborhood of ( $\mu \pm 0.5\sigma$ ).

### E. Total number of stimuli intervals per experimental run

In certain experiments, the duration of a psychometric experiment is a consideration in selecting an appropriate paradigm. The total number of stimulus intervals per run is

TABLE I. Total number of stimuli intervals per experimental run for various psychometric paradigms.

	Yes-No	2 AFC	3 AFC	4 AFC
2-point sampling	100	480	300	400
3-point sampling	150	480	450	600
4-point sampling	200	640	600	800

the product of the number of trials/datum point, the number of data points, and the number of stimulus intervals/trial. The number of stimulus intervals per trial is 1 for a yes-no paradigm, and  $m$  for an  $m$ -alternative forced-choice paradigm. With 50 trials/point for yes-no, three- and four-alternative forced-choice, 120 trials/point for two-alternative forced-choice using 2-point sampling, and 80 trials/point for 2-alternative forced-choice using 3-point or 4-point sampling, the total number of stimulus intervals per run for these paradigms is shown in Table I. To achieve good parameter estimates (i.e., estimated parameter values are within 5% of the “true” values), many more trials per datum point are needed for the 2-alternative forced-choice paradigm than for the 3- or 4-alternative paradigms. As a result, the total number of stimulus intervals (hence, the duration of an experiment) is about the same for the 2-, 3-, and 4-alternative forced-choice paradigms, but much more than for the yes-no paradigm. It is obvious that the time it takes the computer to present each of the  $m$  alternatives is much shorter than the response time of the subject. Nonetheless, assuming the subjects’ response times for various experimental conditions are roughly the same, the total time to complete an  $m$ -alternative paradigm would still be longer than that of a yes-no paradigm because the former requires many more stimulus intervals per run.

### III. DISCUSSION

Psychometric functions have been used to summarize different psychophysical tasks in a variety of settings. Using a constant stimuli method, it is very time consuming to collect data for psychometric function determination if a large number of stimuli values are sampled. We sought to determine the minimum number of data points, the optimal placement of the data points, and the minimum number of trials per datum point needed to accurately estimate the parameters of a psychometric function represented by a logistic function in the form of Eq. (1). For each of these combinations, the parameter values of  $M$  and  $S$  were estimated from the computer-generated data using linear and then nonlinear regression. The simulation process was repeated 1000 times each to obtain the means and standard deviation of the estimated  $\hat{M}$  and  $\hat{S}$ .

The computer simulation results show that for the yes-no paradigm: (1) there are no differences in the estimated parameter values ( $\hat{M}$  and  $\hat{S}$ ) and their standard deviations between the 2-point sampling method and the 3-point sampling method at the optimal data placements (i.e., the parameter estimate that is closest to the known parameter value) of  $[X_{31}, X_{69}]$  and  $[X_{31}, X_{50}, X_{69}]$ , respectively; (2) the 2-point sampling method gives parameter estimates ( $\hat{M}$  and

$\hat{S}$ ) as accurate as the 4-point sampling method, but with a statistically significantly larger standard deviation in the estimated parameter values; (3) in general, the optimal range for stimuli placement is relatively wide, with the best stimuli placement for the 2-point sampling method at  $[X_{31}, X_{69}]$ , and at  $[X_{12}, X_{31}, X_{69}, X_{84}]$  for the 4-point sampling method; and (4) the minimum number of trials per datum point is between 30–50, which agrees with our previous experimental results (Lam *et al.*, 1996).

The optimal range for stimuli placement for the  $m$ -alternative forced-choice paradigm is much more restricted. Out of the seven (for 4-point sampling) or eight (for 2- and 3-point sampling) data locations, only one each yields estimated parameter values within 5% of the known parameter values. They are  $[X_{31}, X_{69}]$ ,  $[X_{31}, X_{50}, X_{69}]$ , and  $[X_{12}, X_{31}, X_{69}, X_{84}]$ , respectively, for 2-point, 3-point, and 4-point sampling methods, with the 4-point sampling method yielding the smallest standard deviation of the estimates. At least 80 trials/point are needed for 3-alternative forced-choice paradigm, and 120 trials/point for the 2-alternative forced-choice paradigm.

Green and Swets (1966), and others, have indicated that the  $m$ -alternative forced-choice techniques are theoretically better than the yes-no procedure because in the latter situation, test subjects are free to set their own criterion in selecting a “yes” response. However, our computer simulation results suggest that when it is not necessary to use a criterion-free paradigm, the yes-no paradigm is a better procedure for obtaining the psychometric function than the  $m$ -alternative forced-choice paradigm because: (1) the standard deviation of the estimates from the yes-no paradigm is statistically smaller than that from the  $m$ -alternative forced-choice paradigm; (2) a smaller number of trials/point is needed; and (3) the range of the location for data placement is much wider, as shown in Figs. 3–5. Rose *et al.* (1970), in comparing the staircase estimates of yes-no and 2-alternative forced-choice paradigms, also concluded from their simulation results that the yes-no staircase estimates are better than those from the 2-alternative forced-choice staircase paradigm because: (1) the 2-alternative forced-choice estimates are more variable than the yes-no estimates; and (2) the yes-no estimator is unbiased.

Schlauch and Rose (1990) concluded from their comparison of 2-, 3-, and 4-interval forced-choice staircase procedures that as the number of alternatives is increased from 2 to 4, the variability of the threshold estimates decreases or remains constant and the accuracy of the estimator improves in most cases. The standard deviations of the estimates from our study also decrease from 2- to 3- to 4-alternative forced-choice and to yes-no paradigms. For example, for 50 trials/point and 2-point sampling, the standard deviations are, respectively, 0.664, 0.467, 0.386, and 0.285 for  $\hat{M}$ , and 0.819, 0.869, 0.507, and 0.174 for  $\hat{S}$ , respectively, for 2-, 3-, 4-alternative forced-choice and yes-no paradigms. The same trend of decreasing standard deviations also occurs for 3-point and 4-point sampling. The decrease in standard de-

violation from our study from 2- (to 3-, to 4-) alternative forced-choice to the yes–no paradigm also agrees with the conclusion of Rose *et al.* (1970) that 2-alternative is more variable than the yes–no paradigm. Thus, it seems that the yes–no paradigm has some advantages over the *m*-alternative forced-choice paradigms, especially the more popular 2-alternative case.

Many researchers (Rose *et al.*, 1970; O'Regan and Humbert, 1989; Schlauch and Rose, 1990) have indicated that there are biases in the values of the parameters estimated from the *m*-alternative forced-choice paradigm, especially with small sample sizes. We have also found biases in the parameter estimates at some stimulus placements and/or using small number of trials per datum point. However, such biases can be minimized when the stimuli are placed at certain optimal locations and with 30–50 trials per point (except 80–120 trials per point for the 2-alternative forced-choice situation). Another reason that biases in the parameter estimates are minimized is that the parameters were estimated by curve fitting (rather than by averaging the reversals for threshold estimate, for example), as was pointed out by Schlauch and Rose (1990).

Because the parameter estimates have smaller standard deviations than those using a 2- or 3-point sampling method, the yes–no paradigm using the 4-point sampling method should probably be the method of choice, especially when it is not necessary to use a criterion-free paradigm.

Finally, it is assumed that the experimenter has some idea about the general shape of the psychometric function for the task at hand (via a pilot study, for example) in order to make use of the optimal stimulus placements determined from this study. Of course, if the adaptive up–down method of Leek *et al.* (1992) and Dai (1995) is used to obtain the psychometric function, the appropriate up–down paradigm should be used to obtain data points as close to the optimal stimulus placements as possible. For example, 4-up–1-down would result in a datum point at  $X_{84}$  which is close to  $X_{88}$ , an optimal placement for 4-point sampling.

## ACKNOWLEDGMENTS

The comments by J. L. Hall and an anonymous reviewer are much appreciated. This research is supported, in part, by Research Grant Nos. P50 DC 00422 and R01 DC 00184 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

- Allen, P., and Wightman, F. (1994). "Psychometric functions for children's detection of tones in the noise," *J. Speech Hear. Res.* **37**, 205–215.
- Berkson, J. (1955). "Maximum likelihood and minimum chi-square estimates of the logistic function," *J. Am. Stat. Assoc.* **50**, 130–162.
- Dai, H. (1995). "On measuring psychometric functions: A comparison of the constant stimulus and adaptive up–down methods," *J. Acoust. Soc. Am.* **98**, 3135–3139.
- Dauman, R., Tyler, R. S., and Aran, J. M. (1993). "Intracochlear electrical tinnitus reduction," *Acta Oto-Laryngol.* **113**, 291–295.
- Dubno, J. R., Lee, F. S., Klein, A. J., Matthews, L. J., and Lam, C. F. (1995). "Confidence limits for maximum word-recognition scores," *J. Speech Hear. Res.* **38**, 490–502.
- Green, D. M. (1993). "A maximum likelihood method for estimating thresholds in a yes–no task," *J. Acoust. Soc. Am.* **93**, 2096–2105.
- Green, D. M., and Swets, J. (1966). *Signal Detection Theory and Psychophysics* (Krieger, Huntington, NY).
- Hall, J. L. (1981). "Hybrid adaptive procedure for estimation of psychometric functions," *J. Acoust. Soc. Am.* **69**, 1763–1769.
- Klein, A. J. (1989). "Assessing speech recognition in noise for listeners with a signal processor hearing aid," *Ear Hear.* **10**, 50–57.
- Lam, C. F., Mills, J. H., and Dubno, J. R. (1996). "Placement of observations for the efficient estimation of a psychometric function," *J. Acoust. Soc. Am.* **99**, 3689–3693.
- Lam, C. F., Mills, J. H., and Dubno, J. R. (1997). "A change of subscripts of the stimulus placements for a four-point sampling method," *J. Acoust. Soc. Am.* **101**, 1742.
- Leek, M. R., Hanna, T. E., and Marshall, L. (1992). "Estimation of psychometric functions from adaptive tracking procedures," *Percept. Psychophys.* **51**, 247–256.
- Levitt, H. (1971). "Transformed up–down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Martin, S. E., Engleman, H. M., Deary, I. J., and Douglas, N. J. (1996). "The effect of sleep fragmentation on daytime function," *Am. J. Resp. Crit. Care Med.* **153**, 1328–1332.
- McDevitt, D. G. (1985). "Beta-blockers and psychometric performance: studies in normal volunteers," *Eur. J. Clin. Pharm.* **28** Suppl., 35–38.
- O'Regan, J. K., and Humbert, R. (1989). "Estimating psychometric functions in forced-choice situations: Significant biases found in threshold and slope estimations when small samples are used," *Percept. Psychophys.* **46**, 434–442.
- Paut, O., Guidon-Attali, C., Viviani, X., Lacarelle, B., Bouffier, C., and Francois, G. (1992). "Pharmacodynamic properties of propofol during recovery from anesthesia," *Acta Anesthes. Scand.* **36**, 62–66.
- Rose, R. M., Teller, D. Y., and Rendelman, P. (1970). "Statistical properties of staircase estimates," *Percept. Psychophys.* **8**, 199–204.
- Sand, T., Bovim, G., Grimse, R., Myhr, G., Helde, G., and Cappelen, J. (1994). "Idiopathic normal pressure hydrocephalus: the CSF tap-test may predict the clinical response to shunting," *Acta Neurol. Scand.* **89**, 311–316.
- Schlauch, R. S., and Rose, R. M. (1990). "Two-, three- and four-interval forced-choice staircase procedures: Estimator bias and efficiency," *J. Acoust. Soc. Am.* **88**, 732–740.
- Wall, M., Maw, R. J., Stanek, K. E., and Chauhan, B. C. (1996). "The psychometric function and reaction times of automated perimetry in normal and abnormal areas of the visual field in patients with glaucoma," *Invest. Ophthalmol. Visual Sci.* **37**, 878–885.
- Wetherill, G. B. (1963). "Sequential estimation of quantal response curves," *J. Royal Stat. Soc.* **B25**, 1–48.



# Methods for estimating the sound pressure at the eardrum

H. Hudde<sup>a)</sup>

Institut für Kommunikationsakustik, Ruhr-Universität Bochum, IC 1/132, D-44780 Bochum, Germany

A. Engel

St. Elisabeth-Hospital, HNO-Klinik der Ruhr-Universität Bochum, Bleichstr. 15, D-44787 Bochum, Germany

A. Ludwig

Fischer-Zoth GmbH, München, formerly Institut für Kommunikationsakustik, Ruhr-Universität Bochum, IC 1/132, D-44780 Bochum, Germany

(Received 16 October 1998; accepted for publication 16 June 1999)

The problem of estimating the sound pressure generated at individual eardrums is systematically investigated. In audiometry, the reference to the pressure at the eardrum is usually realized by using a coupler such as the IEC 711 ear simulator, which is intended to approximate an average ear. The errors caused by individually shaped ear canals are calculated for a typical audiometric earphone (Beyer DT 48) in combination with the IEC 711 ear simulator and with an “ideal” coupler. These errors can reach 15 dB and are clearly more important than deviations of the ear simulator from an average ear. In order to obtain correct estimations, the chain matrices of individual ear canals have to be determined. Best estimates are obtained using the “reflectance phase method,” but the “pressure minima method” also provides surprisingly good results, except in narrow frequency ranges. The reflectance phase method is checked using a physical model of the ear canal and the middle ear. The resulting errors of estimation remain within a limit of 3 dB up to more than 10 kHz.

© 1999 Acoustical Society of America. [S0001-4966(99)00610-4]

PACS numbers: 43.66.Yw, 43.64.Ha, 43.20.Mv [RVS]

## LIST OF SYMBOLS

(an additional index refers to a particular condition)

$A$	area	$m_i$	acoustic mass of the section $i$ of a stepped duct
$c$	free-field speed of sound	$n_i$	acoustic compliance of the section $i$ of a stepped duct
$c_m$	cosine coefficient of a Fourier expansion	$p_D$	sound pressure at the drum (“eardrum pressure”)
$c_{11}, c_{12}, c_{21}, c_{22}$	chain parameters, elements of $C$	$p_{ecE}$	Thevenin source pressure representing a sound field outside the head
$C$	chain matrix of the ear canal between $E$ and $D$	$p_E$	sound pressure at the ear-canal entrance
$c_{t11}, c_{t12}, c_{t21}, c_{t22}$	chain parameters, elements of $C_t$	$p_0$	normal static pressure of the air
$C_t$	chain matrix of the ear canal between $E$ and $D$ including the impedance $Z_t$ of the innermost part	$q_D$	volume velocity at the drum
$f$	frequency	$q_E$	volume velocity at the ear-canal entrance
$f_m$	eigenfrequency of the order $m$	$q_0$	Norton volume velocity source representing an earphone
$f_{\min}$	frequency of a pressure minimum	$r_E$	input reflectance at the entrance of an ear canal
$H_{pE} = p_E/p_{ecE}$	input pressure transfer function of an ear canal	$r_i$	radius of the section $i$ of a stepped duct
$H_{pED} = p_D/p_E$	ear canal pressure-transfer function	$r_0$	radius at the beginning of the innermost part of the ear canal
$H_{qp} = q_0/p_D$	transfer admittance of an earphone put on an ear	$s_m$	sine coefficient of a Fourier expansion
$K = q_0/u_{sp}$	transfer function relating an earphone voltage to the corresponding Norton volume velocity source	$u_{sp}$	earphone voltage
$l_{ec}$	length of an ear canal	$x$	ordinate along the middle axis of an ear canal
$l_{\text{frame}}$	frame length of a spatial Fourier expansion	$Y$	acoustic admittance
		$Z$	acoustic impedance
		$Z_D = 1/Y_D$	eardrum impedance/admittance
		$Z_E = 1/Y_E$	input impedance/admittance of an ear canal
		$Z_{ecE}$	intrinsic impedance of the Thevenin

<sup>a)</sup>Electronic mail: hudde@ika.ruhr-uni-bochum.de



	source representing a sound field outside the head
$Z_t = 1/Y_t$	input impedance/admittance of the innermost part of the ear canal
$Z_{tw} = 1/Y_{tw}$	tube wave impedance/admittance
$\alpha$	real part of the propagation constant $\gamma$ of the ear canal

$\beta$	imaginary part of the propagation constant $\gamma$ of the ear canal
$\gamma$	propagation constant $\gamma$ of the ear canal
$\Delta$	length of a section of a stepped-duct approximation
$\lambda$	wavelength
$\rho_0$	static density of air
$\omega = 2\pi f$	angular frequency

## INTRODUCTION

Audiometric or psychoacoustic measurements usually need a reference to a signal which serves as an input to the ear. The choice of an appropriate input depends on the part of the auditory system to be investigated. A reference to the sound field outside the head is suitable if the ear is regarded as a whole (including scattering and diffraction effects of the head). If the middle ear is to be investigated, the pressure at the drum is the best choice. For examining the inner ear, it is desirable to know the stapes vibration or the vestibular sound pressure. However, input signals in the auditory system of humans that are more central than at the eardrum cannot be monitored *in vivo*, except under particular conditions, such as during otosurgery. Thus, in most situations the eardrum pressure is the most suitable input signal. Actually, one single pressure at the drum does not exist, as the pressure varies considerably over the drum surface at higher frequencies. Therefore, it is necessary to define an "effective pressure"  $p_D$ . We will come back to this item at the beginning of Sec. I.

Recently, Neely and Gorga<sup>1</sup> proposed the acoustic intensity entering the ear canal as a good measure of the sound level that corresponds to hearing. The estimate of the acoustic power flow into the middle ear has the advantage of being almost independent of the location of a measurement in the ear canal because losses in the ear canal are relatively small. But, the criterion for choosing a reference signal is not primarily how easily and uniquely it can be measured, but how closely related it is to hearing.

There is no clear answer to the question of whether pressure or intensity is more closely related to the auditory perception, but in our opinion the pressure at the drum is probably better suited in this respect: A considerable fraction of the incident power is absorbed in the middle ear. In a previously published three-part work,<sup>2-4</sup> two of the authors found a middle-ear efficiency of the middle ear decreasing from optimum values of about 30% to 1% at the highest audible frequencies. Thus, a large fraction of the power flow into the middle ear does not contribute to hearing, but is nevertheless measured in the ear canal. A decision concerning the best choice can only be made on the basis of measurements. The quantity that results in smoother and flatter curves of the threshold of hearing is better suited as a reference.

Usually, earphones are used to deliver the sound because they are easy to handle. Instead of large rooms, comparably simple hearing chambers suffice in this case. Using earphones, furthermore, avoids problems of an insufficient signal-to-noise ratio and of changing the relative positions of

the sound source and ear. In this case, the earphone voltage is usually taken as an input signal. Most of the audiometric earphones are equalized in such a way that they approximate the free-field transfer function of the head from the front. But, considerable differences in the frequency responses of such earphones do exist.<sup>5,6</sup>

In audiometry, the reference to the drum is often approximated by calibrating the equipment by means of an "ear simulator" or an "ear coupler." An ear simulator, for instance IEC 711, is intended to simulate an average ear canal terminated by an average eardrum impedance. Therefore, the voltage measured at the microphone "at the drum" of the ear simulator should correspond to the pressure at the eardrum of an average person using the earphone.

However, the pressure generated at the drum of a subject depends on several individual features, such as the ear-canal shape and length, and the acoustic impedance at the eardrum. Also, the coupling of the sound wave into the ear canal is influenced by individual features of the pinna and the *cavum conchae*. Furthermore, the exact position of the earphone considerably affects the frequency response. Fastl and Fleischer<sup>6</sup> report changes of more than 10 dB due to different positions.

Moreover, it is beyond dispute that none of the couplers in use today is even correct with respect to average ears. Their results are comparable, but are not correct on an absolute scale. This has consequences, for instance, for the fitting of hearing aids. Different corrections have been found empirically. A considerable part of these corrections is probably caused by incorrect couplers. Corrections have been published for particular coupler-earphone combinations<sup>7</sup> and for certain circumstances (otitis media,<sup>8</sup> grommets<sup>9</sup>).

Aiming at better control of the pressure actually generated in an ear canal, so-called "real-ear measurements" or "*in situ* measurements" have come into use, mostly in connection with the fitting of hearing aids. The pressure in the ear canal is directly measured using a probe tube, which is put between the ear mold and the ear-canal wall. The opening of the probe tube is usually placed at a distance of about 5 mm in front of the ear mold. Despite several disadvantages of this method (leakage due to the tube at low frequencies, "undefined" position of the tube opening with respect to the eardrum, risk of miscalibration if the opening comes into contact with the ear-canal wall or cerumen), the real ear measurements have been positively assessed by the majority of users, for instance in Ref. 10.

But, even a pressure correctly measured 5 mm from the mold end differs from the pressure at the drum. This has been thoroughly investigated by Chan and Geisler.<sup>11</sup> These

authors find that up to 6 kHz, a pressure measured at a distance of 6 mm from the “top of the drum” can differ by 8 dB from the “eardrum pressure” which the authors specify at a position between the top of the drum and the umbo. For many applications this accuracy is too poor, including modern hearing aids using an extended frequency range up to 10 kHz.

The estimation or the defined generation of the pressure at the drum is the central task to be solved for the drum-related audiometry. The most important individual feature affecting the eardrum pressure is the length and the shape of the ear canal. Without fairly accurate knowledge of the ear canal, a reliable estimate of the eardrum pressure is impossible.

The ear canal geometry could be derived from a replica made by means of an appropriate plastic material like the one used for ear molds. However, this would have many disadvantages. A cast of the inner ear-canal section near the eardrum cannot be obtained because of the risk of damaging the drum. Thus, the “effective length,” which is the most important parameter of the ear canal, would be only roughly known after taking the replica. Even a good replica would not allow the transformation characteristics to be easily derived. As pointed out by Stinson and Lawton,<sup>12</sup> the determination of the area function with reference to the middle axis of the ear canal is not at all trivial. Furthermore, the unavoidable misalignment between the entrance area actually used during audiometry and the entrance area used for estimating the ear-canal transformation reduces the accuracy.

Instead, it is desirable to use only a single acoustic device to measure the ear-canal transfer characteristics and to generate the specified sound pressure at the eardrum. In fact, such a device has already been developed by Stevens *et al.*<sup>13</sup> However, the technique to be described in this paper differs considerably from the method given in Ref. 13 with respect to both the equipment used and the method of estimating the pressure at the drum. The method of Stevens *et al.* comes closest to our “pressure minima method” (see Sec. II B). We use an acoustical measuring head (AMH) which is able to measure the input impedance and to generate a specified pressure in the ear canal. A possible design of such a measuring head including a loudspeaker and two microphones has been described by the authors in Ref. 14. The advantages of using an AMH, which offers the possibility of measuring the input impedance of an individual ear canal, will become clear in Sec. II C (“reflectance phase method”).

All the results presented in this paper—except for the measurements shown in Fig. 16—have been calculated using the eardrum impedance established in our model.<sup>4</sup> The model is set up as an one-dimensional circuit that comprises only such elements that can be quantified by measurements. A great variety of measurements, mainly focusing on the middle ear, but also referring to the ear canal, was performed. All the different data measured in several preparations are approximated by the model. Therefore, the “model impedance” should be close to an average eardrum impedance. In Ref. 4 a “model ear canal” is also specified which approximately equals the average of the ear-canal data provided by Stinson and Lawton.<sup>12</sup> This model ear canal and the

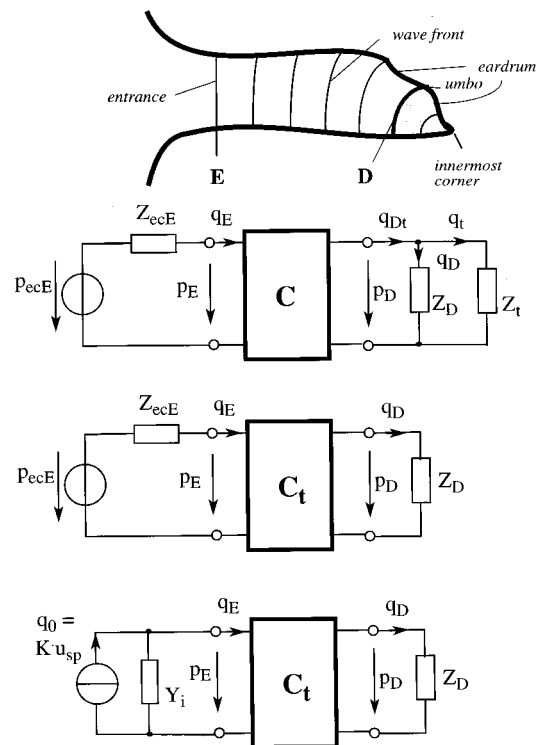


FIG. 1. First panel: Schematic sketch of an ear canal showing some wavefronts and the position of the reference areas for the entrance  $E$  and the drum  $D$ . The section behind  $D$  denotes the innermost part of the ear canal between the umbo and the innermost corner. Second panel: Acoustical circuit representing the ear canal. The sound field outside the head is described by a Thevenin equivalent circuit which consists of the source  $p_{ecE}$  and the intrinsic impedance  $Z_{ecE}$ . The indices denote the reference to the entrance area  $E$  of the ear canal. The impedance  $Z_{ecE}$  equals the radiation impedance out of the ear canal. The two-port  $C$  representing the ear canal between the entrance area  $E$  and the drum reference area  $D$  can be described by the chain matrix  $C$ . The ear canal is terminated by the eardrum impedance  $Z_D$ , which represents the middle ear and the impedance  $Z_t$ , which is the input impedance of the innermost section. Third panel: For convenience, the impedance  $Z_t$  can be treated as a part of the ear-canal matrix ( $C_t$ ). Fourth panel: If an earphone is used, the source is preferably described by a Norton equivalent circuit. Both source parameters, the volume velocity  $q_0$  and the intrinsic admittance  $Y_i$ , must be referred to an appropriate reference area. Here, the entrance area  $E$  is used again. The source volume velocity  $q_0$  is proportional to the earphone voltage  $u_{sp}$ .

individual ear-canal area functions taken from Ref. 12 are used throughout the paper.

## I. EAR-CANAL TRANSFORMATION

The transfer characteristics of an ear canal can be modeled by means of a simple one-dimensional model because the cross-sectional area is small enough to prevent the propagation of higher-order acoustical modes. Using the electroacoustical analogy, the ear canal can be represented by a two-port which links the pressure and volume velocity at the entrance with the corresponding quantities and at the drum.<sup>2-4</sup>

It has already been mentioned that no unique pressure at the drum exists, as the dimensions of the drum are large enough to produce distinct wave effects. As schematically depicted at the top of Fig. 1, the normal to the wavefront of the pressure wave is not perpendicular to the surface of the

drum. For a first approximation, the eardrum serves as a continuation of the ear-canal walls forming an inhomogeneous waveguide. At the highest audio frequencies, this part of the ear canal becomes longer than a quarter wavelength, i.e., the pressure varies along the eardrum from a maximum in the innermost corner via a minimum to a range where the pressure is in opposite phase. This variation of the pressure at the drum also implies that an eardrum impedance  $Z_D$  cannot be defined in the usual way because an acoustical impedance requires an area of constant sound pressure.

In Ref. 2 it was shown that useful definitions of the effective pressure at drum  $p_D$  and the eardrum impedance  $Z_D$  can be found by assuming a load  $Z_D$  concentrated near the umbo (the exact position is not very important). This results in the analog circuit of the second panel in Fig. 1. The ear canal between the entrance area  $E$  and the drum area  $D$  (which is not the surface of the drum, see Fig. 1 for explanation) is an inhomogeneous acoustical duct. As only the fundamental mode can propagate in the audio range, it can be described by a two-port, for instance in terms of the chain matrix  $C$ . Also, the terminating part behind the area  $D$  forms an inhomogeneous duct, but due to its small length the input impedance  $Z_t$  is essentially that of a capacitor representing the volume of this part.

Using temporal bone preparations, it was shown<sup>2</sup> that the transformation of the sound pressure and volume velocity (and thereby of the impedance) inside the ear canal can be fairly well approximated by the given circuit. The impedance derived from measurements at a certain distance from the drum closely approximates a direct measurement of the impedance. "Direct measurement" means that the acoustical measuring device is applied in such a way as to generate sound waves that impinge onto the drum perpendicularly. In this case, the measured impedance is that of the fundamental mode of the sound wave in the measuring tube, and the pressure  $p_D$  obtained by transformation to the drum position is a spatially averaged pressure. This pressure turned out to be useful as the effective pressure at drum  $p_D$  because it is closely related to vibrations of the middle ear. These findings simultaneously solve the problems of defining both the pressure  $p_D$  and the eardrum impedance  $Z_D$ . The usefulness of this notion has also been corroborated in a measurement intended to validate the estimation procedures presented in this paper (see Sec. III).

The chain matrix  $C$  of the ear canal

$$\begin{pmatrix} p_E \\ q_E \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \cdot \begin{pmatrix} p_D \\ q_{Dt} \end{pmatrix} = \mathbf{C} \cdot \begin{pmatrix} p_D \\ q_{Dt} \end{pmatrix}, \quad (1)$$

relates the pressure  $p_E$  and volume velocity  $q_E$  at the entrance to the pressure  $p_D$  at the drum and to the sum volume velocity  $q_{Dt}$  due to the impedances  $Z_D$  and  $Z_t$ . The matrix  $C$  can be calculated from the cross-sectional area function of the ear canal between the entrance and the umbo, specified along the middle axis of the canal.<sup>2</sup>

Only the fraction  $q_D$  of the total volume velocity  $q_{Dt}$  describes the vibrations of the eardrum. Therefore, it is convenient to include the admittance  $Y_t = 1/Z_t$  into the two-port relating ear-canal entrance and eardrum. This is done in the third panel of Fig. 1.

The new matrix

$$\mathbf{C}_t = \begin{pmatrix} c_{t11} & c_{t12} \\ c_{t21} & c_{t22} \end{pmatrix} = \begin{pmatrix} c_{11} + c_{12}Y_t & c_{12} \\ c_{21} + c_{22}Y_t & c_{22} \end{pmatrix}, \quad (2)$$

differs from  $C$  only with respect to the chain parameters  $c_{11}$  and  $c_{21}$ . These two chain parameters are noticeably altered by  $Z_t$  at high frequencies above about 10 kHz only. Using the matrix  $C_t$  and the eardrum admittance  $Y_D = 1/Z_D$  the pressure transfer function from the entrance to the drum is written as

$$H_{pED} = p_D/p_E = 1/(c_{t11} + c_{12}Y_D). \quad (3)$$

The four chain parameters of  $C$  (or  $C_t$ ) are not independent. In acoustical systems, they are always related by the reciprocity condition  $c_{11}c_{22} - c_{12}c_{21} = 1$ . If the ear canal can further be assumed to be lossless, the complete chain matrix is determined by only two elements, one diagonal element and one off-diagonal element. Actually, this fact provides the basis of the reflectance phase method presented in Sec. II C: the input impedance of a rigidly terminated inhomogeneous duct, which is the ratio of two chain parameters, contains all the information concerning the area function.

The chain matrix  $C$  of a simple tube of length  $l_{ec}$  and of constant cross-sectional area  $A$  can be expressed using the tube wave impedance  $Z_{tw}$  and the propagation constant  $\gamma$  of the tube,

$$\mathbf{C} = \begin{pmatrix} \cosh(\gamma l_{ec}) & Z_{tw} \cdot \sinh(\gamma l_{ec}) \\ (1/Z_{tw}) \cdot \sinh(\gamma l_{ec}) & \cosh(\gamma l_{ec}) \end{pmatrix}. \quad (4)$$

As shown in Ref. 3, the propagation losses  $\alpha$  occurring in the propagation constant  $\gamma = \alpha + j\beta$  are about three times as large as in a metal tube of the same dimensions. Nevertheless, the losses remain relatively small compared to the imaginary part of the propagation constant, which approximately equals the wave number  $\beta \approx 2\pi/\lambda = 2\pi f/c = \omega/c$ . For accurate calculations the losses should be taken into account. Also, the tube wave impedance  $Z_{tw}$  is influenced by the losses, but it can be adequately approximated by  $Z_{tw} = \rho_0 c/A$  ( $\rho_0$  is the density of air) in the case of an ear canal.

For inhomogeneous tubes like the ear canal, the simple matrix in Eq. (4) is replaced by a more general form where the hyperbolic sine and cosine terms are warped by the area function to some extent. To evaluate the matrix numerically, the canal is usually divided into a sufficiently large number of sections approximating the actual area function. The total chain matrix is the product of all the partial matrices.

As already mentioned, in common audiometry standardized ear simulators are used to simulate the frequency response of an earphone on an average ear. This means that a fixed transfer function  $H_{up} = (p_D/u_{sp})$  is supposed to relate the input voltage of the earphone to the pressure generated at the drum. Actually, this transfer function depends on the individual ear-canal and eardrum impedance, and on the type and the exact position of the earphone. The errors due to this simplification, exclusively regarding the effect of individual ear-canal shapes, are shown in Fig. 2. In order to calculate the errors, it is necessary to replace the source ( $p_{ecE}, Z_{ecE}$ ) of an external sound field with a source representing the earphone (see bottom panel of Fig. 1). A method to measure

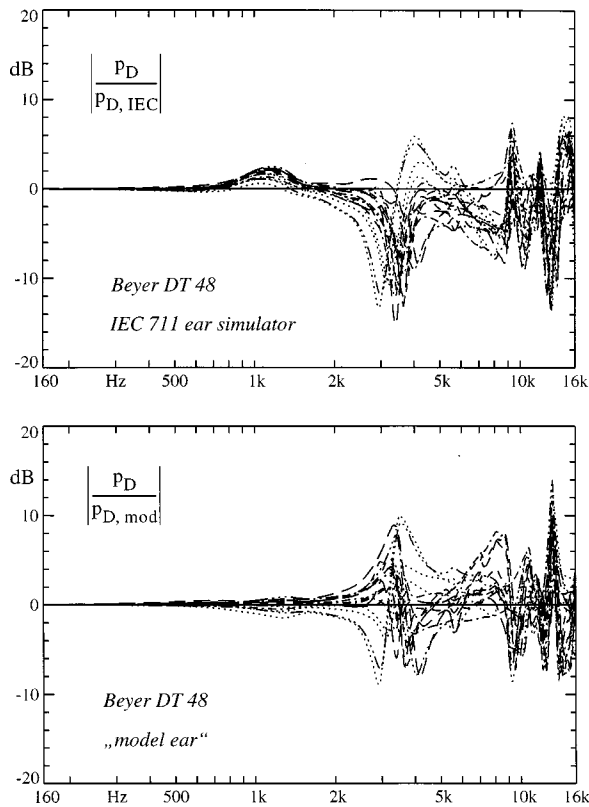


FIG. 2. Effect of individual ear canals on the pressure generated at the eardrum using the earphone Beyer DT-48. In the upper panel the earphone is exactly calibrated to the IEC 711 ear simulator. In the lower panel, the model ear, as specified by our model Ref. 13, is used as a reference which approximates an average ear. The figure shows the errors as ratios,  $(p_D/p_{D,IEC})$  or  $(p_D/p_{D,mod})$ , respectively, of the pressure actually generated at drum referring to the reference case. The errors shown in this figure are exclusively caused by individual ear-canal shapes. Neither individual eardrum admittances nor any effects of the earphone position are taken into account.

these source parameters has been described in Ref. 15. Instead of the Thevenin equivalent source  $(p_{ecE}, Z_{ecE})$  a Norton equivalent source specified by the volume velocity  $q_0$  and an intrinsic shunt admittance  $Y_i$  is used. This choice is not obligatory, but the volume velocity  $q_0$  is more appropriate than the corresponding pressure because the earphones behave more like ideal velocity sources than ideal pressure sources. Since the effect of the acoustical load on the electrical input impedance of the earphone is negligible, the volume velocity generated can be assumed to be proportional to the earphone voltage ( $q_0 = Ku_{sp}$ ).

The earphone volume velocity that generates a certain pressure  $p_D$  at the drum can be expressed as

$$q_0 = H_{qp} \cdot p_D = (Y_i + Y_E) \cdot p_E \\ = (Y_i + Y_E) \cdot (c_{t11} + c_{12}Y_D) \cdot p_D, \quad (5)$$

using the chain parameters of  $C_t$ , the acoustical earphone admittance  $Y_i$ , and the acoustical input admittance  $Y_E$  at the ear-canal entrance

$$Y_E = \frac{c_{t21} + c_{22}Y_D}{c_{t11} + c_{12}Y_D}. \quad (6)$$

The transfer admittance  $H_{qp}$  in Eq. (5) characterizes the resonator established by the ear canal, which is terminated by the eardrum admittance  $Y_D$  and the earphone admittance  $Y_i$  at its ends. This can be seen if the entrance admittance  $Y_E$  occurring in Eq. (5) is replaced by the expression given in Eq. (6). Then, the transfer admittance reads

$$H_{qp} = Y_i c_{t11} + Y_i Y_D c_{12} + c_{t21} + Y_D c_{22}. \quad (7)$$

In the case of an individual ear canal, an earphone volume velocity  $q_0$  generates a certain eardrum pressure  $p_D = q_0/H_{qp}$ . An audiometric earphone is calibrated using an ear simulator, usually the simulator according to IEC 711. In this case, the earphone produces an eardrum pressure  $p_{D,IEC} = q_0/H_{qp,IEC}$ , where  $H_{qp,IEC}$  denotes the transfer admittance of the earphone acting on the IEC 711 simulator.

The resulting deviations, expressed as the pressure ratio

$$\frac{p_D}{p_{D,IEC}} = \frac{H_{qp}}{H_{qp,IEC}}, \quad (8)$$

are shown in the upper part of Fig. 2 using the measured Norton equivalent source of the supra-aural Beyer DT-48 earphone. The chain parameters of the individual ear canals were calculated from the area functions provided by Stinson and Lawton.<sup>12</sup> Individual changes to the eardrum admittances  $Y_D$  and individual effects of the earphone position were not taken into account. Instead, we used the impedance and ear-canal data specified in our model.<sup>2-4</sup> Therefore, the expected errors are in practice higher than those shown in Fig. 2.

The eardrum impedance approximation of the IEC ear simulator is not very realistic and the ear canal is simply approximated by a tube of a constant cross-sectional area. Therefore, we also simulated an "ideal-ear simulator" which would reproduce our model ear, i.e., the transfer admittance of the average ear-canal and the average eardrum admittance. The average ear-transfer admittance  $H_{qp,mod}$  was calculated for the Beyer DT-48 again. The resulting deviations are shown in the lower panel of Fig. 2.

The impact of the ear simulator on the estimated eardrum pressures is surprisingly weak. Only in the important frequency range from 1 to 4 kHz are the errors using the ideal-ear simulator smaller. Similar results, also revealing the relatively small impact of the eardrum impedance, were reported by Gilman and Dirks.<sup>16</sup> This behavior is explained by the favorable influence of the earphone admittance  $Y_i$ . As this admittance is neither very small nor very high compared to the tube wave admittance  $Y_{tw} = A_E/(\rho_0 c)$  belonging to the cross-sectional area  $A_E$  at the ear-canal entrance, the resonator effect of the ear canal, as established in the transfer admittance  $H_{qp}$ , is not very distinct. For the same reason, the type of earphone is of little importance as long as the admittance  $Y_i$  does not become too small.

This consideration suggests an earphone should be utilized that completely absorbs the sound waves which have been reflected at the eardrum end of the ear canal, i.e.,  $Y_i = Y_{tw}$ . In the case of a homogeneous ear canal of length  $l_{ec}$  between the entrance and drum, an ideally absorbing earphone yields a transfer admittance of



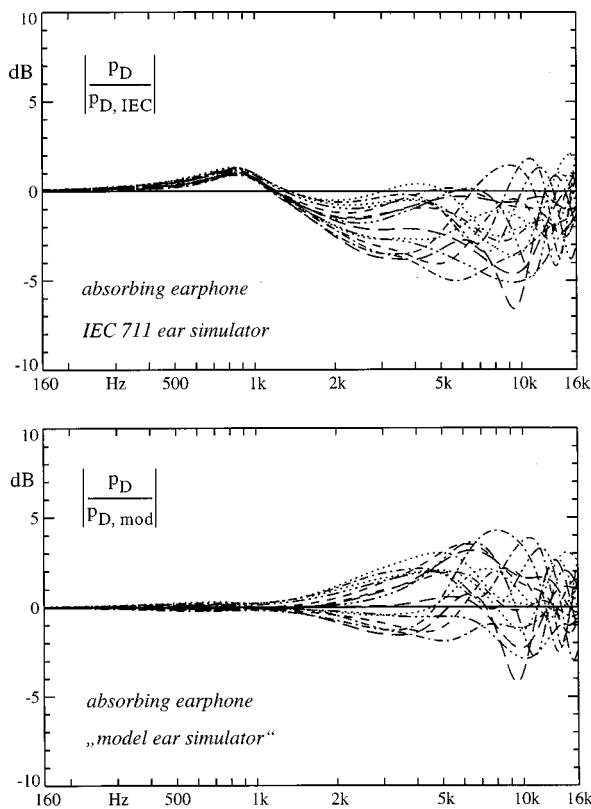


FIG. 3. Effect of individual ear canals, as in Fig. 2, but using an ideally absorbing earphone instead of the Beyer DT-48. Upper panel: IEC 711 ear simulator; lower panel: model ear simulator. Note the change in the scale compared to Fig. 2.

$$H_{qp} = (Y_D + Y_{tw}) \cdot \exp(\gamma l_{ec}). \quad (9)$$

For inhomogeneous ear canals, the result is only slightly different. As the exponential function mainly represents a phase and the eardrum admittances  $Y_D$ , as found in Refs. 2–4, were always smaller than the tube wave admittance  $Y_{tw}$ , the magnitude of the transfer admittance is almost flat and independent of the ear-canal length. For real ear canals with inhomogeneous area functions, this favorable property is maintained to some extent. This can be seen in Fig. 3, which shows the individual deviations referred to the IEC 711 ear simulator and to an ear simulator reproducing the model ear. The resulting errors hardly exceed the limits of  $\pm 5$  dB and could be acceptable in many cases. A slight advantage of using the model-ear simulator exists. An absorbing earphone has been used in the previously cited work of Stevens *et al.*<sup>13</sup> These authors do not calibrate the sound source, but estimate the eardrum pressure in a way similar to our pressure minima method (see Sec. II B).

The deviations reported so far remain incomplete as long as the effects of the unavoidable variations of the position of the earphone are not taken into account. In principle, such variations produce deviations in both source parameters, i.e., in the load impedance  $Y_i$  and in the transfer function  $K = q_0/u_{sp}$  relating the earphone voltage to the volume velocity source changes. The extent of such effects was measured in terms of the source parameters  $q_0$  and  $Y_i$ . The resulting impact on the eardrum pressure calculated for the model ear is shown in Fig. 4. The changes to the source

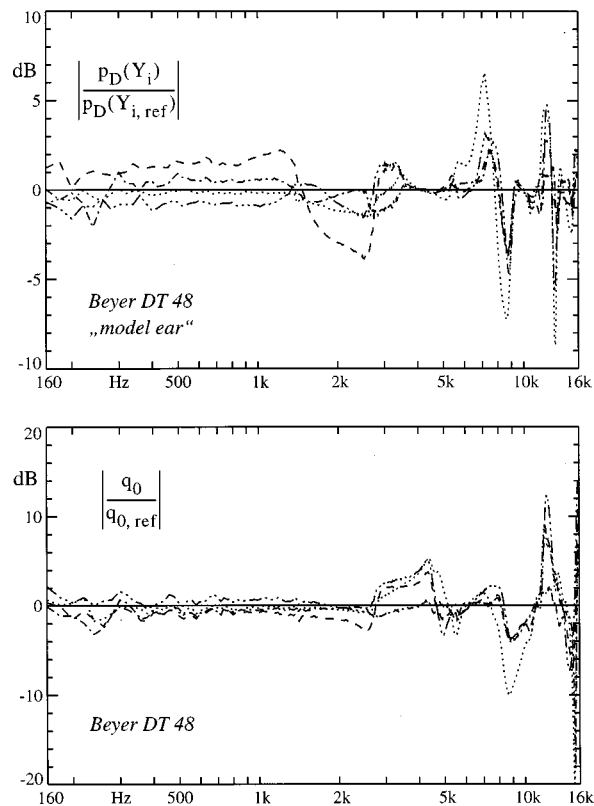


FIG. 4. Effect of the position of the earphone Beyer DT-48, calculated for the model ear on the basis of data measured for the earphone using the method described in Ref. 15. Upper panel: effect of the altered earphone admittance  $Y_i$ ; lower panel: measured change of the volume velocity source. Note the different scales.

volume velocity  $q_0$  need no recalculation because they produce proportional changes to the eardrum pressure. These errors add to all the other deviations. Only at lower frequencies up to 1 kHz can the deviations be accepted; at higher frequencies the errors increase generally. The errors produced by changes in the source volume velocity  $q_0$  dominate. The figure reveals that this error can exceed 5 dB below 5 kHz and 10 dB below 10 kHz.

If the different errors are unfavorably superimposed for certain ear canals and earphone positions, errors of more than 20 dB can result, even below 10 kHz. This is a severe drawback for all audiometric procedures that do not monitor the acoustical quantities actually generated.

## II. ESTIMATION OF THE PRESSURE $p_D$ AT THE DRUM

### A. General considerations

In order to become independent of the properties of audiometric earphones, the estimation of the eardrum pressure must be based on acoustical measurements at the entrance of the ear canal. Here, the pressure  $p_E$  can be measured with little effort using an appropriately calibrated probe tube microphone. To complete the information that can be drawn from acoustical measurements, the volume velocity  $q_E$  must also be measured. This allows the calculation of the input impedance  $Z_E = p_E/q_E$  and the input reflectance

$$r_E = \frac{Z_E - Z_{twE}}{Z_E + Z_{twE}}, \quad (10)$$

where  $Z_{tw}$  denotes the tube wave impedance  $Z_{twE} \approx \rho_0 c / A_E$  at the entrance area  $A_E$ .

The aim of this section is to provide a systematic insight into several possible methods. Therefore, the application of the methods is simulated on the basis of realistic data taken from our model. Only the most promising method, the reflectance phase method presented in Sec. II C, has actually been applied to *in vivo* measurements so far.<sup>17</sup>

The measurement of an input impedance  $Z_E$  in a wide frequency range gives rise to many practical problems. It is outside the scope of this paper to discuss methods for measuring the impedance. Only a few annotations shall be given here. The most severe problem is the exactly specified and constant coupling of the impedance measuring device to the ear canal. The conditions are complicated due to the bend in the ear canal that is frequently found just behind the *cavum conchae*. Therefore, we use an ear mold to realize a well-defined and stable connection between the measuring head and the ear canal. The theoretical background of our method is the same as that used by Keefe *et al.*,<sup>18</sup> Voss and Allen,<sup>19</sup> Sanborn,<sup>20</sup> and also by Neely and Gorga.<sup>1</sup>

It is the main goal of this section to investigate how accurately the eardrum pressure  $p_D$  can be estimated if (a) only the input pressure  $p_E$  (referred to an appropriate reference pressure) and (b) both acoustical input quantities  $p_E$  and  $q_E$  were measured. In principle, both the pressure  $p_D$  and the volume velocity  $q_D$  at the drum can be estimated using the chain matrix  $C_i$  of the ear canal according to Eq. (2),

$$p_{D,est} = c_{22,est} p_E - c_{12,est} q_E, \quad (11)$$

$$q_{D,est} = -c_{21,est} p_E + c_{11,est} q_E. \quad (12)$$

Therefore, the estimation of the chain matrix  $C_{i,est}$  is the fundamental problem to be solved.

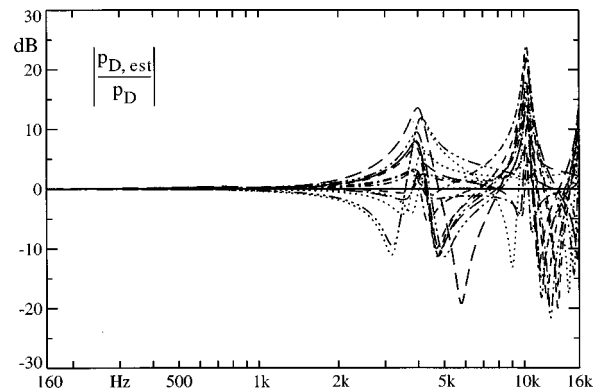


FIG. 5. Estimation of the pressure at the eardrum using the model-ear pressure transformation from the entrance to the drum. In reality, the resulting errors are even larger as no variations of individual eardrum impedances were taken into account.

Let us start by investigating the case of a pure pressure measurement at the ear-canal entrance. In this case, Eq. (11) cannot be applied. Instead, the transfer function of Eq. (3) must be used. This means that not only the ear-canal chain matrix, but also the eardrum admittance  $Y_D$  have to be estimated.

$$p_{D,est} = p_E / (c_{11,est} + c_{12,est} Y_{D,est}) = H_{pED,est} p_E. \quad (13)$$

A very crude estimation is obtained by simply utilizing the average pressure-transfer function  $H_{pED,iv}$ , i.e., that of the model ear, instead of an individual one. However, this produces very large errors (Fig. 5), particularly in the frequency ranges around the  $\lambda/4$  and the  $3\lambda/4$  resonances at about 4 and 12 kHz. This is not surprising, because at these resonance frequencies the transfer function critically depends on the area function of the ear canal. A considerable improvement can be achieved if the ear canal is matched to the measured pressure minima. This is investigated in the next section.

If not only the pressure  $p_E$ , but also the volume velocity  $q_E$  is taken into account by using Eqs. (11) and (12), the

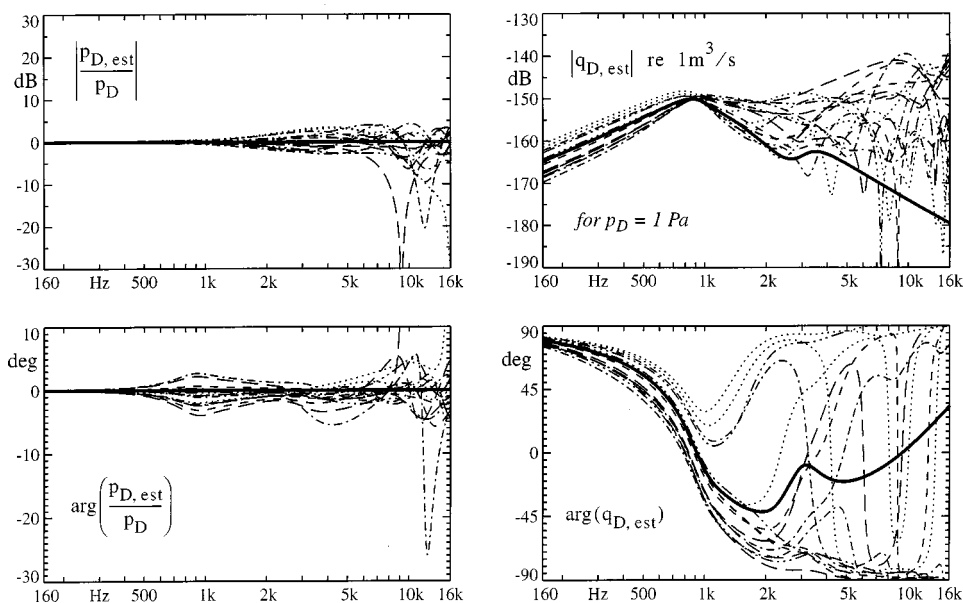


FIG. 6. Estimation of the pressure (left) and the volume velocity (right) at the drum using the model ear-canal chain matrix. For the simulation, an eardrum pressure  $p_D = 1$  Pa and a volume velocity  $q_D = p_D Y_D$  (using the model eardrum admittance  $Y_D$ ) are transformed to the entrance area  $E$  ( $p_E$  and  $q_E$ ) of individual ear canals. These data are transformed back to the drum using the model ear-canal chain matrix. The correct curves are depicted by thick lines.

simple estimation based on an average ear canal yields considerably improved results (Fig. 6). The impact of a wrong ear-canal length is moderated, since no denominator with sharp minima occurs in these equations. Near the  $\lambda/4$  and the  $3\lambda/4$  minimum of the input pressure  $p_E$ , the pressure transformation according to Eq. (13) is very sensitive to errors of any kind. In contrast, the estimate  $p_{D,est}$  according to Eq. (11) is mainly determined by the volume velocity  $q_E$  in that frequency range, because the pressure  $p_E$  is small. The transformation of  $q_E$  into  $p_D$  is as uncritical as the transformation of  $p_E$  into  $p_D$  near the frequencies corresponding to ear-canal lengths of  $\lambda/2$  and  $\lambda$  (see Fig. 5). It is not surprising that large errors can occur at high frequencies if the individual ear-canal length differs greatly from the average one.

In Fig. 6 the volume velocity at drum  $q_{D,est}$  estimated according to Eq. (12) is also represented. At frequencies up to 1 kHz, the errors remain within a tolerance band of 10 dB and  $30^\circ$ . But, at higher frequencies the error in magnitude exceeds 40 dB and the phase seemingly becomes random. Obviously, at higher frequencies the eardrum impedance, which equals the ratio of both estimates, cannot be obtained from the input impedance of the ear canal using this simple method. In Ref. 2 it was even shown that above 2 kHz the accuracy needed for a good estimate of the eardrum impedance is higher than can be realized by means of any known ear-canal estimation technique. Figure 6 reveals that the difficulties in determining the eardrum impedance mainly concern the volume velocity, whereas the eardrum pressure can be estimated fairly well. Due to the high eardrum impedance, the volume velocity  $q_D$  is much smaller than the corresponding velocities at the entrance (except at low frequencies and at the half-wavelength resonance). Therefore, at most frequencies the resulting magnitude of the volume velocity in Eq. (12) is much smaller than both terms on the right-hand side. Small deviations between individual and average chain parameters cause large errors in  $q_{D,est}$ . The magnitude of the eardrum volume velocity is systematically overestimated. This may explain the lower eardrum impedance magnitudes which have been found by many authors, including one of the authors of this paper.<sup>18,19,21-26</sup>

If the sound intensity is to be used as a reference quantity,<sup>1</sup> consequently only the real part of the eardrum impedance will be of interest. In this case, the corresponding component of the volume velocity that is in phase with the pressure  $p_D$  is to be determined. It can be easily obtained from the power flow measured at the entrance and the pressure  $p_{D,est}$  estimated using the methods presented in the following. However, it should be kept in mind that this component of the volume velocity provides little information on the actual velocities at the drum.

Before two possible methods for estimating the eardrum pressure are dealt with in detail, the problem of deriving information on the ear canal from acoustical measurements will be considered from a more general point of view. Equation (6) shows that the admittance measured at the entrance of an ear canal contains information on both the ear canal and the eardrum impedance. But, the effects of the area function, established in the chain parameters, and of the unknown eardrum admittance  $Y_D$  are mingled in the input admittance

$Y_E$ . Separating the two contributions is one of the problems to be solved when the measurements are evaluated in the frequency domain.

Therefore, some authors have tried to measure the area function in the time domain using short pressure pulses. The theory underlying such methods was developed by Sondhi and Gopinath<sup>27</sup> in 1971. The authors derived an integral equation which solves the "inversion problem" of finding the geometry of an acoustical conduit from acoustical quantities. They were interested in the human vocal-tract shape. Much later, in 1993, Joswig<sup>26</sup> applied this method to the ear canal. The main problem of these time-domain methods is the shortness of the time interval that can be evaluated. It is on the order of  $\Delta t = 100 \mu s$ . If this time is exceeded, the influence of the eardrum is mixed with the effect of the ear canal. There are many problems linked to the stability of the sound source and the stability of the deconvolution necessary for removing the impact of the source on the pulse. It seems that an acoustical technique in the time domain is unfavorable because of the shortness of the ear canal.

In the frequency domain, a solution of the inversion problem was found long ago. As early as 1967, Schroeder<sup>28</sup> published a method that relates the shifts of eigenfrequencies of an acoustical duct (he also considered human vocal tracts) to the coefficients of a spatial Fourier expansion of the ear-canal area function.

The eigenfrequencies do not only depend on the ear-canal geometry, but also on the impedances terminating the ear canal at both ends. The zeros of the chain parameter  $c_{11}$  correspond to the eigenfrequencies in the case of an impedance  $Z=0$  at the entrance and a rigid termination ( $Z \rightarrow \infty$ ) at the drum. Similarly, the zeros of  $c_{22}$  correspond to  $Z \rightarrow \infty$  at the entrance and  $Z=0$  at the "drum" end. The chain parameter  $c_{21}$  is associated with infinite impedances, and  $c_{12}$  with zero impedances at both ends. The poles and zeros of the input impedance  $Z_E = c_{11}/c_{121}$  of a rigidly terminated ear canal are identical with the zeros of the chain parameters  $c_{11}$  and  $c_{121}$ , and are therefore identical with the desired eigenfrequencies.

In general, the expansion of an ear-canal area function can be written as

$$A(x) = A_0 \times \left\{ 1 + \sum_m [c_m \cos(m\pi x/l_{ec}) + s_m \sin(m\pi x/l_{ec})] \right\}. \quad (14)$$

Schroeder's calculations were based on a perturbation method applied to a second-order term, the so-called radiation pressure. Hence, the results are only valid for very small deviations from a homogeneous ear canal. Within the bounds of this approach, only the cosine coefficients affect the eigenfrequencies. Following Schroeder's treatment, the relative shifts of eigenfrequencies are simply related to the cosine coefficients according to

$$\Delta f_m / f_m = -c_m / 2, \quad \text{for } m = 1, 2, 3, \dots \quad (15)$$

The odd eigenfrequencies are the zeros of the input impedance  $Z_E$  in the case of a rigid termination; the even eigen-

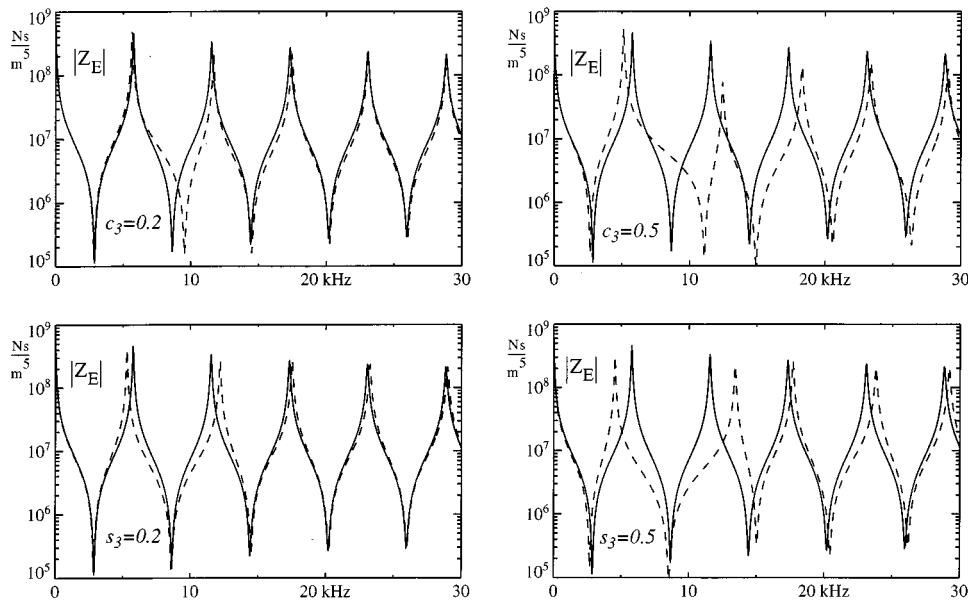


FIG. 7. Shift of eigenfrequencies of an ear canal caused by a nonuniform cross-sectional area. The ear-canal area function is expressed by the Fourier expansion in Eq. (14). The effect of the cosine coefficient  $c_3$  and the sine coefficient  $s_3$  is shown by calculating the magnitude of the ear-canal input impedance  $Z_E$ . For comparison, the impedance of a homogeneous ear canal of the same length is also depicted. For coefficients that are sufficiently smaller than unity, mainly one eigenfrequency (cosine coefficient) or two eigenfrequencies (sine coefficient) are shifted. For larger coefficients the inhomogeneity affects  $Z_E$  in a broad frequency range. To show these effects, an extended frequency range is used.

frequencies are poles. As each Fourier coefficient affects only one eigenfrequency, leaving all the others almost untouched, a single Fourier term only produces changes in a restricted frequency range around the corresponding eigenfrequency. Thus, if the inhomogeneities are sufficiently small, “slower” alterations to the area function only affect lower frequencies of  $Z_E$ , and vice versa.

However, the degree of inhomogeneity in real ear canals is not so small that the perturbation theory is accurate enough. The low-order coefficients actually reach the order of unity. In this case, the inhomogeneity affects a broad frequency range, as can be seen in Fig. 7. The top left-hand panel is in agreement with Eq. (15): the third cosine coefficient  $c_3$  mainly shifts the third eigenfrequency (the second zero), although the value of  $c_3=0.2$  is fairly high. But, the top right-hand panel shows that for  $c_3=0.5$  the inhomogeneity alters  $Z_E$  in a broad frequency range.

The bottom panels reveal that the sine coefficients also have an impact on the input impedance which is similarly pronounced, but differs in detail. As empirically found, small sine coefficients do not shift the eigenfrequencies corresponding to Eq. (15), but they shift the two neighboring eigenfrequencies in opposite directions. The odd Fourier coefficients shift two poles, the even coefficients two zeros (of course, the pole at  $f=0$  Hz cannot be shifted by means of  $s_1$ ). Again, the influence broadens in the frequency range if the value of the sine coefficient is increased.

The question arises as to whether the area function can be uniquely determined from zeros and poles. From the theory used for the time domain, it is clear that the impulse response at the entrance is uniquely related to the area function except for a factor. The factor poses no problem because the absolute cross-sectional area at the entrance can be measured. Hence, disregarding this factor, the impulse response, and therefore also the input impedance, allows a unique determination of the area function. But, when measuring in a bounded working frequency range, details of the area function cannot be resolved. According to Fig. 7, the situation would be simple if the coefficients of the Fourier expansion

were smaller than about 0.2. In this case, the slow fluctuations of an area function could be uniquely derived from measurements in a corresponding frequency range. However, as larger Fourier coefficients can actually occur, the close relationship between the shift of single eigenfrequencies and corresponding Fourier coefficients is weakened.

On the other hand, we are not actually interested in area functions, but only in chain parameters in a restricted frequency range. Therefore, it is not even necessary to obtain the correct area function. As the zeros and poles of the input impedance widely govern the frequency response in the full frequency range, any area function that produces correct eigenfrequencies yields a fairly accurate estimate of the chain matrix. This will be shown later on in Fig. 10. Different frequency responses of the chain parameters outside the working frequency range have no relevance for estimating the eardrum pressure.

Input impedances  $Z_E$  calculated according to our model for various ear canals are shown in Fig. 8. Usually two minima and two maxima can be evaluated. Therefore, only four Fourier coefficients describing the area function can be obtained. “Faster” changes to the area function cannot be resolved. In Fig. 8, the influence of the eardrum impedance  $Z_D$  on the input impedance  $Z_E$  is also elucidated. This is of great importance for estimating the pressure at the drum, because  $Z_D$  is not known *a priori* and cannot be accurately estimated from the input impedance as already mentioned. This item will be discussed at the beginning of Sec. II C.

## B. Pressure-minima method

As the measurement of an acoustic impedance is not a simple task, it is reasonable to look for sources of information which can be obtained with less effort. In this section, it is shown that the frequencies of the pressure minima in the working frequency range suffice to allow a surprisingly good estimation of the ear-canal chain matrix. In real measurements, it is not always clear which of the pressure minima are actually caused by the  $\lambda/4$  and  $3/4 \lambda$  distance to the end



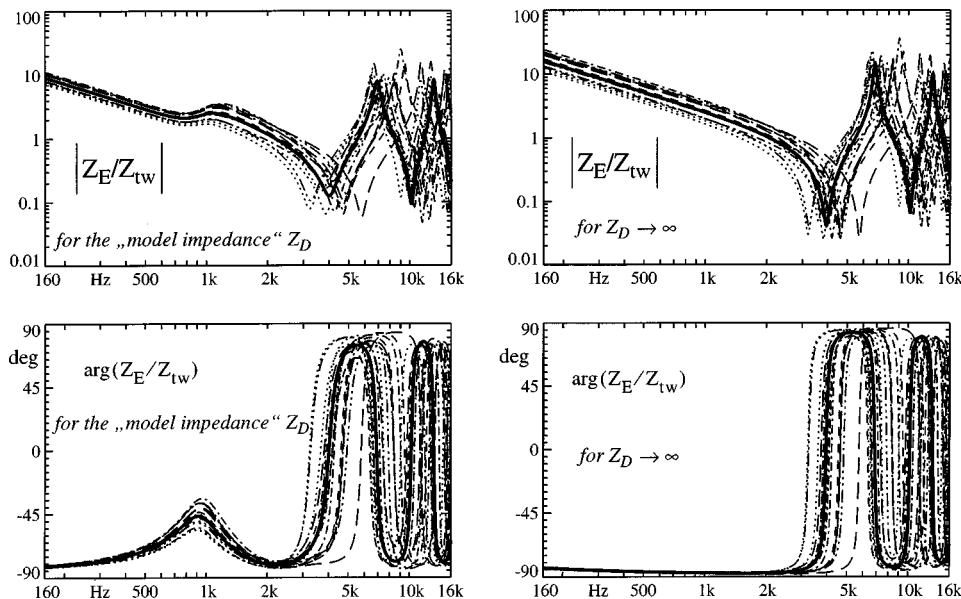


FIG. 8. Input impedances  $Z_E$  at the ear-canal entrance, normalized to the tube wave impedance  $Z_{tw}$  of about  $8 \times 10^6 \text{ Nsm}^{-5}$ . The different impedances correspond to individual ear canals. On the left-hand side the eardrum impedance of our model Ref. 4; on the right-hand side a rigid termination is assumed to terminate the ear canal.

of the ear canal and which are caused by sound-source features and scattering and diffraction effects at the head. This problem can be solved by using the input pressure-transfer function

$$H_{pE} = \frac{P_E}{P_{ecE}} = \frac{Z_E}{Z_E + Z_{ecE}}, \quad (16)$$

which relates the pressure  $p_E$  at the entrance to the external sound source  $p_{ecE}$  (see Fig. 1). Equation (16) is easily derived from the upper circuit in Fig. 1 by using the formula for a voltage divider (here, pressure divider) at the port  $E$ . The transfer function can be measured using only one probe tube microphone. The pressure must be measured in the case of the ear canal rigidly blocked at the entrance ( $p_{ecE}$ ) and in the normal case ( $p_E$ ) preferably using a free sound field. As the same microphone can be used in both cases, no calibration is needed. As only the zeros of  $H_{pE}$  are actually evaluated, moderate errors introduced by possible movements of the subject's head have little impact on the result. We have never performed such measurements, but Hammershoi and Moeller<sup>29</sup> have presented results which clearly show that the pressure zeros can be uniquely found in this way.

Can the two real numbers  $f_{min1}$  and  $f_{min2}$  actually provide enough information to derive the chain parameters needed for estimating the ear-canal transfer function  $H_{pED}$  according to Eq. (3)? From the preceding section, it is known that the two (or three) pressure minima contain only half the information on the area function. One half of the information, namely that of the poles, is missing. But, nevertheless, surprisingly good estimates of the eardrum pressure can be obtained in this way. The estimates are obtained using a gradient method by which the parameters of the area function (length and Fourier coefficients) are altered, aiming at an optimum fit of the zeros. More details of the algorithm are provided in the next section in the context of the reflectance phase method.

In Fig. 9 a typical result is represented. The small panel inserted into the figure shows large deviations between the assumed radius function ("actual") and the estimated one.

The main figure represents the actual and the estimated input pressure-transfer function  $H_{pE}$  and the ear-canal pressure-transfer function  $H_{pED}$  which finally determines the pressure estimated at the drum. In spite of the poor estimation of the radius function, the error of estimating  $H_{pED}$  is smaller than 1 dB at most frequencies. Only near the eigenfrequencies can the error rise to higher values.

As expected, the calculated radius function is not at all unique. Figure 10 shows several estimates of the same radius function, which is depicted using a thicker line. The different solutions were obtained by using different initial values of the ear-canal parameters for the gradient method. All the solutions perfectly reproduce the first two zeros of the input impedance within the frequency resolution of 20 Hz used in the measurements and simulations. The lower panel represents the corresponding errors of the pressure estimated at the drum. In spite of the widely diverging radius functions, the errors are small. The increase in the error magnitude at high frequencies is produced by the third zero, which is only a little outside the working frequency range. This third zero was not taken into account in the estimation procedure.

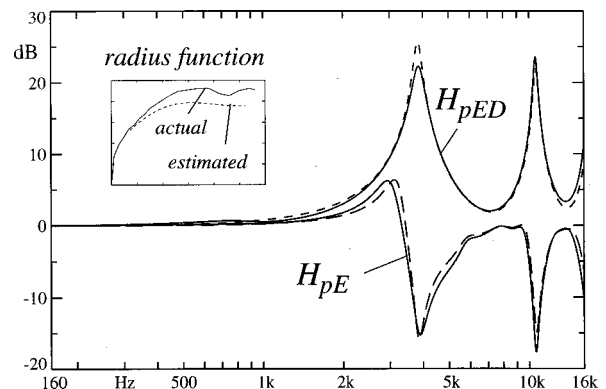


FIG. 9. Estimation of the eardrum pressure using the pressure-minima method. The measured and the approximated input transfer function  $H_{pE} = p_E/p_{ecE}$  and the corresponding ear-canal pressure-transfer functions  $H_{pED} = p_D/p_E$  are depicted. The small figure shows the presumed ear canal (actual) and the estimate.

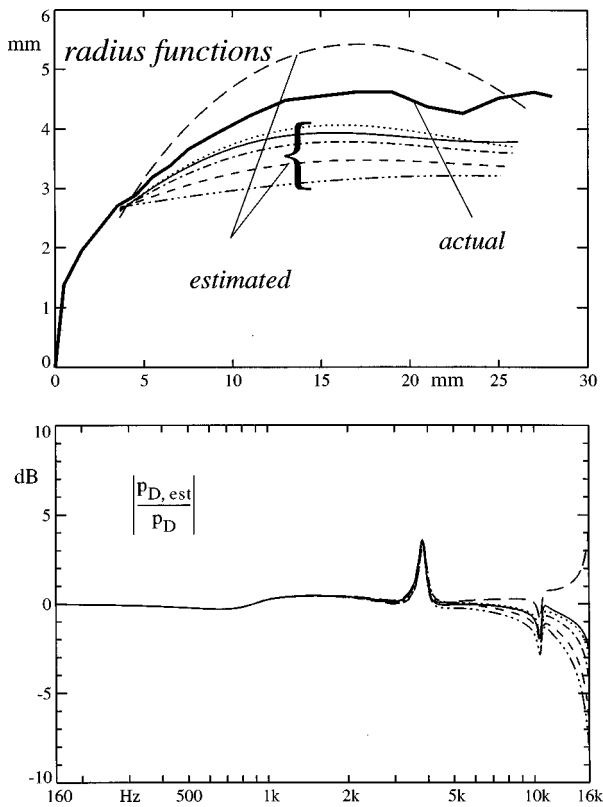


FIG. 10. Although very different radius functions which almost exactly fit the pressure minima can be calculated, the estimates of the pressure  $p_{D,est}$  at the drum are fairly correct. The stronger deviations at high frequencies are a result of the nearby third pressure minimum, which was not taken into account when estimating the radius function.

The good estimation of the pressure  $p_D$  can be explained as follows: According to Eq. (3), the pressure-transfer function  $H_{pED}$  is the inverse of  $(c_{i11} + c_{i12}Y_D)$ . At low frequencies, the chain parameter  $c_{i11}$  predominates because the chain parameter  $c_{i12}$  is small. With increasing frequency the magnitude of  $c_{i11}$  decreases, while that of  $c_{i12}$  increases, but the magnitude of  $Y_D$  also decreases. Therefore, except near the zeros of  $c_{i11}$  this chain parameter dominates, and the transfer function is approximated by  $1/c_{i11}$ . Consequently, merely the chain parameter  $c_{i11}$  has to be accurately determined if only the pressure at the drum has to be estimated. Fortunately, this chain parameter is mainly governed by the zeros of  $Z_E$  and is hardly affected by its poles in the case of a rigid termination ( $Z_E = c_{i11}/c_{i21}$ ).

As already mentioned, the pressure-minima method has some relationship to the work of Stevens *et al.*<sup>13</sup> Also, these authors exploit minima of the pressure generated at the ear-canal entrance. Their technique is based on an approximately absorbing source realized by a sound delivery tube of about 60 cm length and 6.3 mm diameter. In our notation, they evaluate the transfer function  $p_E/u_{sp}$  between the earphone voltage and the pressure at the ear-canal entrance. In contrast to the procedure described in this paper, Stevens *et al.* do not try to obtain an estimate of the area function, but use a general all-pole function to estimate the pressure-transfer function  $H_{pED}$ .

The quality of the estimations, as represented in Figs. 9 and 10, is typical for the pressure-minima method. However,

for particular area functions worse results can also occur, reaching peak errors of up to 8 dB within small frequency ranges around the zeros. Therefore, the pressure-minima method is not sufficiently exact for many applications. The optimum accuracy can only be obtained by measuring and evaluating both the pressure  $p_E$  and volume velocity  $q_E$  at the ear-canal entrance.

### C. Reflectance-phase method

After measuring the impedance  $Z_E$ , the complete set of zeros and poles within the working frequency range is known and can be evaluated with respect to the area function of an ear canal. Matching only the poles and zeros has the disadvantage that disturbances which affect the measured impedance near these discrete eigenfrequencies may have negative effects on the result. Therefore, a continuous frequency response should be evaluated.

Figure 8 shows that both the extrema of the magnitude and the zero crossings of the phase of the input impedance  $Z_E$  mark the eigenfrequencies. Thus, the question arises as to whether it is more favorable to evaluate the complete complex input impedance, or only parts of it, or perhaps a particular mapping of it. The most suitable function is the one that is least influenced by the eardrum impedance, but contains all the relevant information on the area function. In order to examine this feature, the input impedance is calculated for the case of the model eardrum impedance  $Z_D$  terminating the ear canal, as well as for an ideally reflecting termination ( $Z_D \rightarrow \infty$ ). The effect of  $Z_D$  is predominantly observed in the magnitude of the input impedance, mainly at low frequencies up to 2–3 kHz. At higher frequencies, the impact of  $Z_D$  becomes generally very slight. Using the only little-affected phase of  $Z_E$  as an objective function for the estimation procedure yields good results (not shown).

Even better estimations can be obtained by evaluating the phase of the reflectance  $r_E$ , as it remains almost unaffected by the eardrum impedance at higher frequencies. This favorable property is caused by the high reflectivity of the eardrum impedance in this frequency range. Therefore, the reflectance magnitude at the entrance also approaches unity except around the main middle-ear resonance (Fig. 11). The reflectance phase is mainly determined by the linearly decreasing fraction due to the ear canal length  $l_{ec}$ . To focus on the relevant part of the reflectance phase, the phase difference to  $-2\beta l_{ec}$  is depicted in Fig. 11, using a mean value of the ear-canal length  $l_{ec}$ . The impact of the middle-ear resonance on the reflectance phase is clearly seen, but for frequencies above 3 kHz the agreement between the phase with and without the eardrum load is almost perfect. Thus, the effect of the area function is concentrated and almost undisturbed by the eardrum impedance in the reflectance phase above 3 kHz. Therefore, the reflectance phase is ideally suited to deriving the area function. A typical ‘‘unwrapped’’ phase response of the reflectance and its relationship to the impedance magnitude are shown in Fig. 12. The zeros of  $Z_E$  correspond to odd multiples of  $-\pi$ , the poles to even multiples of  $-\pi$ . This figure shows that the eigenfrequencies, i.e., the frequencies where  $\arg(r_E) = -n\pi$  ( $n = 1, 2, 3, \dots$ ), largely determine the complete phase response.

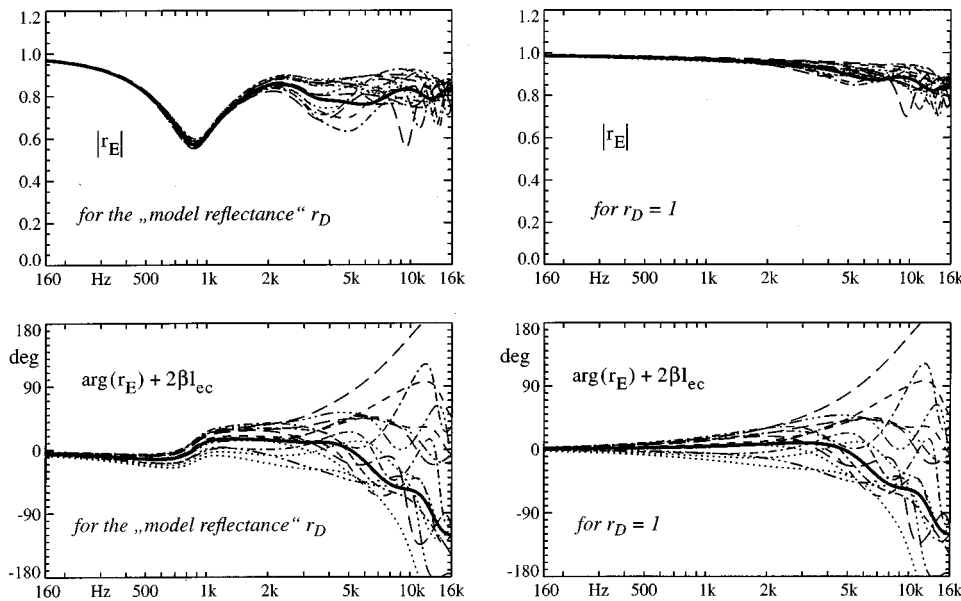


FIG. 11. Magnitudes (upper panels) and phases (lower panels) of the input reflectances  $r_E$  calculated for individual ear canals. The original phases of the reflectances run through a wide range because of the fraction  $2\beta l_{ec}$  which represents the delay of a homogeneous ear canal of length  $l_{ec}$ . The fluctuations of the phase due to the inhomogeneity of ear canals and due to the termination reflectance become more visible when the phase is reduced by the delay term. On the left the model eardrum reflectance  $r_D$  is used; on the right a perfectly rigid termination ( $r_D = 1$ ) is assumed instead.

The only input to the gradient algorithm used for estimating the radius function is the phase  $\arg(r_E)$  of a measured (or simulated) input reflectance of an individual ear. The algorithm calculates the squared phase difference to the reflectance phase of a “comparative ear,” whose ear canal can be numerically altered. The comparative ear comprises an ear canal (usually starting with a homogeneous one) which is terminated by a load. We used our model impedance as a load (reflectance  $r_D$  corresponding to  $Z_D$ ) or alternatively completely omitted the load ( $r_D = 1$ ). The phase differences at all the discrete frequencies are added up beginning at 3 kHz to avoid the impact of the load. For this reason, even impaired middle ears should not have a noticeable impact on the estimation.

It is important to use the unwrapped phase, as depicted in Fig. 12, instead of the phase wrapped in a range of the width  $2\pi$ . In the latter case, discontinuities of  $2\pi$  arise that produce seemingly large phase deviations, although the unwrapped phases almost agree. Usually the unwrapped phase can be uniquely determined. Difficulties can only occur if the reflectance magnitude becomes very small. In this case, small changes can induce the reflectance to circle around the origin of the complex plane or not, which can make the unwrapping procedure ambiguous. In the case of the input reflectance of an ear canal, this problem should never arise as the reflectance magnitude is large enough above the main middle-ear resonance (see Fig. 11).

The rather sharp decrease in size of the cross-sectional area at the end of the ear canal cannot be derived from the measured input impedance. In order to achieve a robust convergence of the estimation algorithm, it is necessary to prescribe the area function in the innermost part in an appropriate way. The part of the radius function which is identical to the one taken into account by the impedance  $Z_i$  in Fig. 1 was simply assumed to decrease linearly from  $r_0 = 2.5$  mm to zero over a distance of 4 mm. This fixed choice based on the data of Stinson and Lawton<sup>12</sup> is only a very rough approximation, but turned out to be completely sufficient for the given purpose.

The ear canal between  $E$  and  $D$  has to be described by a Fourier series. Therefore, the radius at the end of this section must equal  $r_0 = 2.5$  mm to obtain a continuous function. If the origin of the  $x$ -axis is chosen in  $D$ , this means  $A(0) = A_0 = \pi r_0^2$ . This boundary condition is easily realized by using only the sine terms of the Fourier expansion in Eq. (14). Omitting the cosine terms does not impose an inadmissible restriction on possible area functions, but even supports a good convergence, because the continuation of the inner-

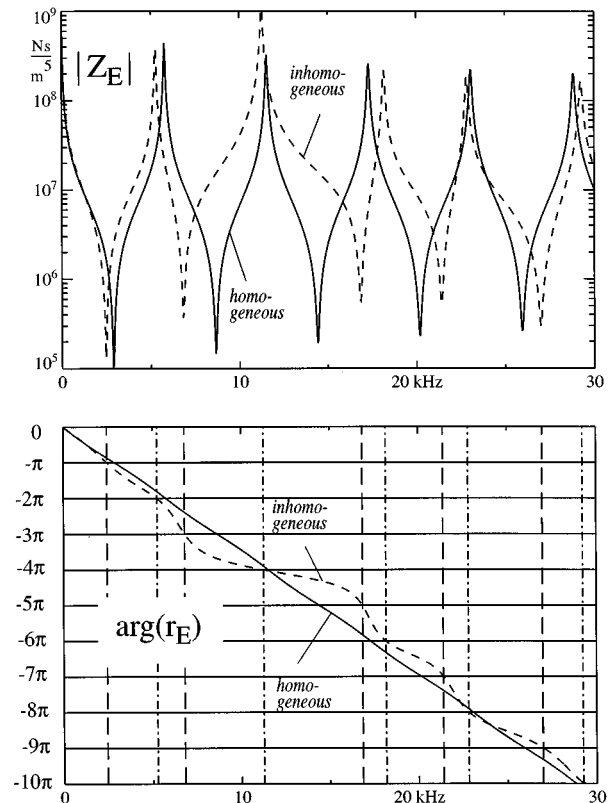


FIG. 12. Relationship between zeros and poles of the input impedance  $Z_E$  and the corresponding phase of the reflectance  $r_E$ . A reflectance phase of  $\arg(r_E) = -n\pi$  corresponds to a zero if  $n$  is odd and to a pole if  $n$  is even.

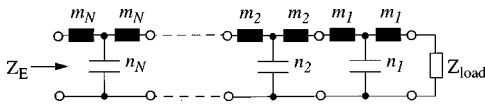


FIG. 13.  $T$  equivalent circuit of the ear canal terminated by a load impedance.

most part of the ear canal is well matched to the maximum slopes of the sine functions in  $x=0$ .

However, the frame length  $l_{\text{frame}}$  of the Fourier expansion must not equal the actual length of the ear canal  $l_{\text{ec}}$  as in this case the entrance area  $A_E$  would also equal  $A_0$ . The actual ear-canal length  $l_{\text{ec}}$  is unknown anyway, and has to be determined by means of the gradient procedure. The frame length  $l_{\text{frame}}$  poses an upper limit of the possible ear-canal lengths. It was chosen to be much larger than an ear canal, namely  $l_{\text{frame}}=50$  mm.

The ear-canal length  $l_{\text{ec}}$  has the greatest effect on the resulting chain parameters. Therefore, the parameters of the gradient method, the length  $l_{\text{ec}}$ , and the sine coefficients  $s_1, \dots, s_n$ , are not treated equally, but in the order of priority just given. This means that after every successful parameter change (producing a reduced phase deviation), further changes start again by first altering  $l_{\text{ec}}$ , then  $s_1$ , and then sine coefficients of a higher order. Therefore, the program is mostly busy adjusting the ear-canal length  $l_{\text{ec}}$ . A further step to help improve the convergence is to use random numbers, which reduce the size of the parameter change actually accepted after a successful trial. This technique is known from learning neural networks.

The calculation of a new input reflectance is the part of the gradient algorithm that is most frequently run. Therefore, it is important to reduce the computing time of this step as much as possible. A simple measure is a reduction of the frequency resolution used. A resolution of 100 Hz is completely sufficient. Furthermore, the calculation of the input impedance can be accelerated by using the  $T$  equivalent circuit of the ear canal (Fig. 13) instead of using chain matrices like the ones in Eq. (4). The acoustical compliances  $n_i$  and masses  $m_i$  of a section of length  $\Delta$  and radius  $r_i$  are given by

$$n_i = \frac{2\pi r_i \Delta}{p_0}, \quad m_i = \frac{\rho_0 \Delta}{4\pi r_i}. \quad (17)$$

Herein,  $p_0$  denotes the (normal) static pressure. Using the  $T$  equivalent circuits, the chain parameters need not be explicitly evaluated. It is much more efficient to calculate the input impedance  $Z_E$  iteratively.

Because all the eigenfrequencies are utilized, the reflectance-phase method is expected to provide a good estimation of the ear-canal shape. This is confirmed in Fig. 14, which represents a typical example. Simulating the "measured" input impedance, an individual eardrum impedance that had actually been measured was used again. The impact of the assumed load reflectance  $r_D$  on the estimated radius functions is insignificant. As no clear advantage of using the model eardrum impedance could be ascertained, the eardrum impedance can be disregarded ( $Z_D \rightarrow \infty$ ). The reflectance phase above 3 kHz is very well matched by the solutions found.

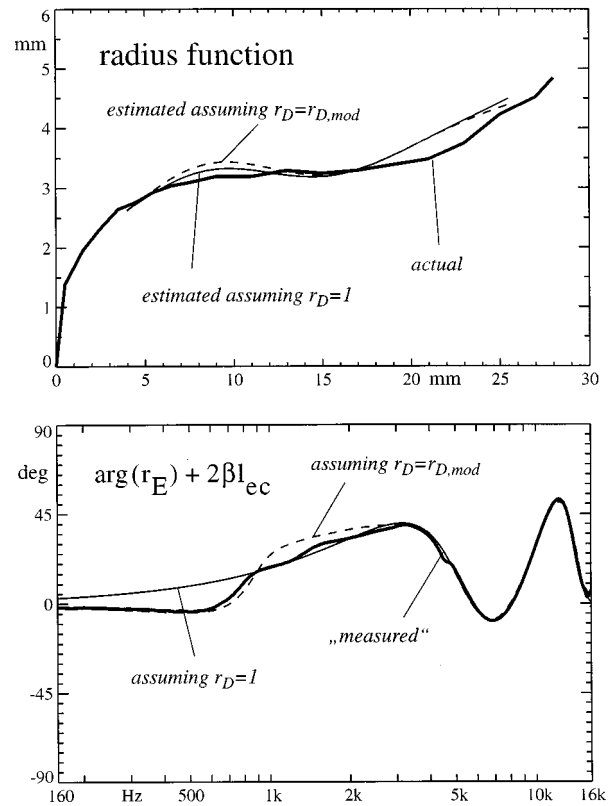


FIG. 14. Estimation of the radius function using the reflectance-phase method. The curves are calculated, but by using a measured individual eardrum impedance which is transformed via an individual ear canal. The upper panel shows the radius function of the individual ear canal and two estimates assuming a rigid eardrum (infinite impedance) and the eardrum impedance according to our model. In the lower panel the phase of the input reflectance  $r_E$  calculated from the measured eardrum impedance (measured) and the two approximations by the fitted ear canals are represented.

The resulting errors of the estimated pressure  $p_D$  and volume velocity  $q_D$  at the drum are represented in Fig. 15. The two quantities are calculated according to Eqs. (11) and (12) using the chain parameters derived from the area function estimated. The errors of the pressure  $p_D$  are very small. They are lower than 2 dB in magnitude and lower than  $2^\circ$  in phase up to more than 10 kHz. This is totally sufficient for most purposes. Also, the error of the volume velocity is considerably decreased compared to Fig. 6, where simply the average ear-canal chain matrix was used. However, as expected, the volume velocity estimate remains inaccurate at high frequencies.

### III. VALIDATION USING A PHYSICAL MODEL

Validating the reflectance phase method in real ears is hardly possible because the actual pressure  $p_D$  at the drum cannot be measured *in vivo*. As already mentioned, the pressure  $p_D$  is specified by the model<sup>2</sup> as an "effective" pressure which governs the mechanical vibrations in the middle ear. It is intended to represent a spatial average over the drum surface which cannot be measured under the circumstances given.

A reliable check of the pressure estimation can be obtained using a physical model of the ear canal and the middle ear. Such a model, described by Taschke *et al.*,<sup>30</sup> was devel-



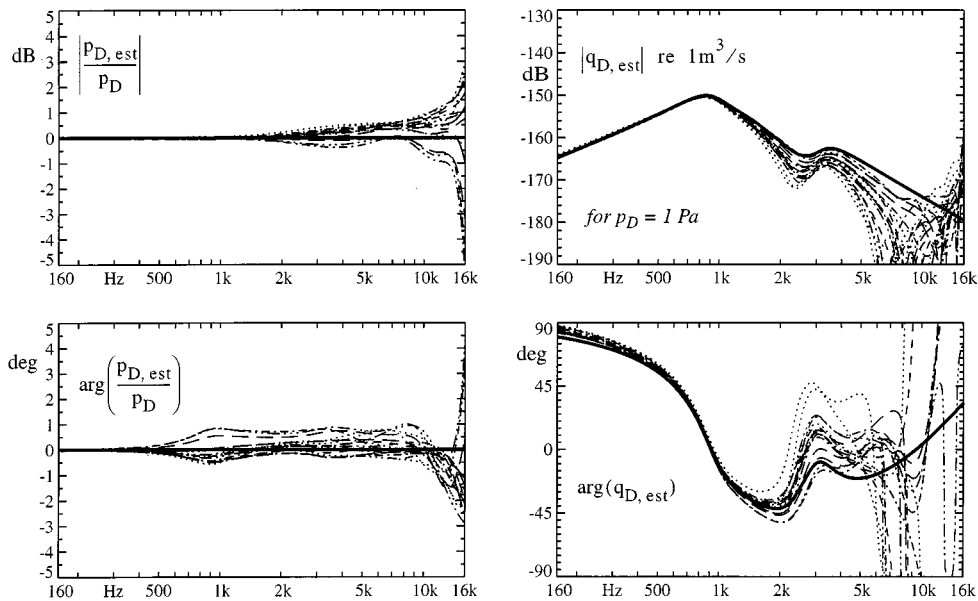


FIG. 15. Estimation of the pressure (left) and the volume velocity (right) at the drum as in Fig. 6, but using the estimated ear-canal chain matrix instead of the model matrix. The same data at the ear canal entrance ( $p_E$  and  $q_E$ ), as used in Fig. 6, are evaluated.

oped at our institute mainly for investigations of the middle-ear mechanics. The model allows a combination of different ear canals made of brass with an artificial middle ear which consists of an eardrum, the auditory ossicles, and the most relevant ligaments. For the application just mentioned, the stapes was disconnected from the incus so that the *processus lenticularis* could be easily accessed. The displacement  $\xi$  of this point was monitored using a laser vibrometer. It can be assumed to be invariably related to the effective eardrum pressure  $p_D$ . In this way, the measurement of  $p_D$  as a spatial average can be avoided. Instead of comparing a direct measurement of the eardrum pressure with an estimate, the corresponding transfer functions  $\xi/p_D$  and  $\xi/p_{D,est}$  are compared.

The acoustical measuring head (AMH) is coupled to the ear canal of the physical model. The internal loudspeaker is driven by white noise in the working frequency range up to 16 kHz. It generates a sound wave in the ear canal and simultaneously measures the acoustical input quantities  $p_E/u_{sp}$  and  $q_E/u_{sp}$  as transfer functions with respect to the loudspeaker voltage. The complex transfer functions are determined by a four-channel spectrum analyzer. Each transfer function contains 800 frequency lines from 20 Hz to 16 kHz with a spacing of 20 Hz. Averaging of the measured cross-correlation functions provides a sufficient signal-to-noise ratio in the full frequency range. Usually 1000 averages are taken.

The ratio of the two transfer functions gives the input impedance  $Z_E = (p_E/u_{sp}) / (q_E/u_{sp})$ . The displacement  $\xi$  of the *processus lenticularis* is simultaneously measured by means of the laser vibrometer, again referred to the loudspeaker voltage ( $\xi/u_{sp}$ ). The transfer function between the pressure at the entrance and the displacement of the *processus lenticularis* ( $\xi/p_E$ ) =  $(\xi/u_{sp}) / (p_E/u_{sp})$  is shown in Fig. 16. It contains two pressure minima as poles. From the input impedance  $Z_E$ , the ear-canal chain matrix and thus the pressure-transfer function  $H_{pED,est} = p_{D,est}/p_E$  is estimated using the reflectance-phase method. Thereby, the transfer function  $\xi/p_E$  can be converted into  $\xi/p_{D,est}$ .

The direct measurement of  $\xi/p_D$ , also represented in Fig. 16, was performed in a changed arrangement. The ear canal was removed, and the AMH was immediately coupled to the artificial eardrum. As the foremost microphone of the AMH has a sufficient distance from the eardrum, no disturbances caused by higher-order modes are possible. No inclination of the eardrum with respect to the ear-canal axis was realized. Therefore, the sound-field distribution over the eardrum surface is considerably altered compared to the reference case. Even so, the measured transfer function  $\xi/p_D$  agrees fairly well with the estimated one. The ratio of the transfer function  $\xi/p_D$  and  $\xi/p_{D,est}$  gives the resulting error as  $p_{D,est}/p_D$ . This error, depicted in the lower panel of Fig. 16, also contains the errors caused by the alteration to the sound field in front of the drum. Probably the larger errors in the frequency range 1–4 kHz are of this type. Therefore, the errors in Fig. 16 obtained by a simulation of the reflectance-phase method can be considered as being in good agreement with the measured errors.

Also, the area function estimated can be checked as the shape of the ear canal of the physical model is known. The accuracy of this estimation (Fig. 17) is roughly the same as found by means of the simulations (Fig. 14). In both cases, there is a considerable error in the ear-canal length estimate. These errors are partly compensated with respect to the chain parameters by a correspondingly altered area function. As observed in many other simulations, which cannot be presented in this paper, the total volume of the ear canal is usually fairly well estimated. To show the good convergence of the gradient method, ten estimates of the radius function are plotted in Fig. 17 in the overlay modus. The deviations stay within a band that corresponds to a moderately thick linestyle.

The measurements given in this section prove the applicability of the method, in principle. When the method was applied to *in vivo* measurements in real ears, some additional problems occurred. To improve the quality of the results, we used individually fitted ear molds to couple the (heated) AMH to ear canals. In this way, several measurements could

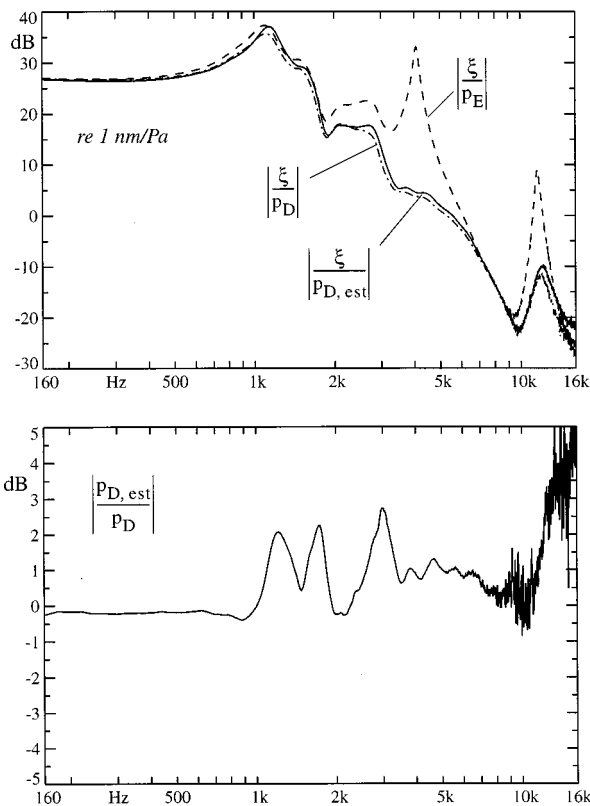


FIG. 16. The middle-ear transfer function  $\xi/p_E$  of a physical ear model (ratio of the displacement of the *processus lenticularis* to the pressure at the entrance of the ear canal) was measured using an acoustical measuring head (AMH) and a laser vibrometer. The transfer function  $\xi/p_{D,est}$  between the pressure at the drum and the displacement of the *processus lenticularis* was estimated from  $\xi/p_E$  and from the additionally measured input impedance  $Z_E$  by using the reflectance-phase method. The result of this estimation is compared to a direct measurement of the transfer function  $\xi/p_D$  described in the text. The lower panel represents the error of the pressure estimation.

be achieved that gave plausible results. But, in other cases, input reflectances were measured that failed to provide a reasonable estimation of the area function. Sometimes even negative resistances were found. We are convinced that the problems in determining area functions are caused by inaccurately measured reflectances. Therefore, we are going to further improve our measuring device. As long as the reli-

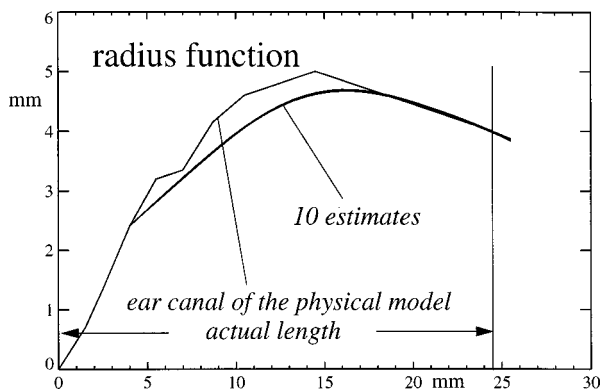


FIG. 17. Estimation of the ear-canal radius function of a physical model based on a measured input reflectance  $r_E$ . Ten estimates of the radius function plotted in the same diagram differ so little that seemingly a thicker linestyle results.

ability of the reflectance measurement is not considerably improved, no results will be presented.

#### IV. CONCLUSIONS

A new method, the reflectance-phase method, has been presented which allows audiological and psychoacoustical measurements to be referred to the sound pressure  $p_D$  actually generated at the eardrum. To apply this method, an acoustical measuring head (AMH), a spectrum analyzer, and some software (mainly the gradient method used for estimating the area function of the ear canal) are necessary. The AMH is used to generate a sound wave and to simultaneously measure the input impedance  $Z_E$  of the ear canal which is needed for the estimation. The pressure at the drum is estimated within an accuracy range of better than 3 dB (mostly much better) up to more than 10 kHz.

Furthermore, it was shown that a surprisingly good estimation of the pressure at the drum can be achieved, even without measuring the input impedance, by only using the minima of the pressure at the ear-canal entrance. This allows an estimation of similar accuracy (pressure-minima method), but not in the full working frequency range. In a band of about 1-kHz width around the pressure zeros, the accuracy is reduced. Here, the errors can reach a magnitude of 8 dB.

However, even this is a considerable improvement compared to the usage of an ear simulator like the IEC 711 coupler. As shown in Figs. 2 and 4, the errors of this simulator used in connection with the Beyer DT 48 earphone can reach 15 dB. These errors are not caused primarily by the coupler's poor design, but by the effect of individual ear canals. Therefore, a more accurately designed, but still invariable, ear simulator cannot provide a solution to the problem.

The best estimates can be obtained using the reflectance-phase method. It has been shown that this method provides very good results if the magnitude of the eardrum reflectance is not too small. All the reflectances we found in temporal bone measurements<sup>2-4</sup> exhibited this important feature. As the reflectance phase is to be evaluated only at frequencies above 3 kHz, the influence of impairments in the middle ear is insignificant—at least as far as can be predicted by our model. Others have measured lower eardrum reflectances at higher frequencies which would decrease the accuracy of the estimations. During unpublished *in vivo* measurements, we have sometimes also found lower reflectance magnitudes, but in these cases slight variations of the AMH calibration had a considerable impact on the result. So, we rather suspect artifacts in these cases. In this respect a clarification is absolutely essential.

This paper focuses on the investigation of methods for estimating the pressure at the drum and on the degree of accuracy achievable. However, for the application in hearing experiments the pressure appearing at the drum has to be not only estimated, but generated in a specified way. This step is easily done and does not require too much effort. Since the AMH is able to measure pressure and volume velocity, it can also be calibrated to generate a specified pressure. The only additional requirement to be met is an absolute calibration of the microphone(s) in the AMH.

The system consisting of the AMH and the coupled ear canal is linear to a good approximation. Even the acoustic reflex excited at higher sound levels has little effect on the ear-canal pressure-transfer function because the eardrum impedance at higher frequencies can hardly be increased by the acoustic reflex. At lower frequencies, the ear-canal transfer function is near unity and therefore not essentially influenced. In conclusion, the transfer function that converts the pressure estimated at the drum into the desired pressure must be measured only once for an individual subject. It may be updated from time to time to take into account time variance. The transfer function specifies a filter which has to be realized by digital signal processing. If only tones are to be used, a calibrated attenuator suffices. In a case where short test signals, like Gaussian pulses, are to be generated at the drum, the corresponding earphone voltage can be calculated by inverse Fourier transformation and realized by a programmable signal generator.

The method has been successfully applied to fitting hearing aids.<sup>17</sup> In this context also, the measurements of the hearing threshold of one subject is described which is referred to the pressure at the drum. The resulting curve is actually smoother than the ones usually measured.

There are several links to the work of Sanborn,<sup>19</sup> for instance, using a Thevenin equivalent circuit for describing the hearing aid. But in contrast to that paper, we refer all the fitting procedures to the eardrum pressure  $p_D$ . This has the advantage of producing objective functions which do not contain resonator effects of the ear canal. After making some changes to our measuring device, we are going to perform *in vivo* measurements of ear-canal input impedances and the hearing threshold, with reference to the pressure at drum and to the incident power flow as well. This should clarify the question as to which quantity is to be preferred as a reference for hearing experiments.

## ACKNOWLEDGMENTS

The work was supported by grants from the Bundesministerium für Bildung, Wissenschaft, Forschung, und Technologie (No. BMBF, 01 VJ 9302/6). We would like to thank Dipl.-Ing. Henning Taschke, who performed the measurements on the physical model of the ear canal and middle ear. Arnold Tubis and an unknown reviewer have given valuable comments and have helped to improve the quality of this paper with their constructive criticism. Also, a personal discussion with Stephen Neely and Hendrikus Duifhuis on the subject of whether the sound-pressure level or the sound-intensity level would be suited better as a reference quantity for hearing experiments was very helpful to us.

<sup>1</sup>S. T. Neely and M. P. Gorga, "Comparison between intensity and pressure as measures of sound level in the ear canal," *J. Acoust. Soc. Am.* **104**, 2925–2934 (1998).

<sup>2</sup>H. Hudde and A. Engel, "Measuring and modeling basic properties of the human middle ear and ear canal. Part I. Model structure and measuring techniques," *Acust./Acta Acust.* **84**, 720–738 (1998).

<sup>3</sup>H. Hudde and A. Engel, "Measuring and modeling basic properties of the human middle ear and ear canal. Part II: Ear canal, middle ear cavities,

eardrum, and ossicles," *Acust./Acta Acust.* **84**, 894–913 (1998).

<sup>4</sup>H. Hudde and A. Engel, "Measuring and modeling basic properties of the human middle ear and ear canal. Part III: Eardrum impedances, transfer functions, and complete model," *Acust./Acta Acust.* **84**, 1091–1108 (1998).

<sup>5</sup>H. Fastl, "Comparison of DT 48, TDH 49, and TDH 39 earphones," *J. Acoust. Soc. Am.* **66**, 702–703 (1979).

<sup>6</sup>H. Fastl and H. Fleischer, "Freifeldübertragungsmaße verschiedener elektrodynamischer und elektrostatischer Kopfhörer," *Acustica* **39**, 182–187 (1978).

<sup>7</sup>R. M. Cox, "NBS-9A coupler-to-eardrum transformation: TDH-39 and TDH-49 earphones," *J. Acoust. Soc. Am.* **79**, 120–123 (1986).

<sup>8</sup>H. C. Martin, G. F. Westwood, and J. M. Bamford, "Real-ear to coupler differences in children having otitis media with effusion," *Br. J. Audiol.* **30**, 71–78 (1996).

<sup>9</sup>H. C. Martin, K. J. Munro, and D. H. Langer, "Real-ear to coupler differences in children with grommets," *Br. J. Audiol.* **31**, 63–69 (1997).

<sup>10</sup>I. R. Swan and S. Gatehouse, "The value of routine in-the-ear measurement of hearing aid gain," *Br. J. Audiol.* **29**, 271–277 (1995).

<sup>11</sup>J. C. K. Chan and C. D. Geisler, "Estimation of eardrum acoustic pressure and of ear canal length from remote points in the canal," *J. Acoust. Soc. Am.* **87**, 1237–1247 (1990).

<sup>12</sup>M. R. Stinson and B. W. Lawton, "Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution," *J. Acoust. Soc. Am.* **85**, 2492–2503 (1989).

<sup>13</sup>K. N. Stevens, R. B. Berkovitz, G. D. Kidd, Jr., and M. Green, "Calibration of ear canals for audiometry at high frequencies," *J. Acoust. Soc. Am.* **81**, 470–484 (1987).

<sup>14</sup>H. Hudde, A. Engel, and A. Ludwig, "A wide-band precision acoustic measuring head," *Acust./Acta Acust.* **82**, 895–904 (1996).

<sup>15</sup>H. Hudde, "Kopfhörer-Impulsantworten im Gehörgang und im Mittelohr," in *Fortschritte der Akustik—DAGA'91 Bad Honnef: DPG-GmbH* (1991), pp. 609–612.

<sup>16</sup>S. Gilman and D. D. Dirks, "Acoustics of ear canal measurement of eardrum SPL in simulators," *J. Acoust. Soc. Am.* **80**, 783–793 (1986).

<sup>17</sup>A. Ludwig, "Ein System zur Anpassung von Hörgeräten: Meßtechnik, patientenbezogene Modellierung, Optimierung der Komponenten," Ph.D. thesis, Faculty of Electrical Engineering, Ruhr-Universität Bochum, Shaker Verlag, Aachen, 1998.

<sup>18</sup>D. H. Keefe, J. C. Bulen, K. H. Arehart, and E. M. Burns, "Ear-canal impedance and reflection coefficient in human infants and adults," *J. Acoust. Soc. Am.* **94**, 2617–2638 (1993).

<sup>19</sup>S. E. Voss and J. B. Allen, "Measurement of acoustic impedance and reflectance in the human ear canal," *J. Acoust. Soc. Am.* **95**, 372–384 (1994).

<sup>20</sup>P.-E. Sanborn, "Predicting hearing aid response in real ears," *J. Acoust. Soc. Am.* **103**, 3407–3417 (1998).

<sup>21</sup>J. J. Zwislöcki, "An acoustic coupler for earphone calibration," Special NASA report. LSC-S-7, Syracuse University, Syracuse, New York (1970).

<sup>22</sup>S. Mehrgardt and V. Mellert, "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.* **61**, 1567–1576 (1977).

<sup>23</sup>W. M. Rabinowitz, "Measurements of the acoustical input immittance of the human ear," *J. Acoust. Soc. Am.* **70**, 1025–1035 (1981).

<sup>24</sup>H. Hudde, "Measurement of the eardrum impedance of human ears," *J. Acoust. Soc. Am.* **73**, 242–247 (1983).

<sup>25</sup>K. Okabe, S. Tanaka, H. Hamada, T. Miura, and H. Funai, "Acoustical impedance measurement on normal ears of children," *J. Acoust. Soc. Jpn.* **9**, 287–294 (1988).

<sup>26</sup>M. Joswig, "Impulse response measurement of individual ear canals and impedances at the eardrum in man," *Acustica* **77**, 270–282 (1993).

<sup>27</sup>M. M. Sondhi and B. Gopinath, "Determination of vocal-tract shape from impulse response at the lips," *J. Acoust. Soc. Am.* **49**, 1867–1873 (1971).

<sup>28</sup>M. R. Schroeder, "Determination of the geometry of the human vocal tract by acoustic measurements," *J. Acoust. Soc. Am.* **41**, 1002–1010 (1967).

<sup>29</sup>D. Hammershoi and H. Moeller, "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Am.* **100**, 408–427 (1996).

<sup>30</sup>H. Taschke, Ch. Weistenhöfer, and H. Hudde, "Ein akustomechanisches Mittelohrmodell in Originalgröße," in *Fortschritte der Akustik-DAGA 98, Dtsch. Ges. Akustik, Oldenburg* (1998), pp. 296–297.

# Geometry, kinematics, and acoustics of Tamil liquid consonants

Shrikanth Narayanan

*AT&T Labs—Research, 180 Park Avenue, Florham Park, New Jersey 07932-0971*

Dani Byrd

*Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511*

Abigail Kaun

*Department of Linguistics, Yale University, New Haven, Connecticut 06520*

(Received 10 December 1997; revised 15 February 1999; accepted 15 May 1999)

Tamil is unusual among the world's languages in that some of its dialects have five contrasting liquids. This paper focuses on the characterization of these sounds in terms of articulatory geometry and kinematics, as well as their articulatory-acoustic relations. This study illustrates the use of multiple techniques—static palatography, magnetic resonance imaging (MRI), and magnetometry (EMMA)—for investigating both static and dynamic articulatory characteristics using a single native speaker of Tamil. Dialectal merger and neutralization phenomena exhibited by the liquids of Tamil are discussed. Comparisons of English /ɹ/ and /l/ with Tamil provide evidence for generality in underlying mechanisms of rhotic and lateral production. The articulatory data justify the postulation of a class of rhotics and a class of laterals in Tamil, but do not provide evidence in favor of a larger class of liquids. Such a superclass appears to have largely an acoustic basis. © 1999 Acoustical Society of America. [S0001-4966(99)03809-6]

PACS numbers: 43.70.Aj [WS]

## INTRODUCTION

Tamil, a Dravidian language of Southern India, is unusual among the world's languages in that many of its dialects have five contrastive liquid consonants (Christdas, 1988, and references cited therein). An accurate articulatory and acoustic characterization of these sounds has been elusive. In this paper, our goal is twofold. First, this work provides a novel use of multiple technologies in speech production research in order to provide a detailed description of the articulatory geometry, kinematics, and acoustic characteristics of these liquids. We have pursued the combined use of multiple methodologies with a single speaker—magnetic resonance imaging (MRI), magnetometry in the form of the electromagnetic midsagittal articulometer (EMMA), static palatography, and acoustic analysis and modeling. This effort continues the exciting recent advances in the use of imaging in speech production research. When combined with more established tools in speech production studies, imaging technology can help identify which aspects of vocal tract behavior are of linguistic significance.

The second goal of this paper is a linguistic one. Liquids have long been recognized as difficult to study, in part because of the diversity of forms these articulations take. Tamil provides a particularly complex set of liquids, including retroflex and nonretroflex articulations. Phonetic study of Tamil thereby provides a unique opportunity to add to the linguistic body of knowledge regarding liquid sounds. This paper contributes to an explicit description of the five liquids of Tamil in the articulatory and acoustic domains, and attempts to identify the linguistically significant aspects of each sound.

The vocal tract geometry of these liquid articulations has been briefly outlined by Narayanan *et al.* (1996). The present study presents a more in-depth description of the articulatory

geometry by providing information on the kinematics and articulatory-acoustic relations of these five consonants. It is hoped that by combining, within the same speaker, three-dimensional information available from static MRI with midsagittal kinematic information provided by magnetometry, a more comprehensive understanding can be obtained of the articulatory mechanisms involved in producing this complex system of linguistic contrast.

Advanced technologies such as MRI and magnetometry now enable detailed investigations of complex speech sounds such as liquid consonants. Until now, such sounds have resisted fully satisfactory articulatory descriptions. Each of these techniques has its advantages and disadvantages. MRI scans require artificial prolongation of the sound but provide information on the shape of the vocal tract not obtainable by other methods. Static palatography measures only the aggregate articulatory contact throughout an utterance. Nonetheless, this technique does show fairly precisely which part of the tongue makes contact with which part of the palate. Magnetometry recordings provide valuable dynamic information but are restricted to tracking just a few points along the midsagittal plane. In spite of their respective shortcomings, when combined, the articulatory data available from all three techniques enable us to obtain an increased understanding of vocal tract and tongue shaping mechanisms.

The five liquids of the relevant Tamil dialects are voiced. Two are described as rhotics and two as laterals (e.g., Christdas, 1988). The fifth has been variously described, sometimes as a rhotacized lateral (Balasubramanian, 1972; Christdas, 1988, and references cited therein). This liquid is placed among the rhotic group in the discussion below. The decision was based on the fact that like other rhotic sounds, [ɹ] involves central airflow. Laterals, by contrast, exhibit airflow along the sides of the tongue. Throughout this paper,



TABLE I. Symbols used to identify the five Tamil liquids.

IPA symbol	Description	Romanized orthography
[l]	dental l	l
[ɭ]	retroflex l	L
[r]	pre-alveolar r	r
[ɻ]	post-alveolar r	R
[ɻ̟]	palatal r	zh

the IPA symbols for these liquids proposed by Narayanan *et al.* (1996), based largely on MRI analyses of their geometric tongue shapes, have been adopted. The five liquids are identified then as in Table I.

In the report below, the results of three experiments designed to investigate the articulation of these Tamil liquids are presented. The first two experiments use MRI and static palatography data collected on two different days in 1995 for a single speaker—MRI at Cedars-Sinai Medical Center (Los Angeles, CA) and palatography at the UCLA Phonetics Laboratory. The third experiment reports data collected for the same speaker in 1996 using the EMMA magnetometry system (Perkell *et al.*, 1992) at Haskins Laboratories (New Haven, CT). Thus all the results reported here are based on a single experimental subject, SN (the first author). The static articulatory geometry of the Tamil liquids is described in Sec. I on the basis of the structural MRI and palatography. Section II presents a kinematic analysis of the liquids using magnetometer data.

The information obtained in the experimental work helps illuminate various issues in speech production and linguistic phonetics. In Sec. III, the articulatory-acoustic relations involved in producing these liquids are modeled based on vocal tract dimensions evidenced in the MRI data. The results are compared to the natural speech spectra. It is found that, despite the complexity of these articulations, by using a simple one-dimensional acoustic model and vocal tract data derived from MRI, we are able to establish the basic relations between vocal tract cavities and formant structure. Then, with a comprehensive picture of the articulatory and acoustic characteristics of these sounds in hand, three issues of significance in linguistic phonetics are discussed in Sec. IV. It is postulated that three-dimensional tongue shaping serves to unify articulations within the rhotic group and within the lateral group. Acoustic characteristics, on the other hand, show overlap between these two classes, suggesting that the broader category “liquid” has an acoustic basis. Next, a discussion is provided on how information regarding articulatory geometry and kinematics, in addition to a consideration of acoustic characteristics, can be brought to bear in understanding phonological merger and substitution that take place among these sounds. Finally, comparison of the Tamil liquids with the phonetic characteristics of the two liquids of American English is provided.

## I. STATIC ARTICULATORY GEOMETRY—MRI AND PALATOGRAPHY FINDINGS

### A. Method

#### 1. Subject

MRI and static palatography data were acquired for one native male speaker of the Brahmin dialect (SN, the first

author). This subject, who speaks English as a second language, was raised in Madras, India, where he spent the first 20 years of his life. In 1988, he moved to the United States, where he resided at the time of these experiments.

### 2. Structural MRI

Information about “static” vocal tract shapes came from MRI scans (GE 1.5 T scanner) at contiguous 3-mm intervals in the sagittal and coronal anatomical planes. This allowed the construction of three dimensional views of the vocal tract in a computer representation. The subject, in a supine position in the scanner, produced each consonant preceded by /pa/ (i.e., /paC/) and continued sustaining the final consonant for about 13 s, thereby enabling four contiguous image slices to be recorded (3.2 s/slice). The above procedure was repeated until the entire vocal tract region was imaged. (Spectrographic analysis, although not discussed below, indicated that formant values were near steady-state for the sustained utterances.) Details of image acquisition, measurement, and analysis are similar to those given in Narayanan *et al.* (1995). Measurements of vocal tract dimensions and cavity volumes were obtained both from raw image scans and computer reconstructions of the 3-D vocal tract. Area functions reported in this paper (Sec. III B and Appendix B) were obtained by re-sampling the 3-D vocal tract at 0.43-cm contiguous intervals along, and in a plane perpendicular to, the vocal tract midline specified in a midsagittal reference image. The cross-sectional areas were calculated directly by pixel counting.

### 3. Static palatography

Static palatography was used to register graphically the contact of the tongue with the palate, alveolar ridge, and inner margins of the teeth (Ladefoged, 1957). Carbon powder was coated on the tongue surface prior to speaking, and after articulation the resulting contact patterns on both the tongue and palate were captured with video imaging. This method records any and all palatal areas at which lingual contact occurred. A subset of the words (without a carrier phrase) used for the magnetometry recording (Appendix A) was used for palatography. The resulting (video) palatograms and linguograms provided data that were useful in inferring tongue shapes.

## B. Static articulatory geometry—Results

Midsagittal MRI scans for the five (artificially sustained) Tamil liquids are shown in Figs. 1–3. Example tongue-palate contact patterns for the Tamil liquids [l], [ɭ], and [ɻ̟] are given in Figs. 1(e), (f), and 3(b), respectively.

### 1. Laterals

For the laterals, Fig. 1 shows that [l] was characterized by tongue-tip contact in the dental region, a somewhat high-posterior tongue body position, and retraction of the tongue root toward the posterior pharyngeal wall [Fig. 1(a)]. There was a flat anterior surface, but the tongue was generally convex. The curved sides of the posterior tongue create the inward lateral compression characteristic of the consonant (cf.

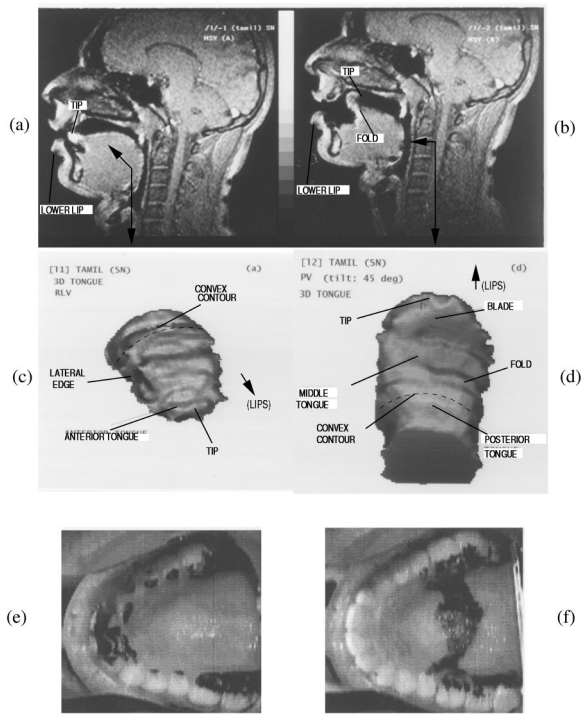


FIG. 1. (a) Midsagittal MR image for [l] (b) Midsagittal MR image for [ɭ]. (c) 3-D tongue shape for [l]. The 3-D tongue is viewed from the side [refer to the view orientation arrow shown in the midsagittal image of panel (a)] and tilted toward the front for better display. The tongue tip appears toward the lower right of the panel. (d) 3-D tongue shape for [ɭ]. The 3-D tongue is viewed from the posterior pharyngeal wall [refer to the view orientation arrow shown in the midsagittal image of panel (b)] and given a forward tilt of 35° for a better view. The tongue tip is toward the top of the panel. (e) Linguopalatal contact for [l], dark region is area of contact, front of mouth is toward the left. (f) Linguopalatal contact for [ɭ] (dark region is area of contact, front of mouth is toward the left).

Stone *et al.*, 1992; Narayanan *et al.*, 1997). Studying the tongue-palate contact by means of palatography in conjunction with the midsagittal MR images provided further important information. [l] was apical, characterized by medial tongue-tip closure at and behind the central incisors and lat-

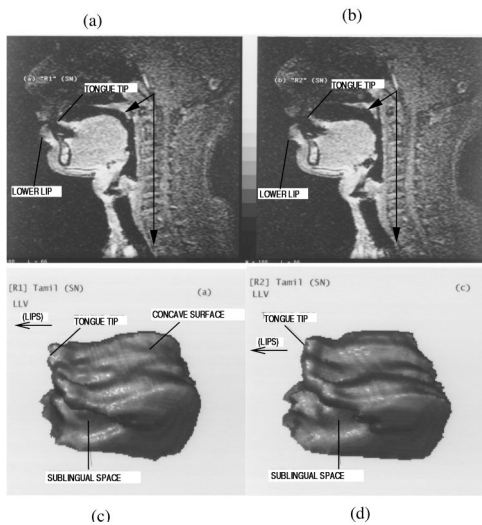


FIG. 2. (a) Midsagittal MR image for [r]. (b) Midsagittal MR image for [ɽ]. (c) 3-D tongue shape for [r] (tongue tip toward the left of panel). (d) 3-D tongue shape for [ɽ] (tongue tip toward the left of panel).

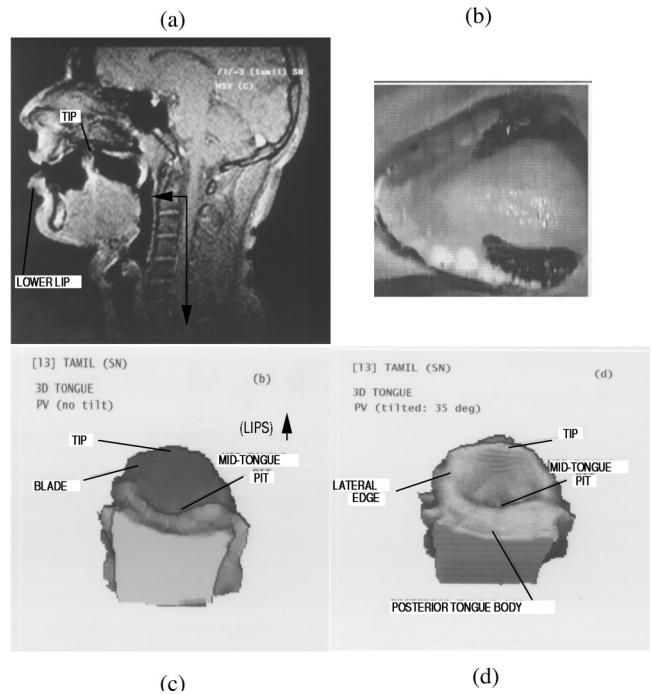


FIG. 3. (a) Midsagittal MR image for [ɻ]. (b) Linguopalatal contact for [ɻ] (dark region is area of contact along the sides of the middle tongue, front of mouth is toward the left). (c) 3-D tongue shape for [ɻ]. The 3-D tongue is viewed from behind, looking in from the posterior pharyngeal wall toward the direction of the lips [refer to the view orientation arrow shown in panel (a)]. The tongue tip is toward the upper middle of panel. (d) Same as in (c), but with a 45° front-to-back tilt of the tongue body to enable a better view of the mid-tongue pit.

eral contact in the postpalatal/velar region (starting near the first molar). Although these patterns suggest lateral airflow paths in the anterior region, prediction of actual cross-sectional tongue shapes and area functions is not straightforward; midsagittal and tongue contact information alone is insufficient.

The anterior tongue body for [ɻ] [Fig. 1(b)] was drawn upward and well inside the oral cavity, with the medial tongue occlusion appearing in the palatal region. The anterior tongue body was flat and raised upward while the posterior tongue body was convex. [ɻ] was subapical with contact made along the edge of the underside of the tongue in the palatal region. It should be noted that [ɻ] in syllable-initial cases is realized as a flap, and often may not involve complete (subapical) palatal contact in fluent speech. In cases where there was complete linguopalatal closure for [ɻ], such as in syllable final position, the anterior contact pattern was more extensive (laterally) when compared to [l].

## 2. Rhotics

For the rhotics, the overall midsagittal tongues shapes for [r] and [ɽ] were very similar (Fig. 2). The tongue-tip constriction for [r] was in the prealveolar region. The tongue surface contour was slightly concave or flat and had lateral bracing in the palatal region. Lingua-palatal contact showed no medial closure. The post alveolar [ɽ] had a midsagittal tongue shape similar to [r], but the narrowest tongue-tip constriction was more posterior. The posterior tongue body for [ɽ] was somewhat lower than for [r]. The tongue-palate con-

tact patterns obtained by palatography for [r] and [ɾ], which are both apical, were not very distinct. In both cases, lateral linguapalatal bracing in the palatal region played a role in tongue shaping.

Lastly, for [ɹ], the anterior tongue body [Fig. 3(a)] was drawn upward and well inside the oral cavity. The narrowest tongue constriction appeared in the palatal region, although the exact location was inconsistent, as was demonstrated by the magnetometer data (see Sec. II). There was no medial linguapalatal contact and central airflow occurred. In addition, there was lateral contact with the tongue body middle in the palatal region (extending for about 1 cm). Crucially, the upward raised and inward pulled anterior tongue position produced a pitlike cavity in the middle of the tongue body that was supported by bracing of the sides of the mid-tongue region against the palate [see Fig. 3(d)]. The posterior tongue had no bracing and was somewhat flat. This liquid was strikingly distinguished from the others by the pitted tongue shape and a correspondingly greater back-cavity volume, the significance of which will be discussed in Sec. III B.

## II. ARTICULATORY KINEMATICS—QUANTITATIVE MAGNETOMETRY FINDINGS

### A. EMMA method

#### 1. Stimuli

This investigation of the production characteristics of the Tamil liquids focused on the kinematic behavior of the tongue tip. The tongue tip is the articulator that creates the narrowest constriction for all of the liquids. The stimuli included each of the five liquids in the following contexts: /kaCi, paCi, vaC, aCai, paCam/ where C was {[I], [l], [r], [ɾ], [ɹ]}. Because three of these stimuli were nonsense words, three additional meaningful words with parallel segmental structure were added to supplement the list. This yielded a total corpus of 28 words. These words were randomized and presented in the carrier phrase “*Andha vakyam \_\_ perusu*” (*The utterance \_\_ is big.*) Ten repetitions of this randomized list were recorded. These words are shown in Appendix A as they were presented to the subject. In all, 280 sentences were recorded.

#### 2. Subject and data collection

The EMMA magnetometer system (Perkell *et al.*, 1992) was used to track the movement of the articulators. These data were acquired from the same speaker (see Sec. IA 1) who recorded for the MRI and static palatography experiments, i.e., a native male speaker of the Brahmin dialect (SN). This experiment was part of a longer magnetometer recording session, and was the second of three recordings in that session.

The EMMA magnetometer system was used to transduce the horizontal ( $x$ ) and vertical ( $y$ ) movements of small coils attached to the articulators in the midsagittal plane. The technical specifications of the EMMA magnetometer system are outlined in Perkell *et al.* (1992) (see also Gracco and Nye, 1993; Löfqvist, 1993). Single transducers were placed on the nose, upper and lower gumline (maxilla and jaw, respectively), upper and lower lips at the vermilion border,

tongue tip, and three transducers were placed on the tongue body. The EMMA data were sampled at 625 Hz after low-pass filtering at 200 Hz before voltage-to-distance conversion. After voltage to distance conversion (with a filter cutoff of 17 Hz), correction for head movement (using the nose and maxillary reference transducers), and rotation to the occlusal plane, the position signals were subject to 25-point smoothing by a triangular filter.

### 3. Data analysis

The measurements made in this experiment were of the spatiotemporal behavior of the transducer placed on the tongue tip (approximately 7 mm, with the tongue somewhat extended, from the tongue-tip apex on the superior surface).

*a. Signal analysis.* Horizontal ( $x$ ) and vertical ( $y$ ) position signals for the tongue-tip transducer were used to calculate the tangential velocity of the tongue tip according to the following formula:

$$tvel = \sqrt{((\dot{x})^2 + (\dot{y})^2)},$$

where  $tvel$  = tangential velocity of the tongue-tip transducer,  $\dot{x}$  = velocity in the  $x$ -coordinate of the tongue-tip transducer,  $\dot{y}$  = velocity in the  $y$ -coordinate of the tongue-tip transducer. This signal was also smoothed at 25 points.

The beginning, extremum, and end of the production of the liquid consonant were defined by algorithmically identifying minima in the tangential velocity signal using the HADES signal analysis program (Rubin, 1995). The time and tongue-tip  $x$ - and  $y$ -positions at each of these points were recorded. [See Löfqvist *et al.* (1993) for a description of the use of tangential velocity in the segmentation of magnetometer signals.] Additionally, for each opening and closing movement, time and magnitude of the peak tangential velocities were collected. Finally for each constriction formation (interval between beginning and extremum) and constriction release (interval between extremum and end), the pathlength and average curvature were calculated. Pathlength is the actual (midsagittal) distance traveled by the transducer, and is calculated by summing the consecutive Euclidean distances between each sample. Curvature is indicative of the direction of movement with negative values for clockwise movement and positive values for counterclockwise movement, assuming that the subject is facing to the left. [See Löfqvist *et al.* (1993) for a more extensive discussion of the use of curvature information in the analysis of EMMA data.] This calculation was made according to the following formula:

$$\text{curvature} = \frac{\sum((\dot{x}\ddot{y} - \dot{y}\ddot{x})/tvel^3)}{\text{constriction interval}}$$

where  $curvature$  is the average curvature value over constriction formation (or release),  $\dot{x}$  (or  $\dot{y}$ ) is the  $x$  (or  $y$ ) component velocity,  $\ddot{x}$  (or  $\ddot{y}$ ) is the  $x$  (or  $y$ ) component acceleration,  $tvel$  is the tangential velocity, and  $constriction\ interval$  is the duration of the constriction formation (or release). Figure 4 shows an example of measurements made for a token of [palam]. A movement token was excluded if multiple velocity minima occurred at the movement's extremum or if multiple velocity maxima occurred during a single constriction



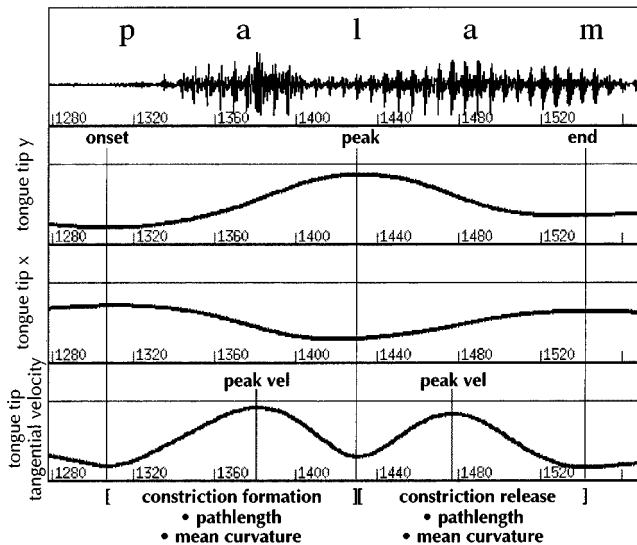


FIG. 4. A sample token of EMMA data indicating the measurements made for experiment 3.

formation or release. Such exclusions were rare, amounting to 21 tokens (out of 280 stimuli). Of these, 17 were word-final liquids, including all 10 tokens of the word final [l].<sup>1</sup> Lastly, the automated algorithm for calculating curvature failed for 13 tokens, 2–4 tokens of each liquid type.

*b. Quantitative analysis.* The complete list of kinematic variables considered in the statistical analysis is shown below. (Recall that constriction beginning, extremum, and end points are defined by the tangential velocity minima; that constriction formation is defined as the interval between beginning and extremum; and that constriction release is defined as the interval between extremum and end.)

- the tongue-tip y-position at extremum constriction;
- the tongue-tip x-position at extremum constriction;
- peak tangential velocity for constriction formation and release;
- pathlength for constriction formation and release;
- average curvature for constriction formation and release;
- duration of constriction formation and release;
- time from onset to peak constriction velocity;
- proportional time from beginning to peak constriction velocity (i.e., time to peak velocity/duration of constriction formation).

These measurements reflect the duration and magnitude of the formation and release of the liquid articulations and help illuminate certain aspects of the intragesural dynamics.

Three statistical analyses were conducted using a  $p \leq 0.05$  criterion for significance. The first analysis was a two-factor ANOVA with the factors LIQUID (dental [l], retroflex [ɭ], prealveolar [r], postalveolar [ɻ], palatal [ɹ]) and VOCALIC CONTEXT (a\_a, a\_i, a\_ai, a\_#). For this analysis, *post hoc* Scheffé's S-tests (Scheffé, 1953) were used to test *post hoc* comparisons among the liquids. (Because of the many possible crossings of the two factors, discussion of vocalic context effects are left for the second analysis.) This analysis examines all the dependent variables (a)–(h) listed

above. The second analysis was a two-factor ANOVA testing for effects of LIQUID and VOWEL HEIGHT (lowV\_lowV and lowV\_hiV). Main effects of contextual vowel height are reported, as well as any crossover interactions between LIQUID and VOWEL HEIGHT. The final analysis was a two-factor ANOVA testing for effects of LIQUID and WORD POSITION (word final and intervocalic). (Note that there are generally many fewer word final tokens than intervocalic tokens, and no data for the retroflex [ɭ] could be included.) Main effect of WORD POSITION and any crossover interactions with LIQUID are reported. The final two analyses examined only a subset of the dependent variables.

## B. Articulatory kinematics—Results

### 1. Analysis one—Main effects of liquid identity

There was a main effect of liquid on both peak x-position [ $F(4,241) = 460.774, p = 0.0001$ ] and peak y-position [ $F(4,241) = 1937.996, p = 0.0001$ ]. *Post hoc* tests showed all the liquids to differ significantly in peak tongue-tip x-position, and all except the dental [l] and post alveolar [ɻ] to differ significantly in peak tongue-tip y-position.<sup>2</sup> The five liquids were highly differentiable in x-y space by their tongue-tip positions at peak constriction. The dental [l], pre-alveolar [r], and postalveolar [ɻ] moved progressively back in the x-dimension, with little difference in the y-dimension. The palatal [ɹ] and retroflex [ɭ] were considerably back and somewhat higher than the other liquids. The palatal [ɹ] showed the most variability in y. This variability in tongue-tip position reflected a greater propensity to coarticulation with neighboring vowels. The variability in [ɭ] was less than in [ɹ] ([ɭ] showed some fronting in the x-dimension in the /a\_i/ context). The relative x,y-positions for the liquids are shown in Fig. 5. Recall that examination of static MRI and palatography data outside the midsagittal plane demonstrated that the palatal [ɹ] and the retroflex [ɭ] also differed in that the former had tongue side bracing that the latter lacks. While the tongue sides were braced, the tongue tip was free, giving rise to an expectation of positional variability on the part of the tongue tip. This expectation was borne out by the magnetometer findings—observe the greater scattering of the y tongue-tip position for [ɹ] in Fig. 5.

Apart from the positional characteristics of liquid production, dynamic characteristics such as the constriction formation and release velocities also provided useful information. The constriction formation peak velocities were significantly affected by liquid identity [ $F(4,241) = 127.1307, p = 0.0001$ ] with all liquids differing from one another except the dental [l] and the palatal [ɹ]. The retroflex [ɭ] and the postalveolar [ɻ] had the highest constriction formation peak velocities, and these are the sounds that are typically described as flap and tap articulations, respectively. The constriction release peak velocities were also significantly affected [ $F(4,241) = 1623.2642, p = 0.0001$ ] and differed for all liquids except the dental [l] and the postalveolar [ɻ]. These velocities were much higher for the two backmost articulations—the palatal [ɹ] and the retroflex [ɭ]—with the backer of the two, i.e., the retroflex [ɭ], having the highest peak release velocities. The backmost [ɭ] had about three



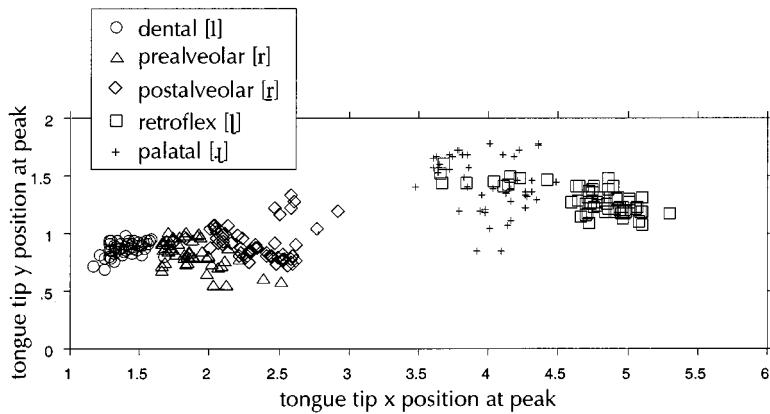


FIG. 5. Tongue-tip x- and y-positions with respect to the occlusal plane at y-peak for the five liquids.

times the release velocity of the fronter three articulations, [l], [r], and [ɾ], which differed only minimally among themselves. The slightly fronter [ɟ] had about two times the release velocity of [l], [r], and [ɾ]. Thus more posterior tongue-tip constriction position correlated with increased release velocity.

Constriction formation duration and pathlength exhibited similar patterns [ $F(4,241)=46.1668, p=0.0001$ ;  $F(4,241)=176.8743, p=0.0001$ ]. *Post hoc* tests showed that constriction formation duration is shortest for fronter articulations—the dental [l], prealveolar [r], postalveolar [ɾ]—and longest for the backer ones—palatal [ɟ] and the retroflex [ɻ]. Constriction formation pathlength was shortest for the prealveolar [r], next shortest for dental [l], intermediate for the postalveolar [ɾ], longer for the palatal [ɟ], and longest for the retroflex [ɻ]. Here, greater constriction formation pathlengths and durations tend to correlate with posterior tongue-tip position.

Release duration and pathlength displayed similar patterns [ $F(4,241)=115.0354, p=0.0001$ ;  $F(4,241)=1444.7914, p=0.0001$ ]. Just like constriction formation duration, release duration was shortest for fronter articulations—the dental [l], prealveolar [r], postalveolar [ɾ]—and longest for articulations further back—palatal [ɟ] and the retroflex [ɻ]. Release pathlength was shortest for the dental [l] and prealveolar [r], slightly longer for the postalveolar [ɾ], longer for the palatal [ɟ], and longest for the retroflex [ɻ]. Overall, the constriction and release duration and pathlength data follow the general pattern of the-farther-the-longer behavior that has been found to be typical of both speech and limb movements, whereby durations tended to be longer for movements with larger displacements (e.g., Kelso *et al.*, 1985; Ostry and Munhall, 1985; Saltzman *et al.*, in press).

Recall that constriction formation curvature refers to the direction of tongue-tip movement. This measure was significantly affected by liquid identity [ $F(4,228)=59.2735, p=0.0001$ ] (see Fig. 6). It was positive with low variability for the retroflex [ɻ] and palatal [ɟ]. This means that the two high back articulations [ɻ] and [ɟ] had counterclockwise movements (head facing to the left). In other words, the x-component peak was attained before the y-component peak. Curvature was negative for [r] and [ɾ], with [ɾ] exhibiting low variability and [r] exhibiting moderate variability. For the dental [l], this measure tended to be negative but

quite variable, with many tokens having a near zero value. *Post hoc* tests indicated that the positive-curvature articulations differ significantly from the negative-curvature articulations.

Liquid identity also affected release curvatures [ $F(4,228)=26.5673, p=0.0001$ ]. Positive values for the retroflex [ɻ] and negative values for postalveolar [ɾ] were observed. It was fairly consistent that release curvature was slightly positive with low variability for [ɟ] and slightly negative for prealveolar [r]. It was quite variable for the dental [l], including many positive and negative values. The patterns for curvature can be summarized as follows. The dental [l] was formed with a straight (uncurved) or slightly clockwise movement and may be released in either a clockwise or counterclockwise direction. The retroflex [ɻ] and palatal [ɟ] were formed and released in the counterclockwise direction, i.e., back-to-front. The [r] and [ɾ] were formed and released with a basically clockwise direction, front-to-back. Note that

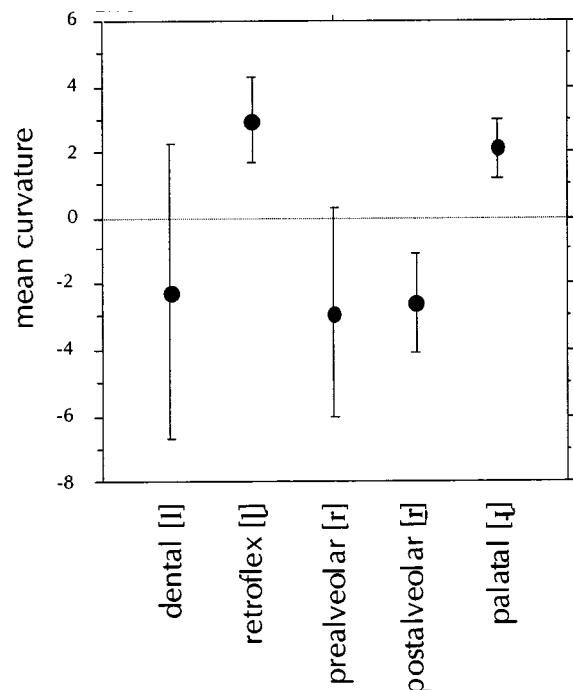


FIG. 6. Mean constriction formation curvature for the five liquids with one standard deviation error bars.

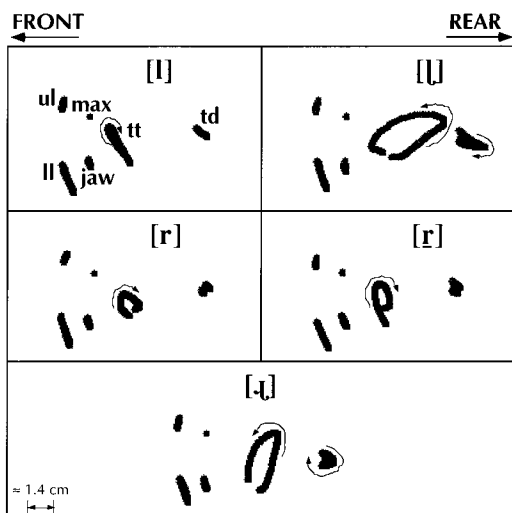


FIG. 7. Sample position trajectories for the tongue tip, with the upper lip, lower lip, maxilla, jaw, and tongue dorsum shown for reference, for tokens of each of the five liquids in the [a\_a] context. Arrows indicate the direction of movement of the tongue tip, and where necessary the tongue dorsum. The trajectories are displayed for the interval between the points of minimum tongue-tip height preceding and following the constriction.

while the tongue tip for palatal [ɹ] was quite variable in the  $x,y$ -position, it was remarkably consistent in curvature. Figure 7 provides a qualitative visual aid for observing the overall location and direction of tongue tip movement for each liquid.

Finally, consider the time from the onset of constriction formation to peak velocity, as well as this interval normalized for the duration of the constriction formation interval. The latter is a measure of the skewness of the velocity profile. These variables provide insight into the temporal dynamics underlying the constriction formation. These variables will be referred to as time to peak velocity and proportional time to peak velocity. These variables are of specific interest in considering the contextual effects covered in the following two sections, but some differences do exist among the liquids independent of context [ $F(4,241) = 14.7912, p = 0.0001$  for time to peak velocity and  $F(4,241) = 52.0319, p = 0.0001$  for proportional time to peak velocity]. *Post hoc* tests showed the retroflex [ʎ] to have a longer time to peak velocity than the other liquids. The post-alveolar [ɹ] was also found to have a longer time to peak velocity than dental [l]. For proportional time to peak velocity, *post hoc* tests found almost all pairwise comparisons to differ significantly. Peak velocity occurs proportionally earlier for the palatal [ɹ], intermediate for the [l] and [ʎ], and latest for the [r] and [ɹ].

## 2. Analysis two—Main effects of contextual vowel height

In this analysis the stimuli were assigned to two vowel height groupings. The Lo\_Lo group included the *aCai* and *paCam* words and the Lo\_Hi group included the *kaCi* and *paCi* words. Note that these groupings are a function of following vowel height. (Interesting effects on the liquid articulations due to preceding vowel height are likely to exist; unfortunately, the stimuli did not allow us to investigate

these effects because in all cases the liquid in question was preceded by the vowel [a].) Two-factor ANOVAs with LIQUID and VOWEL HEIGHT as factors were conducted. Below are described the main effects of vowel height on (1) peak  $x$ - and  $y$ -positions, (2) constriction formation and release durations, and (3) constriction formation and release pathlengths.

Small but significant effects of following vowel height existed for peak positions [ $F(1,217) = 49.8508, p = 0.0001$  for  $x$ -position, and  $F(1,217) = 121.1377, p = 0.0001$  for  $y$ -position] such that the Lo\_Hi context had fronter and higher tongue-tip positions at peak for all liquids. These differences were small for [l], and relatively large in the  $y$ -dimension for [ɹ]. No significant effect was found for pathlength or release duration. The constriction formation duration is shorter [ $F(1,217) = 9.0261, p = 0.003$ ] for the Lo\_Hi context, especially for the nonfront liquids.

## 3. Analysis three—Main effects of word position

Byrd and Saltzman (1998) have demonstrated that increased absolute (and for some speakers, proportional) time-to-peak velocity are informative dynamic “signatures” of phrase-final lengthening. (They relate this to a lowering of the stiffness parameter in a critically damped mass-spring gestural model.) This present analysis investigated whether these articulatory signatures of lengthening existed for word-final liquids. In this analysis the stimuli were assigned to two groupings: intervocalic and word-final. Note that for the palatal [ɹ] this included only six tokens, for the postalveolar [ɹ] this included only seven tokens, and no tokens of retroflex [ʎ] are included. For this reason the results below, while robust, must be considered preliminary. Two-factor ANOVAs with LIQUID and WORD POSITION as factors were conducted. Tests for main effects of word position on constriction formation, release duration, and pathlength, as well as absolute and proportional time-to-peak velocity are presented below. These dependent variables were chosen for examination in the expectation that they would be the most likely to exhibit positional effects such as word-final lengthening.

Word position had no significant effect on either constriction release duration or pathlength. However, word position did have a significant effect on both constriction formation duration and pathlength. Constriction formation duration was longer word-finally for all liquids [ $F(1,194) = 141.903, p = 0.0001$ ], and pathlength was longer word-finally [ $F(1,194) = 175.4454, p = 0.0001$ ] for [l], [r], and [ɹ] but not for [ɹ]. Finally, proportional time to peak velocity was longer word-finally [ $F(1,194) = 28.0729, p = 0.0001$ ] for all the liquids and absolute time-to-peak velocity was longer word-finally [ $F(1,194) = 30.2707, p = 0.0001$ ] for [l], [r], and [ɹ] but again not for [ɹ]. While it is conceivable that the speaker employed a phrase boundary after the target word, indicating phrasal rather than word level lengthening, the experimenters, having listened to the sentences, believe this to be unlikely. It is more likely that these data contain instances of word-final lengthening.<sup>3</sup>

## III. ARTICULATORY—ACOUSTIC RELATIONS

In this section, a description of the acoustic characteristics of Tamil liquid consonants and an investigation of the

basic underlying articulatory-acoustic relations are provided. As seen in the previous sections, Tamil has lateral and rhotic approximants, occurring as both retroflex and nonretroflex. In general, both laterals and rhotics tend to have a formant structure similar to that of vowels (Ladefoged, 1982). The lateral sounds of the world's languages exhibit a wide variety of articulation and of concomitant acoustic characteristics (Ladefoged and Maddieson, 1996). Both lateral consonants in Tamil are voiced lateral approximants. Ladefoged and Maddieson (1996) summarize the general acoustic characteristics of voiced lateral approximants as follows: an  $F_1$  rather low in frequency, an  $F_2$  that may have a center frequency anywhere within a fairly wide range depending on the location of the occlusion and the profile of the tongue, an  $F_3$  with a relatively strong amplitude and high frequency, and possibly several closely spaced formants above the frequency of  $F_3$ .

There is an even wider range of articulatory variation among the rhotic sounds of the world's languages. Rhotic sounds have been associated with lowered third formant frequency (Lindau, 1985) based mainly on data from English (e.g., Delattre and Freeman, 1968; Espy-Wilson, 1992). (See for examples of exceptions, such as in uvular and dental r-sounds, Ladefoged and Maddieson, 1996, pp. 244–245.) The acoustic correlate of the retroflex consonantal posture is the general lowering of the third and fourth formants (Ladefoged, 1982). More specifically, Fant (1968) associates retroflex modification of alveolar sounds to the lowering of  $F_4$  frequency so that it comes close to  $F_3$ , and the retroflex modification of palatal sounds to the lowering of  $F_3$  frequency so that it comes close to  $F_2$ .

### A. Acoustic characteristics of Tamil liquids

Sample acoustic spectra of the five Tamil liquids are shown in Fig. 8. These acoustic spectra were obtained from /paC/ utterances recorded in a quiet sound booth by subject SN. The data were directly digitized onto a SUN Workstation at 32 kHz and later down-sampled to 8 kHz. The final liquid consonant in these utterances was artificially sustained in a procedure similar to that followed during the MRI experiments. DFT and 14th-order LPC spectra (Fig. 8) were calculated from 25-ms Hanning-windowed segments taken at approximately 50 ms into the production of the liquid. Formant values of the sustained liquids were found to be consistent across repetitions.

All five liquids were characterized by an  $F_1$  in the 400–450 Hz range. The dental [l] had a clearly defined  $F_2$  at 1200 Hz, a relatively broad spectral peak signifying  $F_3$  around 2400 Hz, and an  $F_4$  peak around 3850 Hz.  $F_2$  and  $F_3$  of the retroflex lateral [ɭ], on the other hand, appeared close to each other, around 1460 and 1800 Hz.  $F_4$  and  $F_5$  for [ɭ] were around 2500 Hz and 3600 Hz, respectively. In addition, there appeared to be a spectral zero around 3500 Hz for [l] and around 3300 Hz for [ɭ]. It may be that this zero was responsible for the decrease in the prominence of  $F_4$  and  $F_5$ . In general, these formant frequency values for this subject's laterals are comparable to the data from a male Tamil speaker reported by Ramasubramanian and Thosar (1971): average  $F_1$ ,  $F_2$ , and  $F_3$  frequency values (in Hz) of

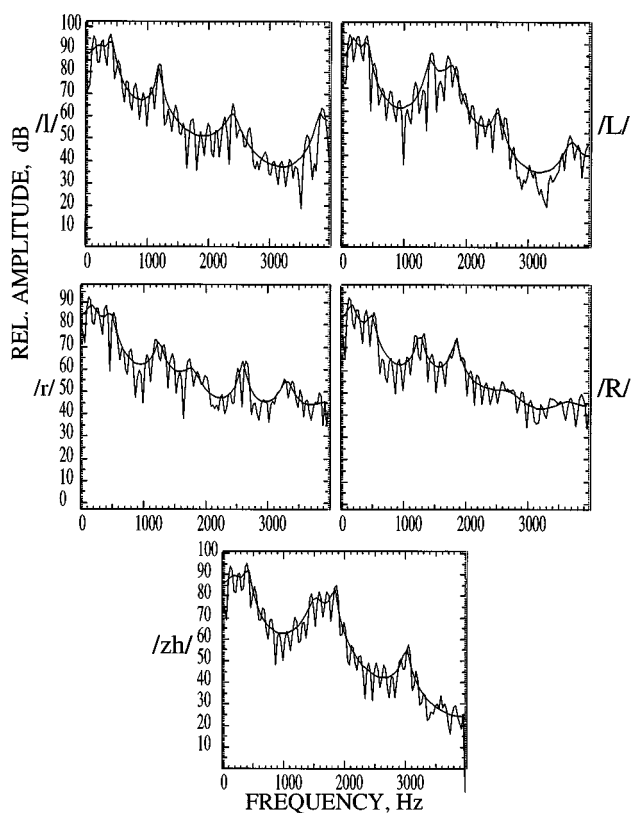


FIG. 8. Sample acoustic spectra of five Tamil liquids (DFT with an LPC overlay).

300, 1300, and 3000, respectively, for /l/; and 280, 1450, and 1600, respectively, for /ɭ/ (both produced in inter-vocalic position).

The acoustic characteristics of the prealveolar and the postalveolar rhotics were very similar to one another except for a slight difference in  $F_3$  values: 1700 Hz versus 1850 Hz, respectively. In general, the average difference between the  $F_3$  values of the [r] and [ɽ] was about 100 Hz.  $F_2$ ,  $F_4$ , and  $F_5$  are around 1200, 2600, and 3600 Hz for both [r] and [ɽ].

The low-frequency acoustic characteristics of the retroflex approximant /ɻ/ were strikingly similar to those of the retroflex lateral [ɭ]:  $F_2$  and  $F_3$  were close to each other, around 1500 and 1850 Hz, respectively.  $F_4$ , however, was around 3000 Hz, about 500 Hz higher than the  $F_4$  of [ɭ]. No significant spectral peaks or valleys were evident above 3000 Hz.

### B. Articulatory-acoustic relations in Tamil liquids

In this section, the basic articulatory-acoustic relations in Tamil liquids are outlined using vocal tract dimensions derived from MRI data. Figure 9 shows the area functions for the five Tamil liquids of subject SN. The numerical area-function values are given in Appendix B. For the analysis presented in this section, planar acoustic wave propagation in the vocal tract was assumed. Furthermore, source-filter separability was assumed, and the effects of the vocal tract bend and vocal tract losses were ignored. Under these assumptions, the vocal tract cavities can be roughly approximated as concatenated uniform cylindrical tube sections. As a further approximation, only averaged areas of the vocal tract cavities

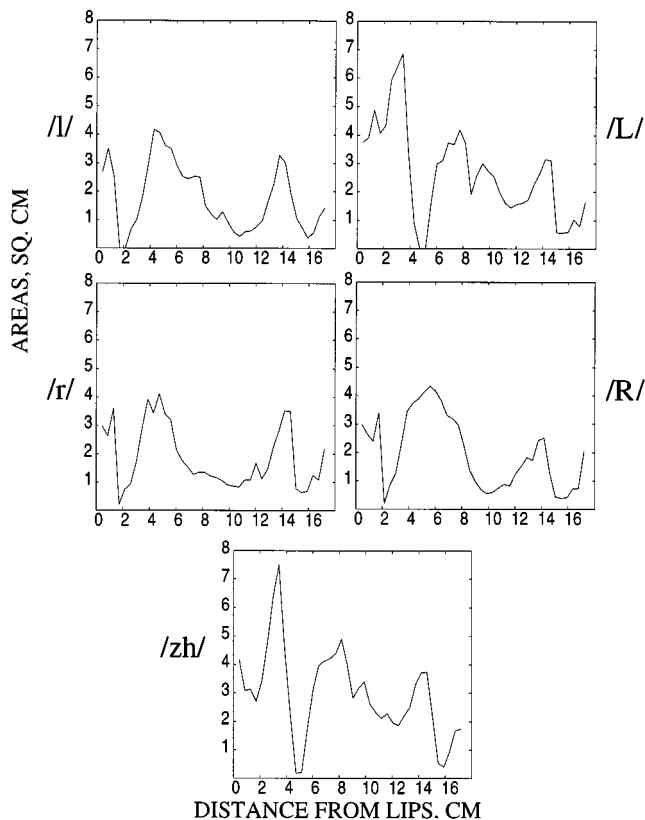


FIG. 9. MRI-derived area functions for five Tamil liquids.

to the front and back of the oral constriction were considered. The following abbreviations are used:  $L_b$  and  $A_b$  refer to back-cavity length and area;  $L_f$  and  $A_f$  refer to front-cavity length and area;  $L_c$  and  $A_c$  refer to constriction length and area.  $L_s$  and  $A_s$  refer to sublingual cavity length and area (for rhotics) or side-cavity length and area (for laterals). Although a somewhat oversimplified approximation, the analysis of such tube models is aimed at providing insights into the origin of the poles and zeroes of the vocal tract transfer function. [Acoustic modeling experiments wherein area functions of Tamil liquids are directly input to an articulatory synthesizer are reported in Narayanan and Kaun (1999).] The analysis presented here followed the general acoustic modeling principles proposed in Flanagan (1972) and Stevens (1998).

The tube dimensions obtained from the MRI data for each sound (Fig. 9) are listed in Table II. For the laterals, the

TABLE II. Vocal tract tube dimensions estimated from MRI-derived area functions.  $L_b$  and  $A_b$  refer to back-cavity length and area,  $L_f$  and  $A_f$  to front-cavity length and area, and  $L_c$  and  $A_c$  to the constriction length and area.  $L_s$  and  $A_s$  refer to the sublingual length and area (for [r] and [ɾ]) or the lateral channel length and area (for [l] and [ɭ]).

Liquid	$L_b$ (cm)	$A_b$ (cm <sup>2</sup> )	$L_f$ (cm)	$A_f$ (cm <sup>2</sup> )	$L_s$ (cm)	$A_s$ (cm <sup>2</sup> )	$L_c$ (cm)	$A_c$ (cm <sup>2</sup> )
[l]	14.62	1.83	1.72	3.3	2.5	2.5	0.86	0.5
[ɭ]	11.61	2.4	4.73	4.75	2.58	2.5	0.86	0.4
[r]	14.6	1.89	1.72	1.89	2.8	0.85	0.86	0.2
[ɾ]	14.19	2.02	2.15	2.02	2.15	1.26	0.86	0.2
[ɻ]	11.61	2.88	4.73	2.88	-	-	0.86	0.18

TABLE III. Average frequency values of the first four formants (in Hz) for the five Tamil liquids. Obsv. refers to values measured from spectra of natural speech and Est. refers to values estimated by the tube models (Sec. IV B).

Liquid	F1		F2		F3		F4	
	Obsv.	Est.	Obsv.	Est.	Obsv.	Est.	Obsv.	Est.
[l]	400	480	1200	1180	2400	2360	3850	3540
[ɭ]	400	540	1460	1480	1800	1822	2500	2460
[r]	460	507	1200	1200	1700	1765	2600	2400
[ɾ]	460	507	1200	1233	1850	1935	2600	2466
[ɻ]	450	433	1500	1485	1850	1822	3025	2970

areas of both the side channels were combined and modeled as a single channel.  $L_c$  for the laterals was taken to be the length of the medial tongue occlusion, while  $A_c$  was the average combined side channel area over the occlusion region. For the rhotic sounds, the tongue-tip constriction region, defined by the minimum area in the area function, was taken to be approximately two area function sections, that is, 0.86 cm. Once the constriction region was defined, the front- and back-cavity lengths were measured, respectively, from the lip opening to the constriction-region beginning and from the glottis to the constriction-region ending. With the aid of cross-sectional MRI data, the side-cavity length  $L_s$  for the laterals was defined to be the length along which a space existed between the sides of the convex tongue surface contour and the oral cavity wall. The sublingual length  $L_s$  for the alveolar rhotics was defined to be the oral cavity length for which a distinct airspace area under the tongue surface was measurable.

The vocal tract resonance frequency values were estimated from the tube model and were compared to the measured values obtained from the subject's natural speech. These results are summarized in Table III. Note that for all the calculations reported below, the speed of sound in air,  $c$ , equals 34480 cm/s.

F1 for all the liquids can be attributed to the natural frequency of the Helmholtz resonator formed by the back cavity and the constriction created by the anterior tongue. The estimated Helmholtz resonance values  $[(c/2\pi) \times \sqrt{A_c/(L_b^* A_b^* L_c)}]$  ranged between 400 and 500 Hz. In spite of significant differences in back-cavity lengths across the five liquids, differences in the volume of the Helmholtz resonator were relatively minimal. Tongue shaping helped achieve this back volume "normalization:" the back-cavity volume in the case of the retroflex sounds was made larger by either concave or pitted surface contour (as in the case of [ɾ]) or a lowered posterior tongue body (as in the case of [ɭ]). On the other hand, the rhotics and the dental [l] had a relatively higher mid- and posterior tongue body height and either a convex contour ([l]) or a less concave contour (rhotics).

In a highly simplified approximation, the vocal tract configuration for the lateral sounds can be modeled by just two tubes representing the back and front cavities that are separated by a constriction. In such a case, the back-cavity resonances  $[(N^*c)/2L_b, N=1,2,3,\dots]$  for ([l]) were estimated at 1180 Hz, 2358 Hz, 3540 Hz, etc.; while for [ɭ], they



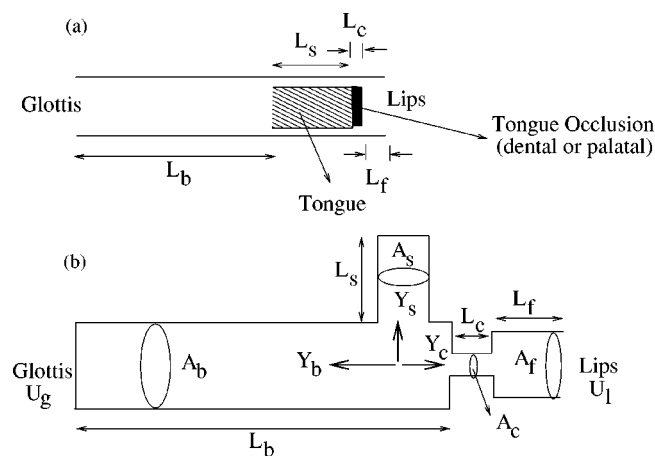


FIG. 10. (a) Schematic of the oral cavity configuration (as viewed looking down from the palate) for the lateral sounds. (b) A tube model approximation for the lateral sounds (see Sec. III for explanation of the various symbols).

were at 1485 Hz, 2975 Hz, 4455 Hz, etc. Frequency of the lowest front-cavity resonance  $[=c/4*L_f]$  was above 5000 Hz for [l], while the relatively longer front cavity for [ɭ] yielded a smaller frequency value of around 1820 Hz for the lowest front-cavity resonance. Hence for /l/,  $F_2$ ,  $F_3$ , and  $F_4$  are attributed to the first three back-cavity resonances. For /ɭ/,  $F_2$  and  $F_4$  are attributed to back-cavity resonances, while  $F_3$  is the lowest front-cavity resonance. While these estimated values provide some gross indications of the vocal tract cavity affiliations of the peaks in the natural speech spectra, they are not adequate for explaining the zeroes in the spectrum.

Consider now a more detailed model that includes the effect of the lateral channel [a schematic is shown in Fig. 10(a)]. The corresponding tube model is shown in Fig. 10(b). Stevens (1998) has proposed a similar model for laterals that explicitly includes a side channel. As a result of the presence of lateral channels along the sides of the tongue due to the medial tongue occlusion, a cavity of approximately the length of the lateral channel is created behind the medial tongue occlusion. This cavity extends between the medial occlusion created by the anterior tongue (in the dental region for [l] and the palatal region for [ɭ]) on one end and is open to the oral cavity in the direction of the pharynx. This cavity is modeled as a shunt branch while the flow along the sides of the tongue is combined and modeled as a single channel representing the main airway path. The modified back-cavity length is  $(L_b - L_s)$ . Assuming that the front and back cavities can be decoupled, the poles of the transfer function  $U_l/U_g$  between the volume velocity at the lips,  $U_l$ , and at the glottis,  $U_g$ , would include the natural resonances of the combined side channel and the modified back cavity and the natural resonances of the front cavity. The zeroes for the transfer function  $U_l/U_g$  occur at frequencies for which the acoustic impedance,  $Z_s$ , looking into the cavity created as a result of the medial tongue occlusion and side airway branch creation, is zero. If  $L_s$  is the length of the uniform tube approximation of this cavity, then the frequency of the lowest zero occurs at  $[=c/4*L_s]$ .

In summary, the analysis for the Tamil lateral approxi-

ments yielded the following results (Table III). For [l], the lowest zero frequency was at 3450 Hz. This zero is at the frequency for which the impedance looking into the side branch is zero. The poles of the transfer function were at 1180 Hz, 2360 Hz, 3540 Hz, and 4720 Hz. Due to the relatively short length of the cavity anterior to oral constriction for the dental [l], there was no contribution to the spectrum from the resonances of the front cavity in the range of interest below 5 kHz (the frequency of the lowest front-cavity resonance,  $[=c/4*L_f]$ , is around 5 kHz). Approximate analysis suggests that  $F_1$  is the Helmholtz resonance between the constriction and the back-cavity volume, while  $F_2$ ,  $F_3$  and  $F_4$  are half-wavelength resonances of the combined back and side cavities.

For [ɭ], the lowest zero frequency was around 3340 Hz. Estimated poles from the cavities behind the oral constriction were at 1485 Hz, 2970 Hz, and 4454 Hz. The lowest front-cavity resonance was at 1822 Hz. Hence the formant frequencies for [ɭ] can be summarized as  $F_1=540$  Hz,  $F_2=1480$  Hz,  $F_3=1822$  Hz,  $F_4=2460$  Hz, and  $F_5=3650$  Hz.  $F_2$ ,  $F_3$ , and  $F_5$  are combined back- and side-cavity resonances, while  $F_3$  is attributed to the front cavity. Compared to [l],  $F_2$  and  $F_3$  in [ɭ] were closer to each other mainly due to the longer front cavity resulting from the constriction in the palatal region created by the retroflexed tongue.

As a first approximation, a two cavity tube can be used to model the acoustics of the Tamil rhotic sounds [r] and [ɽ]. Notice that the front cavity is modified by a significant (sublingual) area under the raised tongue tip (Fig. 2). As a further approximation, the region in front of the tongue tip constriction can be modeled as a single cavity that extends between the lip opening and the end of the sublingual space. A justification for this approximation comes from the fact that the sublingual opening for the vocal tract configurations of Tamil [r] and [ɽ] is relatively large compared to the cross-sectional areas of the sublingual cavity. Under this approximation, the effective front-cavity length is increased by the amount of the sublingual cavity length. This results in an effective front-cavity resonance frequency,  $[=c/4*L_f]$ , of 1907 Hz and 2004 Hz for [r] and [ɽ], respectively. A similar approach wherein the sublingual space is modeled as an extension of the front cavity has been found to produce satisfactory prediction of  $F_3$  frequencies for American English /r/ (Espy-Wilson *et al.*, 1997). An additional correction for lip rounding (which is not accurately measurable from the MRI data) will further lower the front-cavity resonance by about 100 Hz. The back-cavity resonances for [r] and [ɽ],  $[(N*c)/2L_b, N=1,2,3,\dots]$ , are around 1200 Hz, 2400 Hz, 3600 Hz, etc. In summary (Table III),  $F_2$  for the rhotics [r] and [ɽ] corresponds to the lowest back-cavity resonance (around 1200 Hz).  $F_3$  corresponds to the lowest sublingual-inclusive front-cavity resonance (about 1800 Hz for [r] and 1900 Hz for [ɽ]).  $F_4$  corresponds to the back-cavity resonance at 2400 Hz.

A more accurate model would include a separate side branch cavity to model the sublingual effect, cf., sublingual models proposed for postalveolar fricatives (Narayanan, 1996), American English /r/ (Espy-Wilson *et al.*, 1997), and

“retroflex /r/” (Stevens, 1998). Such a model would also yield a zero in the frequency range of interest (3100 Hz and 4000 Hz for the sublingual dimensions of [r] and [ɽ] noted in Table III, respectively). This zero may contribute to lowering the amplitude of the high-frequency formants  $F4$  and  $F5$ , as seen in the spectrum shown for [ɽ] (Fig. 8). However, it should be noted that the lack of a clear indication of zeroes in the spectra of voiced sounds is not uncommon, given their harmonic structure, as in the case of the spectrum of [r] in Fig. 8.

Finally, the retroflex approximant /ɻ/ is considered. The front- and back-cavity dimensions are very similar to those of /l/ explaining the similarity in their low-frequency spectral characteristics ( $F2$  and  $F3$  close to each other at 1485 and 1800 Hz, due to the lowest back- and front-cavity resonances). The only other significant peak around 2975 Hz is attributed to a back-cavity resonance. The lack of significant lateral channels is reflected by the lack of zeros in the high-frequency region (2.5–5 kHz), distinct from the spectrum of /l/.

The formant frequency values estimated from the tube models, assuming decoupled front and back cavities, correspond fairly well with the values measured from the natural speech of subject SN, as shown in Table III. In summary, using a simple one-dimensional acoustic model and vocal tract data derived from MRI, it was possible to establish the basic relations between vocal tract cavities and formant structure.

## IV. DISCUSSION

### A. Regarding the significance of tongue shaping

Our data lead to an interesting conclusion: tongue shape mechanisms are the unifying characteristic within the phonetic class of laterals and within the phonetic class of rhotics; constriction location serves to differentiate members within each class.

In Tamil phonology, the laterals pattern differently from the rhotics. With respect to their phonotactic distribution, the laterals are systematically more consonant-like. For example, the laterals along with nasal, stop, and fricative consonants occur singly or as geminates; whereas the rhotics, like the glides, occur only singly. Furthermore, rhotics and glides can occur before a tautosyllabic consonant. Laterals and all other consonants are banned from this position. The rhotics, like vowels, lack medial tongue-palate closure (or near closure as seen in fricatives) and instantiate simple articulatory-acoustic relations that are predictable, a characteristic associated with vowels more than with consonants.

#### 1. Descriptive classification

Let us begin with the class of laterals. The mechanisms of [l] production are the same as those of [l̥]: an oral constriction along the midsagittal line is formed, lateral channels along the sides of the tongue are created, and a convex posterior tongue body surface facilitates coupling between the front and back cavities. In addition, [l̥] and [l] share similarly shaped (i.e., skewed) velocity profiles (intermediate values of proportional time to peak velocity). It is proposed that

these mechanisms characterize the class Lateral. The Tamil laterals are distinguished from each other by means of the location of the tongue blade constriction: the medial oral constriction for [l̥] occurs in the palatal region, while the constriction for [l] is alveolar. In terms of dynamics, the constriction formation and release durations, pathlengths, and velocities of [l̥] are greater than those of [l]. The retroflex lateral [ɻ] has a larger front cavity and a correspondingly low front-cavity resonance value ( $F3$ ), thus distinguishing it acoustically from [l].

The similarity in tongue shaping for the rhotic class is even more striking. In particular, it was seen above that tongue shaping is essentially identical for [r] and [ɽ]: a raised tongue-tip (apical) constriction, a tongue body gradually lowered in the antero-posterior direction, and slight concavity of the tongue surface. Additionally, [r] and [ɽ] share similarly shaped (i.e., skewed) velocity profiles (late proportional time-to-peak velocity). The members of this category are distinguished from one another by constriction location, [ɽ] being articulated with a slightly more anterior tongue constriction than [r]. In terms of contrast in dynamics, [ɽ] has greater constriction pathlengths and velocity values than [r], but somewhat overlapping duration values.

The articulatory data justify the postulation of a class of rhotics and a class of laterals in Tamil, but do not provide evidence in favor of a larger class of liquids. Such a super-class appears to have largely an acoustic basis. Both rhotics and laterals are voiced sonorants and have similarities in formant structure. The first formant (400–500 Hz) is due to the Helmholtz resonance of back-cavity constriction. The second formant (1200–1600 Hz) is the lowest resonance of the back cavity. (Note that, as discussed in Sec. III, the value of the third formant, which distinguishes members of the liquid class from one another—around 2400 Hz for [l] and significantly less for the other liquids—is influenced by where and how the oral constriction is formed.) In the liquid class the first three formants are most prominent, while higher formants are strongly reduced in intensity, typically due to zeroes in the spectra. The articulatory origin of the zeroes varies: they may arise due to sublingual space, as in the case of certain rhotics, or due to side-channel effects, as in case of laterals. However, the presence of a zero in the region of high formants, along with voicing and a common  $F1$  and  $F2$  structure, is another acoustic characteristic of the class of liquids.

Let us next turn to the more elusive consonant /ɻ/. Based on the theory outlined above, we are now in a position to place /ɻ/ within one of the segment classes of Tamil in a principled manner. The overall tongue shape mechanisms of /ɻ/ clearly are not those associated with the production of the lateral class. The laterals are characterized by medial tongue closure, a laterally inward drawn tongue body, and a convex posterior tongue surface. The segment /ɻ/ involves no medial closure, a lateral bracing of the tongue body, and a concave posterior tongue surface. The rhotics /r/ and /ɽ/ have in common with /ɻ/ a medial tongue-tip constriction (but no closure) and an absence of lateral channels. Furthermore, the tongue surface contour behind the oral constriction is concave in /r/ and /ɽ/, and even more so in /ɻ/. From this it can be con-

cluded that the Tamil /ɻ/ is properly categorized as a rhotic.

Our data have provided us with hitherto unavailable information regarding the production of Tamil liquids. This information has allowed us to propose a principled system of liquid categorization and demonstrates that the classification of /ɻ/ as anything other than a rhotic is without justification.

Two questions may be posed now. First, why has /ɻ/ often been categorized as a lateral? And, second, what articulatory-acoustic relations play a role in dialectal substitutions and mergers that occur among these liquids? The latter question is addressed in the following section. In regards to the first question, it is our view that the affinity between /l/ and /ɻ/ is based on their extreme retroflexion. This retroflexion gives rise to a large front cavity and, consequently, an acoustic similarity between /l/ and /ɻ/. In particular, both exhibit proximity of  $F2$  and  $F3$ . There is a dynamic affinity between /l/ and /ɻ/ as well. While in /r/ and /r̥/, tongue-tip peak velocity is attained proportionally late, in /ɻ/ and the laterals (particularly /l/), peak velocity is reached proportionally early. Also, due to their retroflex articulation, the tongue-tip constriction formation and release curvature is counterclockwise (i.e., back-to-front) in both /l/ and /ɻ/. It is thus understandable and correct that /l/ and /ɻ/ should be grouped together in classification—both are retroflex sounds.

## 2. Linguistic substitutions and mergers

Recall that dialectal substitutions and mergers are known to take place among these liquids. In some dialects the acoustically similar /r/ and /r̥/ have merged or are in the process of merging (Christdas, 1988, p. 131). It was shown that for this speaker these sounds are produced with a tongue-tip constriction that differs only minimally in horizontal position (see Fig. 5), and that these sounds share a similar formant structure. Overall for this speaker the distinction between /r/ and /r̥/ is fragile in both the articulatory and auditory domains. Small variations in the production of either or both sounds would be likely to obliterate this subtle distinction. Thus we speculate that the linguistic merger of Tamil /r/ and /r̥/ might be a consequence of minimal articulatory distinction and substantial auditory overlap.

Furthermore, certain Tamil speakers exhibit substitution of /l/ for /ɻ/ (Christdas, 1988, p. 160). Recall that for the speaker in the present study, /l/ and /ɻ/ share similar acoustic spectral characteristics, particularly in their first three formants. In particular, both /l/ and /ɻ/ have a low  $F3$  frequency close to that of  $F2$ , due to a relatively long front cavity. Recall also from the MRI and palatography data that the production of /ɻ/ involves, in addition to creating the retroflex constriction, a pitted posterior tongue cavity. The posterior tongue body shaping for /ɻ/ involves coordination of both internal tongue muscles and external forces provided by linguopalatal bracing. For /l/, on the other hand, the posterior tongue body shaping is primarily executed by internal muscular action. Further, both the absolute time-to-peak velocity and the proportional time-to-peak velocity of the tongue tip are earlier in /ɻ/ than /l/. The remaining liquids, /r/ and /r̥/, which have the latest proportional time-to-peak velocity of the tongue tip, are characterized by the simplest tongue

shapes. This is consistent with the notion that tip peak velocity is reached comparatively earlier in sounds that require more complex tongue body shaping. We speculate that it is perhaps the more complex tongue-shaping requirement, above and beyond the creation of a retroflex constriction, for producing /ɻ/ that is compromised in the neutralization of the /l/-/ɻ/ contrast by some speakers yielding the single variant /l/.

## 3. Dimensions of contrast

The diversity of tongue shapes and dynamics are made possible, at least in part, through different lingua-palatal bracing mechanisms (see Stone *et al.*, 1991, 1992; Narayanan *et al.*, 1997; Alwan *et al.*, 1997). The tongue shape geometry and the temporal characteristics of their realization give rise to systematic differences in the acoustic resonance behavior of the vocal tract. These data suggest that speech production targets go beyond simple constriction targets in the midsagittal plane. This is not surprising considering the complexity of behavior possible for the tongue, “a boneless, jointless structure...[that] can elevate, depress, widen, narrow, extend, and retract...and move differentially, both laterally-to-medially and left-to-right” (Stone *et al.*, 1991). The targets that the speech production system works to attain by calling on these many degrees of freedom must be able to yield complex tongue shapes, such as the pitted tongue surface contour observed for [ɻ] articulation. Furthermore, our data also suggest that different aspects of vocal tract configuration (e.g., tongue position and shape) are constrained in a phoneme-specific manner. For example, among the five liquids, [ɻ] exhibits the most variability in tongue-tip position while being fairly consistent in tongue shaping. When compared to the laterals this is understandable because the tongue tip is free in [ɻ] and not in the laterals. Rhotics [r] and [r̥] could in principle exhibit variability, but in fact, only [ɻ] shows significant tongue-tip position variability. The two anterior rhotics are only subtly distinct; any divergence from prototypical tongue-tip position would be likely to increase the chance of confusion between the two phones. (In those dialects exhibiting [r]-[r̥] merger, one might speculate that tongue-tip freedom has given rise to acoustic overlap.) In summary, this suggests that for certain sounds the realization of a particular constriction position may be less narrowly constrained than is the realization of tongue shape for that sound.

## B. Comparison to English liquids

The liquids of Tamil and English can be compared using published articulatory and acoustic data on English liquids. Articulatory data on English /l/'s have been obtained using various techniques such as x-ray imaging (Giles and Moll, 1975), magnetometry (Sproat and Fujimura, 1993), ultrasound imaging (Stone *et al.*, 1992), MRI (Narayanan *et al.*, 1997), and electropalatography (Stone *et al.*, 1992; Narayanan *et al.*, 1997). Espy-Wilson (1992) provides a comprehensive characterization of the acoustics of American English /l/. In English, the lateral approximant is voiced and has been broadly classified into light and dark varieties, occur-



ring typically in pre-vocalic and post-vocalic contexts, respectively. Both types, in general, show contact of the anterior tongue in the dental or alveolar region and an inward lateral compression of the posterior tongue body (toward the midsagittal plane) that facilitates the creation of lateral channels along the sides of the tongue. The areas of these channels are typically unequal. The contours of the tongue surface cross-sections are convex.

The articulatory and acoustic characteristics of the Tamil dental [ɭ] and the American English (AE) [ɭ] appear very similar. Both have tongue-tip constrictions in the dental region, lateral linguapalatal contacts of the anterior tongue, convex tongue surface contours, and side airflow channels created by a laterally inward-drawn tongue body. Like the AE [ɭ], the dental [ɭ] in Tamil can appear in both syllable-initial and -final positions, word-initially, and word-finally. The dark AE [ɭ], usually appearing syllable-finally, is typically characterized by narrow upper pharyngeal areas due to high position of the posterior tongue body raising and retraction of the tongue root (Sproat and Fujimura, 1993; Narayanan *et al.*, 1997). Such “pharyngealization” contributes to *F2* lowering in dark [ɭ] when compared to light [ɭ]. The MRI data of the Tamil [ɭ] also shows such pharyngealization. Recall that the Tamil [ɭ] was produced in the /paC/ context during the MRI experiment. Unfortunately, it is not possible to tell from the MRI data whether there are any systematic differences in the phonetic implementation of [ɭ] depending on its relative position within the syllable. (The magnetometer procedure in turn offers no information on tongue root position.)

The production of English /ɹ/ has been observed to exhibit considerable variability in its articulatory configurations (Delattre and Freeman, 1968; Alwan *et al.*, 1997; Westbury *et al.*, 1998). Nonetheless, /ɹ/ shows acoustic stability, i.e., different vocal tract shapes yield essentially equivalent acoustic output. English /ɹ/ is characterized by a stable acoustic pattern of lowering of the third formant frequency close to that of the second formant (Boyce and Espy-Wilson, 1997). It has long been claimed, for instance, that American English /ɹ/ can be either “retroflex” or “bunched” (Ladefoged and Maddieson, 1996). In fact, the /ɹ/ of English exhibits a continuum of vocal tract shapes across speakers that yields a regular acoustic output (Delattre and Freeman, 1968; Lindau, 1985; Alwan *et al.*, 1997; Boyce and Espy-Wilson, 1997; Westbury *et al.*, 1998; see also Guenther *et al.*, 1999).

The pre- and post-alveolar /r/'s of Tamil do not show much variation in their vocal tract configurations. (As linguistically contrasting, it is not a surprise that the articulations of these sounds are largely kept distinct.) Interestingly, these sounds occupy adjacent and somewhat overlapping regions of the articulatory continuum identified for English. The relevant region is that which characterizes the so-called “tip-up” American English /ɹ/ (cf. Delattre and Freeman’s British “retroflex” /ɹ/). Given that the tongue shape continuum for English corresponds to a unified acoustic outcome, and given that the Tamil /r/ and /ɹ/ fall on this continuum, these Tamil sounds are expected to share a similar formant structure. This expectation is borne out.

American English lacks strongly retroflexed sounds,

nonetheless a fruitful comparison between English /ɹ/ and Tamil /ɻ/ can be made. Such a comparison serves to demystify the elusive /ɻ/. The Tamil /ɻ/ shares important properties with the English rhotic—particularly the “bunched” variety—although it does not fall directly on the same vocal tract shape continuum. In both sounds, the raised anterior tongue (bunched in AE /ɹ/, retroflexed in Tamil /ɻ/) is accompanied by a prominently concave posterior tongue region that appears as a pitted cavity. In both AE (Alwan *et al.*, 1997) and Tamil, the posterior tongue body shaping is facilitated by bracing of the tongue body against the palate. The dimensions of this pitted back cavity are similar in both sounds and yield similar Helmholtz resonances (*F1*). Low *F3* values in both languages result from a relatively large front cavity. Both languages achieve this by means of an inward-drawn anterior tongue (for AE Alwan *et al.*, 1997), although the front-cavity volume for Tamil /ɻ/ is slightly larger because the tongue tip, rather than the tongue body, is retracted. Thus while both the AE /ɹ/ and Tamil /ɻ/ are rather exotic sounds cross-linguistically, they turn out to be surprisingly similar both articulatorily and acoustically.

## V. CONCLUDING REMARKS

The articulatory characterization of Tamil’s unusual set of five contrasting liquids has been examined both in terms of articulatory geometry and kinematics. Although the study depended on only a single experimental subject, the use of multiple techniques for investigating both static and dynamic articulatory characteristics—static palatography, magnetic resonance imaging (MRI), and magnetometry (EMMA)—has helped provide a better understanding of the production dynamics and articulatory-acoustic mappings for these liquids.

Current models of speech production often assume that constriction position, defined in the midsagittal plane, is the main “place of articulation” parameter. However, this study suggests that articulation cannot be characterized solely by identifying constriction position. Rather, three-dimensional tongue shape and the dynamics underlying shape formation are critical to understanding natural linguistic classes and phonological phenomena.

## ACKNOWLEDGMENTS

Work supported in part by NSF Grant No. IRI-9503089 (S.N.) and NIH Grants Nos. DC-03172, DC-00016 (D.B., Haskins Laboratories). The authors wish to acknowledge the technical support of Kate Haker (Cedars-Sinai Hospital, Los Angeles, CA), and Louis Goldstein and Walter Naito (Haskins Laboratories, New Haven, CT). The authors thank Abeer Alwan and Peter Ladefoged for their expertise and advice; and Pierre Badin, Carol Espy-Wilson, and an anonymous reviewer for their helpful comments.



**APPENDIX A: STIMULI AS PRESENTED TO THE SUBJECT FOR THE EMMA EXPERIMENT. AN ENGLISH GLOSS HAS BEEN ADDED TO RIGHT**

Andha vakyam “kali” perusu.	The utterance	9.46	128.76	300.37	106.79	68.55	316.41
	“fate” is-big.	9.89	90.75	273.56	90.97	54.49	340.14
Andha vakyam “kaLi” perusu.	...“pudding”...	10.32	57.57	254.25	86.57	59.77	260.74
Andha vakyam “kazhi” perusu.	...“pole”...	10.75	41.53	198.85	83.06	74.71	232.47
Andha vakyam “kari” perusu.	...“coal”...	11.18	58.67	160.62	108.11	88.11	210.94
Andha vakyam “kaRi” perusu.	...“curry”...	11.61	61.52	144.14	108.54	82.84	228.52
Andha vakyam “pali” perusu.	...“sacrifice”...	12.04	76.46	156.67	167.43	124.37	196.44
Andha vakyam “paLi” perusu.	...nonce...	12.47	97.12	161.23	112.72	151.22	186.11
Andha vakyam “pazhi” perusu.	...“blame”...	12.9	165.89	172.49	144.58	183.03	220
Andha vakyam “pari” perusu.	...“horse”...	13.33	225.43	224.34	222.8	172.49	249.39
Andha vakyam “paRi” perusu.	...“pluck”...	13.76	327.54	264.99	281.03	242.58	329.59
Andha vakyam “val” perusu.	...“tail”...	14.19	303	315.75	352	251.59	371.78
Andha vakyam “vaL” perusu.	...“sword”...	14.62	193.8	310.63	349.58	128.1	373.97
Andha vakyam “vazh” perusu.	...“live”...	15.05	108.98	57.55	77.12	43.73	218.63
Andha vakyam “var” perusu.	...“strap”...	15.48	72.73	55.81	64.16	39.11	53.83
Andha vakyam “vaR(pu)” perusu.	...“mold”...	15.91	36.47	59.33	68.33	40.43	40
Andha vakyam “alai” perusu.	...“wave”...	16.34	55.37	102.83	124.58	72.29	94.48
Andha vakyam “aLai” perusu.	...“sift”...	16.77	111.4	78.66	108.98	73.61	167.87
Andha vakyam “azhai” perusu.	...“invite”...	17.2	141.94	164.36	217.53	205.44	174.02
Andha vakyam “arai” perusu.	...“half”...						
Andha vakyam “aRai” perusu.	...“room”...						
Andha vakyam “palam” perusu.	...“strength”...						
Andha vakyam “paLam” perusu.	...nonce...						
Andha vakyam “pazham” perusu.	...“fruit”...						
Andha vakyam “param” perusu.	...“almighty”...						
Andha vakyam “paRam” perusu.	...nonce...						
Andha vakyam “vaLi” perusu.	...“wind”...						
Andha vakyam “vaLam” perusu.	...“fertile”...						
Andha vakyam “paRavai” perusu.	...“bird”...						

<sup>a</sup> $D_{lips}$  (in cm) is distance from the lip opening.

$D_{lips}^a$ (cm)	Sublingual areas (mm <sup>2</sup> ) [r]	$D_{lips}^a$ (cm)	Sublingual areas (mm <sup>2</sup> ) [r]
1.72	258.40	2.15	269.58
2.15	181.39	2.58	166.11
2.58	164.79	3.01	119.97
3.01	75.81	3.44	107.23
3.44	33.18	3.87	22.63

<sup>a</sup> $D_{lips}$  (in cm) is distance from the lip opening.

**APPENDIX B: AREA FUNCTIONS FROM MRI DATA FOR THE FIVE TAMIL LIQUIDS**

A graphical representation is provided in Fig. 9.

$D_{lips}^a$ (cm)	Cross-sectional areas (mm <sup>2</sup> )				
	[l]	[ʃ]	[r]	[ɾ]	[ɻ]
0.43	267.63	375.29	299	300	418.8
0.86	351.78	391.99	264.11	264.33	309.16
1.29	254.44	487.79	358.59	239.94	312.23
1.72	0	406.71	20.27	340.14	271.36
2.15	0	434.18	74.27	20.25	342.55
2.58	68.45	595.02	93.1	86.46	480.32
3.01	101.23	637.43	171.61	126.86	639.62
3.44	182.81	687.08	289.6	231.81	750.17
3.87	296.63	335.23	392.43	346.07	455.05
4.30	417.7	89.87	343.65	375.07	223.87
4.73	407.15	0	412.21	390.01	17.23
5.16	361.01	0	341.89	414.38	19.95
5.59	350.9	166	320.14	434.72	173
6.02	291.8	299.21	215.99	415.94	311.8
6.45	252.91	310.91	177.1	382.1	396.3
6.88	243.92	373.97	154.69	330.25	412.21
7.31	252.47	367.16	128.1	319.04	421.44
7.74	250.49	418.8	135.13	296.19	440
8.17	148.97	372.22	134.69	221.7	489.12
8.6	120.85	191.6	121.95	136.89	399.68
9.03	101.95	258.4	116.24	96.9	283.45

$D_{lips}^a$ (cm)	Lateral channel areas (mm <sup>2</sup> ) [l]	$D_{lips}^a$ (cm)	Lateral channel areas (mm <sup>2</sup> ) [ʃ]
1.72	47.90	4.73	34.94
2.15	64.60	5.16	59.21
2.58	161.94	5.59	90.70
3.01	156.88	6.02	118.43
3.44	162.38	6.45	103.24
3.87	113.15	6.88	77.13
4.30	92.72	7.31	67.24
4.73	68.33	7.74	44.38

<sup>a</sup> $D_{lips}$  (in cm) is distance from the lip opening.

<sup>1</sup>Note that this means that no tokens of word final retroflex [ʃ] could be included in the analysis. Such multiple velocity extrema are typical of slowed speech. Why this particular liquid should be especially subject to this effect is an open question.

<sup>2</sup>The fact that their y-values are the same does not mean that they are effectively equivalent. Due to the rising slope of the alveolar ridge, [l] achieves complete closure while [ɾ], with the same y-value, achieves only approximation.

<sup>3</sup>Recall also that 17 word final tokens were excluded due to multiple tangential velocity minima at the peak. This velocity profile pattern is characteristic of slowed articulations. Thus, it seems likely that the 17 excluded tokens also underwent considerable word-final lengthening.

Alwan, A., Narayanan, S., and Haker, K. (1997). “Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part

- II. The rhotics," J. Acoust. Soc. Am. **101**, 1078–1089.
- Balasubramanian, T. (1972). "The phonetics of colloquial Tamil," Ph.D. dissertation, University of Edinburgh, Edinburgh.
- Boyce, S., and Espy-Wilson, C. (1997). "Coarticulatory stability in American English /r/," J. Acoust. Soc. Am. **101**, 3741–3753.
- Byrd, D., and Saltzman, E. (1998). "Intragestural dynamics of multiple phrasal boundaries," J. Phonetics **26**, 173–199.
- Christdas, P. (1988). "The phonology and morphology of Tamil," Ph.D. dissertation, Cornell University, Ithaca, NY.
- Delattre, P., and Freeman, D. C. (1968). "A dialect study of American r's by x-ray motion picture," Linguistics, An international review **44**, 29–68.
- Espy-Wilson, C. (1992). "Acoustic measures for linguistic features distinguishing the semivowels in American English," J. Acoust. Soc. Am. **92**, 736–757.
- Espy-Wilson, C., Narayanan, S., Boyce, S., and Alwan, A. (1997). "Acoustic modelling of American English /r/," Proceedings of Eurospeech 97, Rhodes, Greece, pp. 393–396.
- Fant, G. (1968). "Analysis and synthesis of speech processes," in *Manual of Phonetics*, edited by B. Malmberg (North-Holland, Amsterdam), pp. 171–272.
- Flanagan, J. (1972). *Speech Analysis, Synthesis and Perception* (Springer-Verlag, New York).
- Giles, S. B., and Moll, K. L. (1975). "Cinefluorographic study of selected allophones of English [l]," *Phonetica* **31**, 206–227.
- Gracco, V. L., and Nye, P. W. (1993). "Magnetometry in speech articulation research: Some misadventures on the road to enlightenment," Forschungsberichte des Institut für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) **31**, 91–104.
- Guenther, F. H., Espy-Wilson, C., Boyce, S., Matthies, M., Zandipour, M., and Perkell, J. (1999). "Articulatory tradeoffs reduce acoustic variability during American English /r/ production," J. Acoust. Soc. Am. **105**, 2854–2865.
- Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., and Kay, B. (1985). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling," J. Acoust. Soc. Am. **77**, 266–280.
- Ladefoged, P. (1957). "Use of palatography," *Journal of Speech and Hearing Disorders* **22**, 764–774.
- Ladefoged, P. (1982). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York).
- Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages* (Blackwell Publishers, Cambridge, MA).
- Lindau, M. (1985). "The story of r," in *Phonetic Linguistics*, edited by V. Fromkin (Academic, Orlando, FL), pp. 157–168.
- Löfqvist, A. (1993). "Electromagnetic transduction techniques in the study of speech motor control," Reports from the Department of Phonetics, University Umeå, PHONUM **2**, 87–106.
- Löfqvist, A., Gracco, V., and Nye, P. (1993). "Recording speech movements using magnetometry: One laboratory's experience," Forschungsberichte des Institut für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM), **31**, 143–162.
- Narayanan, S., Alwan, A., and Haker, K. (1995). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**, 1325–1347.
- Narayanan, S., Alwan, A., and Haker, K. (1997). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals," J. Acoust. Soc. Am. **101**, 1064–1077.
- Narayanan, S., and Alwan, A. (1996). "Parametric hybrid source models for voiced and voiceless fricative consonants," Proceedings of ICASSP **96**, pp. 337–340.
- Narayanan, S., Kaun, A., Byrd, D., Ladefoged, P., and Alwan, A. (1996). "Liquids in Tamil," edited by H. T. Bunnell and W. Idsardi, Proceedings of ICSLP 96, pp. 797–800.
- Narayanan, S., and Kaun, A. (1999). "Acoustic modeling of retroflex Tamil liquids," Proceedings of XIV ICPhS, San Francisco, CA (in press).
- Ostry, D. J., and Munhall, K. (1985). "Control of rate and duration of speech movements," J. Acoust. Soc. Am. **77**, 640–648.
- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabeta, I., and Jackson, M. (1992). "Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.
- Ramasubramanian, N., and Thosar, R. (1971). "Synthesis by rule of some retroflex speech sounds," *Language and Speech* **14**, 65–85.
- Rubin, P. E. (1995). "HADES: A case study of the development of a signal analysis system," in *Applied Speech Technology* (CRC Press, Boca Raton, FL), pp. 501–520.
- Saltzman, E., Löfqvist, A., and Mitra, S. (in press). "'Clocks' and 'glue'—Global timing and intergestural cohesion," *Papers in Laboratory Phonology V*.
- Scheffé, H. (1953). "A method for judging all contrasts in the analysis of variance," *Biometrika* **40**, 87–104.
- Sproat, R., and Fujimura, O. (1993). "Allophonic variation in English /l/ and its implications for phonetic implementation," J. Phonetics **21**, 291–311.
- Stevens, K. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
- Stone, M., Faber, A., Cordaro, M. (1991). "Cross-sectional tongue movement and tongue-palate movement in [s] and [ʃ] syllables," Proceedings of the XII ICPhS, pp. 354–357.
- Stone, M., Faber, A., Raphael, L., and Shawker, T. (1992). "Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [ʃ] and [l] syllables," J. Phonetics **20**, 253–270.
- Westbury, J. R., Hashi, M., and Lindstrom, M. J. (1998). "Differences among speakers in lingual articulations for American English /'trnr'/," *Speech Commun.* **26**, 203–226.

# Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results

Roger W. Chan<sup>a)</sup> and Ingo R. Titze

National Center for Voice and Speech, Department of Speech Pathology and Audiology,  
The University of Iowa, Iowa City, Iowa 52242

(Received 6 November 1998; revised 5 March 1999; accepted 29 June 1999)

A standard method for the empirical rheological characterization of viscoelastic materials was adopted to measure the viscoelastic shear properties of human vocal-fold mucosal tissues (the superficial layer of lamina propria). A parallel-plate rotational rheometer was employed to measure shear deformation of viscoelastic tissue samples, which were deformed between two rigid circular plates rotating in small-amplitude sinusoidal oscillations. Elastic and viscous shear moduli of the samples were then quantified as a function of oscillation frequency (0.01 to 15 Hz) based on shear stresses and strains recorded by the rheometer. Data were obtained from 15 excised human larynges (10 male and 5 female). Results showed that the elastic shear modulus  $\mu$  and the damping ratio  $\zeta$  of human vocal-fold mucosa were relatively constant across the range of frequencies observed, while the dynamic viscosity  $\eta$  decreased monotonically with frequency (i.e., shear thinning). Intersubject differences in  $\mu$  and  $\eta$  as large as an order of magnitude were observed, part of which may reflect age-related and gender-related differences. Some molecular interpretations of the findings are discussed. © 1999 Acoustical Society of America. [S0001-4966(99)03710-8]

PACS numbers: 43.70.Aj, 43.35.Mr, 43.75.Rs [AL]

## INTRODUCTION

Mechanical properties of vocal-fold tissues are important in the study of the acoustics and biomechanics of voice production. This has been demonstrated by computer simulation of speech and voice, where knowledge of tissue mechanical properties is required to predict tissue forces and deformation through a set of constitutive equations. For example, Titze (1976) and Titze and Talkin (1979) assumed transverse isotropy in a linear, elastic continuum to model vocal-fold tissues, which in principle requires five independent tissue mechanical constants for the constitutive equations. These elastic constants include the longitudinal Young's modulus  $E'$  (an indication of tensile elasticity or stiffness along the direction of tissue fibers), the transverse and the longitudinal Poisson's ratios ( $\nu$  and  $\nu'$ , respectively, as indications of tissue compressibility or the amount of tissue volume change upon deformation), and the transverse and the longitudinal shear moduli ( $\mu$  and  $\mu'$ , respectively, as indications of tangential elasticity or stiffness). Berry and Titze (1996) later showed that the two shear moduli were sufficient if tissue is assumed to be incompressible, such that the Poisson's ratios are known *a priori*, and if there is no vibration in the direction of tissue fibers, such that the equations of motion do not involve  $E'$ .

Among the different layers of vocal-fold tissues, the vocal ligament and the thyroarytenoid (TA) muscle may be assumed to be transversely isotropic, where tissue mechanical properties are independent of the orientation of deformation in the plane transverse to the essentially parallel tissue fibers, but different in the (longitudinal) direction of the fibers. On the other hand, the superficial layer of lamina pro-

pria (mucosa) can be assumed to be isotropic, where tissue properties are independent of the orientation of deformation in all planes, because of the basically random orientation of macromolecules in the lamina propria extracellular matrix (Gray *et al.*, 1999; Hammond *et al.*, 1997; Hirano, 1981; Pawlak *et al.*, 1996). Hence, only *one* shear modulus is needed for the superficial layer, as the transverse and the longitudinal shear moduli become identical ( $\mu = \mu'$ ). A fair amount of empirical data on tissue fiber tension has been gathered for the vocal ligament and the TA muscle (e.g., Alipour-Haghighi and Titze, 1991; Min *et al.*, 1995). This tension can be related to  $\mu'$ . However, little is known about  $\mu$ , neither for the isotropic mucosa nor the transversely isotropic ligament and muscle. The present study is directed toward this deficiency in the knowledge of vocal-fold tissue properties. In fact, the shear properties of other articulators necessary for solving the mechanics of other speech production models are also lacking (e.g., finite element models of the tongue and the velum, Berry *et al.*, 1999; Wilhelms-Tricarico, 1995).

Besides computer modeling, mechanical properties of vocal-fold tissues are also required to establish the threshold conditions for vocal-fold oscillation, which can be quantified by the *phonation threshold pressure*. Phonation threshold pressure has been defined as the minimum subglottal pressure required to initiate and sustain vocal-fold oscillation (Titze, 1988, 1992). It is an objective indication of vocal function, in particular the ease of phonation and the amount of vocal effort required. Phonation threshold pressure is associated with phonation onset or offset where energy supplied to the vocal folds by the glottal airflow is just large enough to overcome energy loss in the vocal folds due to tissue damping, such that the oscillation can be self-sustained. Based on a theoretical analysis of linear small-

<sup>a)</sup>Electronic mail: roger-chan@uiowa.edu

amplitude oscillation of the vocal folds, Titze (1988) showed that phonation threshold pressure is directly related to the viscosity of the vibratory portion of the vocal fold, i.e., the vocal-fold mucosa in small-amplitude oscillation. The validity of this analytical relationship has been verified by different kinds of experiments, using excised larynges (Finkelhor *et al.*, 1987), a physical model of the vocal-fold mucosa (Chan *et al.*, 1997; Titze *et al.*, 1995), as well as human subjects (Verdolini *et al.*, 1990, 1994).

Knowledge of the mechanical properties of vocal-fold tissues is also important clinically in the surgical management of vocal-fold disorders (phonosurgery) (Chan and Titze, 1998, 1999a; Gray *et al.*, 1999). The effect of tissue properties on vocal-fold oscillation should be considered in phonosurgery, where various types of synthetic or biological substitute materials have been used as surgical implants to replace or augment natural tissues in pathological vocal folds. In the management of vocal-fold mucosal defects, e.g., scarring and atrophy, biomaterials such as collagen and fat are often implanted directly into the vocal-fold mucosa as a filler substance, so as to repair focal defects, to smooth the vocal-fold margin, and to soften any stiffened scar tissue (Benninger *et al.*, 1996; Ford *et al.*, 1992; Sataloff *et al.*, 1997; Shaw *et al.*, 1997). It is obvious that the introduction of these implantable biomaterials into the vocal fold can change its mechanical properties and alter the mechanics of vocal-fold oscillation, which could make phonation easier or more difficult. This is particularly true when the vocal-fold mucosa is directly involved in repair, because the mucosa is the major vibratory portion of the vocal fold, especially in small-amplitude oscillations like phonation onset and offset (Fukuda *et al.*, 1991; Hirano, 1975; Saito *et al.*, 1985). *Shear* properties of vocal-fold tissues and implantable biomaterials are the most relevant in this context, because oscillation of the mucosa involves the propagation of a surface mucosal wave, which is a shear wave (Baer, 1975; Hirano, 1975). Therefore, shear properties become important factors in the choice of optimal biomaterials for mucosal repair (Chan and Titze, 1998, 1999a). In other situations, like the management of unilateral vocal-fold paralysis, where the main objective of surgery is to restore glottal competence by medializing and augmenting the paralyzed vocal fold, the properties of biomaterials like Teflon and gelatin are less important because they are usually implanted lateral to the vibratory portion of the vocal fold (Arnold, 1963; Schramm *et al.*, 1978).

Given the lack of information on the shear properties of vocal-fold tissues in the literature, this paper attempts to provide some empirical data on the shear properties of human vocal-fold mucosal tissues. A standard method commonly used in rheology (see the definition below) was adopted to characterize the shear properties of human vocal-fold mucosa, namely *sinusoidal oscillatory shear deformation*. Details of this method for the empirical measurement of shear properties *in vitro*, as well as the underlying linear viscoelastic theory are described below. Empirical data obtained with oscillatory shear experiments are interpreted in terms of some molecular findings on vocal-fold tissues in the literature. Theoretical modeling of viscoelastic shear properties, helpful for data extrapolation and parametric descriptions of

vocal-fold tissues, is addressed in a separate paper (Chan and Titze, 1999b). Clinical applications of our findings in phonosurgery are also discussed elsewhere (Chan and Titze, 1999a).

## I. MEASUREMENT OF VISCOELASTIC SHEAR PROPERTIES

In mechanics, ideal solids are described as ‘‘Hookean elastic’’ and ideal fluids as ‘‘Newtonian viscous.’’ Many materials, however, are *viscoelastic*, exhibiting both elastic and viscous properties. The behaviors of viscoelastic materials depend on the rate of deformation, in addition to thermodynamic (temperature and pressure) influences (Ferry, 1980; Fung, 1993). *Rheology*, defined by the Society of Rheology as the scientific study of the deformation and flow of matter, is related to the study of viscoelasticity, particularly viscoelastic shear properties. Standard rheological techniques can be employed to empirically determine the shear properties of viscoelastic materials (Barnes *et al.*, 1989; Bird *et al.*, 1977; Whorlow, 1980). Two such standard methods are discussed below: *steady shear* deformation and *oscillatory shear* deformation.

### A. Steady-shear deformation

Under steady-shear (i.e., continuous, uniform and nonoscillatory shear) deformation, the viscosity (steady-shear viscosity) of a purely viscous (Newtonian) fluid is defined as follows:

$$\tau = \eta_s \dot{\gamma}, \quad (1)$$

where  $\tau$  is shear stress (in Pa),  $\eta_s$  is steady-shear viscosity (in Pascal-second, Pa-s),  $\gamma$  is shear strain, and  $\dot{\gamma}$  is strain rate (in  $s^{-1}$ ). The ‘‘dot over’’ notation represents a time derivative. The SI unit Pascal-second is ten times larger than the somewhat more commonly used unit Poise (P). Water has a steady-shear viscosity of 0.001 Pa-s, or 0.01 P. For a Newtonian fluid like water, the steady-shear viscosity is a constant independent of strain rate (but dependent on temperature). For non-Newtonian fluids (including most polymeric materials), however, viscosity is a function of strain rate, and there can also be an elastic component not described by Eq. (1). Hence, steady-shear deformation alone is not the condition sufficient to characterize the rheological or shear properties of viscoelastic materials (Bird *et al.*, 1977; Ferry, 1980).

Biological tissues and materials are all viscoelastic (Fung, 1993). Their elastic and viscous shear properties are revealed when they are subject to sinusoidal oscillatory shear deformation, which is summarized next.

### B. Oscillatory shear deformation

Viscoelastic shear properties are quantified by the *complex shear modulus* (or dynamic shear modulus), which includes an elastic component, the *elastic shear modulus*, and a viscous component, the *dynamic viscosity*. These components are additive for viscoelastic materials, according to the theory of *linear* viscoelasticity (Barnes *et al.*, 1989; Ferry, 1980; Fung, 1993; Whorlow, 1980). Therefore, for a vis-



coelastic material, the general linear constitutive equation relating shear stress with shear strain and strain rate is

$$\tau = \mu \gamma + \eta \dot{\gamma}, \quad (2)$$

where  $\mu$  is elastic shear modulus or storage shear modulus (in Pa) and  $\eta$  is dynamic viscosity (in Pa-s) of the viscoelastic material. Consider the application of a sinusoidal oscillatory shear strain (with a given amplitude and frequency) to a sample of the material. According to the constitutive equation, it can be shown that steady-state conditions of oscillation will be reached for *small-amplitude* oscillations, for which the sinusoidal shear strain (input) will result in a sinusoidal shear stress (output) of the material at the *same* frequency of oscillation. The resulting shear stress will have an amplitude proportional to the strain amplitude, and a phase lag independent of stress or strain amplitude (Ferry, 1980; Fung, 1993; Whorlow, 1980). The same linear theory also applies when a sinusoidal shear *stress* is being applied, resulting in a sinusoidal shear strain. Using complex number notations, the shear strain and shear stress can be represented as

$$\gamma^* = \gamma_0 e^{i\omega t}, \quad (3)$$

$$\tau^* = \tau_0 e^{i(\omega t + \delta)}, \quad (4)$$

where  $\gamma^*$  is complex shear strain (the asterisk superscript represents complex number notation),  $\gamma_0$  is shear strain amplitude,  $\tau^*$  is complex shear stress,  $\tau_0$  is shear stress amplitude,  $i$  is the imaginary number  $\sqrt{-1}$ ,  $\omega$  is angular frequency of oscillation,  $t$  is time, and  $\delta$  is the phase shift (lead) of shear stress relative to shear strain (Fig. 1). The ratio of the shear amplitudes ( $\tau_0/\gamma_0$ ) and the phase shift ( $\delta$ ) are related to the elastic shear modulus  $\mu$  and the dynamic viscosity  $\eta$ , which are functions of the frequency of oscillation, and thermodynamic variables like temperature and pressure.

By time differentiation of the complex shear strain, the complex strain rate is

$$\dot{\gamma}^* = i\omega \gamma_0 e^{i\omega t}. \quad (5)$$

According to the constitutive equation [Eq. (2)], application of the sinusoidal shear strain [Eq. (3)] on the viscoelastic material sample yields a shear stress of

$$\tau^* = \mu \gamma_0 e^{i\omega t} + i\omega \eta \gamma_0 e^{i\omega t} = (\mu + i\omega \eta) \gamma_0 e^{i\omega t} = \mu^* \gamma^*, \quad (6)$$

where  $\mu^*$  is the complex shear modulus (also called the dynamic shear modulus or the modulus of rigidity; Fung, 1993)

$$\mu^* = \mu + i\omega \eta. \quad (7)$$

It is clear from Eq. (6) that the complex shear modulus can also be defined as the ratio of complex shear stress to complex shear strain. Using a polar coordinate representation,  $\mu^*$  can be expressed in terms of its *magnitude*  $|\mu^*|$  and phase angle  $\delta$ , which is just the phase shift between stress and strain (Ferry, 1980; Fung, 1993)

$$\mu^* = |\mu^*| e^{i\delta}, \quad (8)$$

where

$$|\mu^*| = \sqrt{\mu^2 + (\omega \eta)^2}, \quad (9)$$

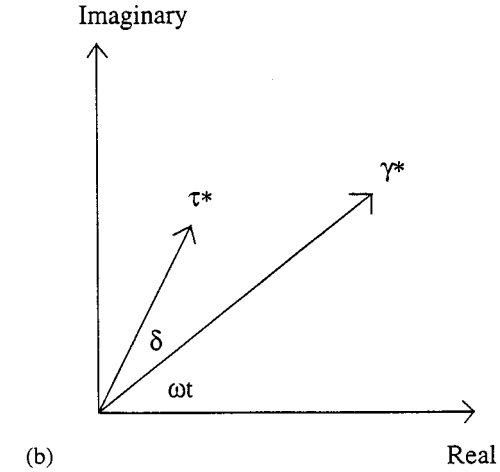
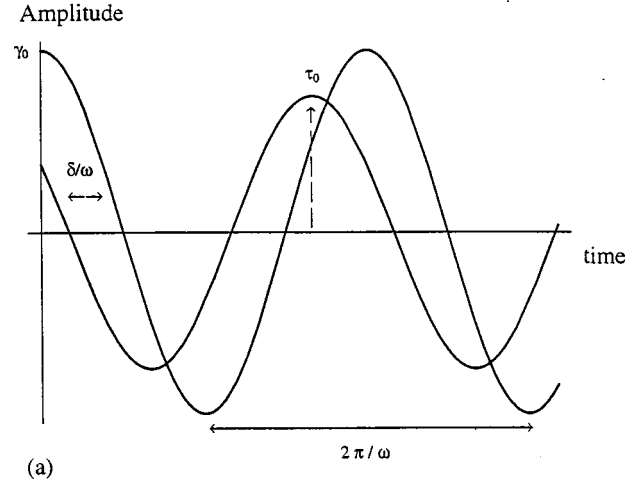


FIG. 1. (a) Sinusoidal representation and (b) Phasor representation of shear stress ( $\tau^*$ ) and shear strain ( $\gamma^*$ ).

and

$$\delta = \tan^{-1} \frac{\omega \eta}{\mu}. \quad (10)$$

Expressing the complex shear stress in terms of  $|\mu^*|$  and  $\delta$  according to Eqs. (6) and (8),

$$\tau^* = |\mu^*| \gamma_0 e^{i(\omega t + \delta)}. \quad (11)$$

Thus, the application of a sinusoidal shear strain on a viscoelastic material described by the linear constitutive equation does result in a sinusoidal shear stress at the same frequency of oscillation ( $\omega$ ). The resulting shear stress is expressed exactly in the form shown in Eq. (4), with

$$\tau_0 = |\mu^*| \gamma_0. \quad (12)$$

Note that the shear stress amplitude  $\tau_0$  is proportional to the shear strain amplitude  $\gamma_0$ , with the proportionality constant being the magnitude of the complex shear modulus  $|\mu^*|$ . In other words, the magnitude of the complex shear modulus turns out to be the ratio of the shear amplitudes ( $\tau_0/\gamma_0$ ).

The elastic shear modulus  $\mu$  and the dynamic viscosity  $\eta$  can also be expressed in terms of the shear amplitudes and the phase shift which are measurable by oscillatory shear

deformation. According to Eq. (12) and Euler's relation, the polar form of the complex shear modulus [Eq. (8)] can be expanded as

$$\mu^* = \frac{\tau_0 \cos \delta}{\gamma_0} + i \frac{\tau_0 \sin \delta}{\gamma_0}, \quad (13)$$

which yields the following expressions for its real part and imaginary part [cf. Eq. (7)]:

$$\mu = \frac{\tau_0 \cos \delta}{\gamma_0}, \quad (14)$$

$$\eta = \frac{\tau_0 \sin \delta}{\omega \gamma_0}. \quad (15)$$

The elastic shear modulus  $\mu$  is proportional to the elastically stored strain energy (or internal energy) in the viscoelastic material over one cycle of oscillation. It quantifies the elasticity or stiffness of the material in shear. The dynamic viscosity  $\eta$ , on the other hand, is proportional to the energy dissipated or lost in the viscoelastic material, typically as heat. It characterizes the mechanical opposition to (shear) flow in the material and is often a monotonically decreasing function of frequency for viscoelastic and polymeric materials.

The symbols for viscoelastic shear properties introduced above have been commonly used in the bioengineering literature, as well as in the acoustics (speech production) literature (e.g., Alipour-Haghighi and Titze, 1988; Berry and Titze, 1996; Titze and Talkin, 1979). However, a different nomenclature has been more widely used in the rheology literature, where the symbol  $G$  is used to represent the shear modulus (Barnes *et al.*, 1989; Bird *et al.*, 1977; Ferry, 1980; Whorlow, 1980):

$$G^* = G' + iG'', \quad (16)$$

where  $G^* = \mu^*$ ,  $G' = \mu$  and  $G'' = \omega \eta$ .  $G''$  is called the *viscous shear modulus* or loss shear modulus, defined as the product of angular frequency and dynamic viscosity. The dynamic viscosity  $\eta$  is commonly represented by  $\eta'$  because it can be defined as the real part of the complex dynamic viscosity  $\eta^*$ . Using the symbol  $G$  in polar form,

$$G^* = |G^*| e^{i\delta}. \quad (17)$$

The magnitude of the complex shear modulus has been used as an indication of the overall shear elasticity, stiffness, or rigidity of a viscoelastic material, while the phase shift is an indication of the amount of energy loss relative to the amount of energy stored in the material.

The *damping ratio*  $\zeta$  (also called *loss tangent* or *loss factor*) can be defined as the tangent of the phase shift  $\delta$ , which is simply the ratio of the viscous shear modulus to the elastic shear modulus ( $G''/G'$ ) [Eq. (10)],

$$\zeta = \tan \delta = \frac{\omega \eta}{\mu}. \quad (18)$$

A material is purely elastic if  $\zeta = 0$  (when  $\eta = 0$ ), purely viscous if  $\zeta = \infty$  (when  $\mu = 0$ ), and viscoelastic if  $\zeta$  lies somewhere in between (when  $0^\circ < \delta < 90^\circ$ , i.e., when there are both viscous and elastic components). In the context of

oscillation, a system is underdamped if  $\zeta < 1.0$ , critically damped if  $\zeta = 1.0$ , and overdamped if  $\zeta > 1.0$ . The typical values of damping ratio for human vocal-fold tissues have been estimated to be on the order of 0.1–0.2 at 30–40 Hz (Kaneko *et al.*, 1972) or 0.2–0.4 at 130 Hz (Isshiki, 1977). Vocal folds of excised human larynges were perturbed with mechanical impulses and their damped oscillations were measured in those studies. Accordingly, in the two-mass model of vocal-fold oscillation, Ishizaka and Flanagan (1972) used a damping ratio of 0.1 for the lower mass and 0.6 for the upper mass (an equivalent value of 0.16 for the bulk of the vocal fold) to simulate self-sustained oscillation at a reasonable range of model parameters (e.g., subglottal pressure, vocal-fold length, mass, and spring constants). However, these data on damping ratio were based on damped oscillations of whole vocal folds, not individual tissue layers. No data are available on the damping ratios of different vocal-fold tissue layers, particularly the vocal-fold mucosa which is the major vibratory portion in small-amplitude oscillations. Also, it is known that damping ratio is a function of frequency (albeit a weak function for most biological soft tissues; cf. Fung, 1993), yet data are only available at a few frequencies. Therefore, the present study attempts to provide a more comprehensive set of data on the damping ratio of vocal-fold mucosal tissues, as well as their elastic shear modulus and dynamic viscosity.

## II. METHOD

### A. Vocal-fold tissue samples

A rheometer was used to measure the complex shear modulus of human vocal-fold mucosal tissues. Tissue samples were obtained from two sources, (1) fresh excised larynges from the University of Iowa Hospitals and Clinics (UIHC) autopsy unit within 24 h postmortem; and (2) quick-frozen excised larynges shipped from the University of Utah (courtesy of Dr. Steven Gray). Most of the larynges from UIHC autopsy were obtained within 18 to 20 h postmortem. All of the larynges from Utah were quick-frozen with liquid nitrogen within 18 h postmortem and stored at a temperature below  $-20^\circ\text{C}$ . Larynges were harvested from ten male and five female cadaver subjects. None of them had a history of laryngeal pathology, laryngeal intubation, or smoking. Table I summarizes the subject information.

Since it was recognized that postmortem changes, freezing, and/or thawing might have introduced significant differences in tissue mechanical properties, an experiment was done to compare the shear properties of fresh (viable) canine vocal-fold tissues with those of frozen tissues (Chan, 1998). Canine tissues were used because they could be obtained immediately postmortem from cardiovascular laboratories in the UIHC, while it was not feasible to obtain fresh enough human larynges from UIHC autopsy. Results showed that up to 24 h of postmortem changes at room temperature and up to 1 month of freezing storage following quick freezing of mucosal tissues (using liquid nitrogen) did not affect tissue shear properties significantly (Chan, 1998).

For each excised larynx, the vocal-fold epithelium and the superficial layer of lamina propria were carefully dis-

TABLE I. Characteristics of human cadaver subjects.

Subject #	Gender	Race	Age	Source
1	M	Polynesian	28	Utah
2	M	white	30	Utah
3	M	white	31	Utah
4	M	n.a. <sup>a</sup>	34	UIHC autopsy
5	M	white	36	Utah
6	M	white	50	UIHC autopsy
7	M	white	59	UIHC autopsy
8	M	white	60	UIHC autopsy
9	M	n.a.	62	UIHC autopsy
10	M	n.a.	72	UIHC autopsy
11	F	white/Hispanic	25	UIHC autopsy
12	F	n.a.	55	UIHC autopsy
13	F	n.a.	55	UIHC autopsy
14	F	white	71	UIHC autopsy
15	F	n.a.	86	UIHC autopsy

<sup>a</sup>n.a.=race information not available.

sected. A small incision on the superior surface of the vocal fold epithelium, just above the boundary between the superficial layer of lamina propria and the vocal ligament, was made by a surgical blade (No. 15). A fine tissue grasping forcep was used to slightly lift up the mucosa. The superficial layer of the lamina propria was separated from the intermediate layer (the vocal ligament) by a Bouchayer spatula for blunt dissection. The dissection process was similar to the lateral microflap technique in phonosurgery (Courey and Ossoff, 1995). Under this procedure, the epithelium always remained attached to the superficial layer during dissection, as well as during the rheometric experiments. Its purpose was to serve as a natural boundary of attachment between the rheometer and the superficial layer of lamina propria.

Except for some frozen hemilarynges from Utah, two samples of mucosal tissues were obtained from each larynx, one from the left and the other from the right vocal fold. Symmetry in shear properties between two vocal folds of the same larynx was assessed in a few subjects. The volume of each tissue sample was approximately 0.1 cc, and the gap size between the two plates of the rheometer was adjusted to fit this sample volume (see below).

## B. Instrumentation and rheometric measurements

### 1. The rheometer

A computer-controlled rheometer (Bohlin CS-50, Bohlin Instruments, Inc., Cranbury, NJ) was used for data acquisition and analysis. A parallel-plate (plate-on-plate) testing geometry was used, consisting of a stationary lower plate and a rotating upper plate (diameter=20 mm) (Fig. 2). With this geometry, the sample volume required to fill the gap between the two plates could be as small as 0.1 cc when a gap size of 0.3 mm was used. Since the sample volume of vocal-fold mucosa obtained from one single vocal fold was typically around 0.1 cc, the gap size of 0.3 mm was appropriate to ensure complete sample filling between the two plates (Fig. 2). Occasionally, when there were smaller samples, a smaller gap size was used until complete sample filling was achieved (0.2 or 0.25 mm). Larger gap sizes were not used as all of the samples tested fit a gap size of 0.3 mm or smaller.

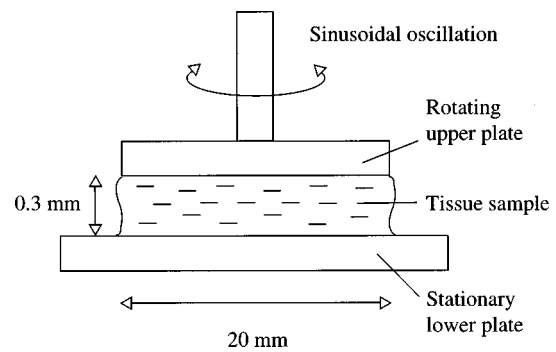


FIG. 2. Schematic of the experimental setup, showing a tissue sample between the two plates of the rheometer (cross-sectional view not to scale).

In the rheometer, the sample was subject to a precisely controlled sinusoidal torque from the upper plate, driven by a drag cup motor with a torque range of 0.001 to 10 mNm (milli Newton-meter) and a resolution of 0.0002 mNm. The resulting angular displacement and angular velocity of the upper plate were monitored by a sensitive transducer (with a fully linear angular deflection range, and a resolution of  $2 \times 10^{-5}$  rad) as functions of time. Sinusoidally varying shear stress, shear strain, and strain rate were calculated from the prescribed torque and the measured angular velocity by the computer, based on conversion constants of the testing system geometry in use. For the parallel-plate geometry with a diameter of 20 mm and a gap size of 0.3 mm, the shear stress–torque conversion constant was 477 465 Pa/Nm. The strain rate–angular velocity conversion constant was  $7.5 \text{ s}^{-1}/\text{rad s}^{-1}$ . The rotational moment of inertia of the upper plate about the long axis was  $2.2 \times 10^7 \text{ kgm}^2$ . From the shear stress and strain functions, the shear amplitudes ( $\tau_0$  and  $\gamma_0$ ) and phase shift ( $\delta$ ) were measured by the controlling software. From these, the elastic shear modulus, the dynamic viscosity, and the damping ratio were computed according to the analytical expressions discussed in the previous section. The software used a window of approximately 20 cycles of oscillation for the calculations.

### 2. Oscillatory shear experiments

Measurements of shear properties were made at a frequency range of 0.01 to 15 Hz, covering 32 frequencies over three decades. Testing at higher frequencies was not meaningful, because linearity of oscillations could not be ensured when rotor (upper plate) and sample inertial effects became significant at high frequency (see below). Both upward and downward frequency sweeps were performed for a few samples in order to check for possible differences between the two. No significant differences were observed between them (Chan, 1998). Hence, in order to be efficient, only upward frequency sweeps were performed for all measurements.

Calibration of the rheometer was done by the manufacturer and was double-checked by measuring the steady-shear viscosities of (1) standard Newtonian silicone (polydimethylsiloxane) solutions (Dow Corning 360 Medical Fluids, Dow Corning, Midland, MI) with known viscosities of 0.1, 0.35, 1, and 12.5 Pa-s at 20 °C; (2) standard fluid asphalt

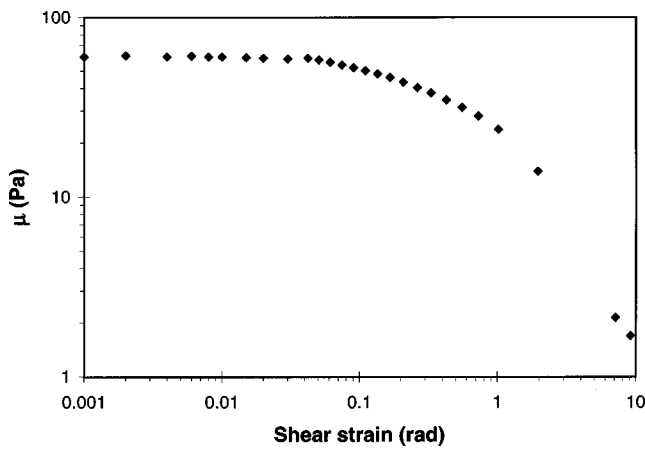


FIG. 3. Elastic shear modulus  $\mu$  of human vocal-fold mucosa as a function of shear strain amplitude  $\gamma_0$  (60-year-old male); frequency=1 Hz.

(also Newtonian) supplied by the manufacturer (Bohlin Instruments, Inc.) with a known viscosity of 35 Pa-s at 20 °C. The measured viscosities showed only a small deviation (<6%) from the stated viscosities in all cases.

Throughout the experiments, samples in the rheometer were maintained at a temperature of  $37\text{ }^\circ\text{C} \pm 0.1\text{ }^\circ\text{C}$  by a Bohlin Temperature Control Unit which circulated distilled water into the mounting of the lower plate. Hydration levels of vocal-fold tissue samples were also kept relatively constant by bathing them in physiological (0.9%) saline solution during the experiments.

### 3. Linearity of oscillation

In order to ensure stress-strain linearity, small-amplitude oscillation was maintained by setting the shear strain amplitude ( $\gamma_0$ ) at a value small enough for the linear constitutive equation [Eq. (2)] to hold. Recall from Sec. I that linearity is characterized by a proportionality between the shear amplitudes, and a phase shift being independent of the shear amplitudes (Ferry, 1980; Fung, 1993; Whorlow, 1980). That is, viscoelastic shear properties are independent of the strain amplitude in the linear region, when shear strain is small. The target shear strain amplitude was located by strain sweep tests done prior to frequency-sweep experiments, where the amplitude of shear strain was slowly varied and the viscoelastic responses ( $\mu$  and  $\eta$ ) were obtained as a function of strain amplitude at selected frequencies.

Figure 3 shows the typical results of a strain sweep test on a sample of human vocal-fold mucosa (from Subject 8). It can be seen that the elastic shear modulus  $\mu$  was fairly independent of the shear strain amplitude  $\gamma_0$  at small strain, up to an amplitude of around 0.01–0.05 rad. Therefore, a target strain amplitude of 0.01 rad was set for all small-amplitude frequency-sweep experiments. It was maintained by the controlling software through automatic adjustments of the shear stress amplitude at each oscillation frequency.

Another feature of the software that helped to ensure linearity was its automatic introduction of delay time periods up to 100 s prior to the calculations of shear stress and strain rate functions at each frequency of oscillation. This delay

period guaranteed that measurements were made from steady-state sinusoidal oscillations, rather than from onset transients.

### 4. Sample inertia

An assumption underlying the linear viscoelastic theory is that the inertia of the sample is negligible in oscillatory shear deformation. It requires that the sample thickness be small compared with the wavelength of shear waves propagated through the medium at the frequency of oscillation (Ferry, 1980). This is called the “gap loading” condition, as the driving upper plate would “feel” like loading an empty gap with no sample inertia. This condition is achieved when there are negligible inertial forces in the sample during oscillation. According to Ferry (1980), inertial forces are small when

$$h \ll \sqrt{\frac{2\pi\eta}{\rho f}}, \quad (19)$$

where  $h$  is sample thickness (in m),  $\rho$  is density of the sample (in  $\text{kg/m}^3$ ), and  $f$  is frequency of oscillation (in Hz). Thus, a thick sample (a large gap size), a low sample viscosity, or a high-frequency oscillation could all invalidate the assumption of negligible sample inertial effects. Assuming a sample dynamic viscosity of 0.5 Pa-s, which was approximately the smallest viscosity value observed for the average vocal-fold mucosa, a sample density of  $1100\text{ kg/m}^3$  (vocal-fold tissue density reported by Perlman, 1985), and an oscillation frequency of 15 Hz, the sample thickness ( $h$ ) should be much smaller than 13.8 mm in order for inertial effects to become negligible. For the 0.3-mm gap size of the rheometer, it was about 46 times smaller than 13.8 mm. In order not to significantly violate the assumption of negligible inertial effects, it was arbitrarily decided that the cutoff ratio would be 50 times. With this cutoff, meaningful measurements were obtainable at frequencies up to 10 Hz, with measurements at 15 Hz considered marginally acceptable.

### 5. Estimation of experimental error

Experimental error associated with the rheological experimental procedure was estimated by mounting and testing similar material samples for more than one trial. Figure 4 shows the elastic shear modulus of three similar samples of Teflon (polytetrafluoroethylene) as a function of oscillation frequency, plotted on a log-log scale. Because the Teflon samples came from the same source (the same tube ready for vocal-fold injection, from Mentor, Inc., Hingham, MA), discrepancies in results between the samples reflected an estimation of the repeatability or experimental error of the oscillatory shear experiments. Possible sources of errors included differences in the process of sample mounting (e.g., incomplete sample filling between the plates, excessive removal of excess sample material) and errors of the rheometer (e.g., inertia of the rotating plate and the sample, errors in rotor torque output, errors in measurement of angular velocity). According to the largest magnitude of differences observed in Fig. 4, experimental error associated with the mea-



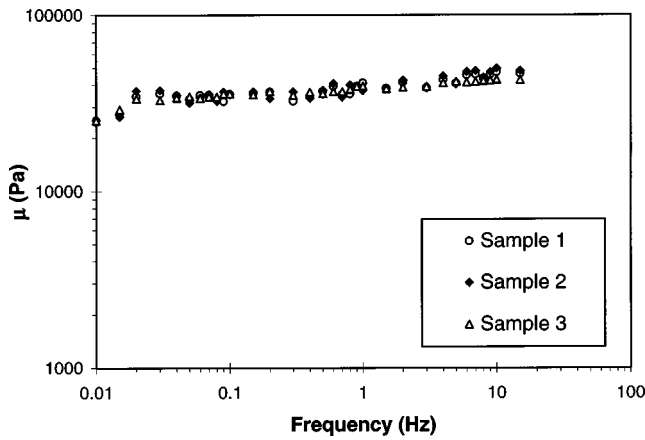


FIG. 4. Elastic shear modulus  $\mu$  of three similar Teflon samples as a function of frequency. Discrepancies between the samples represent an estimation of the magnitude of experimental error.

surement procedure was estimated to be within 10% of the measured viscoelastic data values across all frequencies.

Repeated trials of experiments were also performed on samples of human vocal-fold mucosa. Any differences in shear properties between the trials would partly reflect the above experimental error, and partly reflect the effect of the rheological testing procedure *itself* on shear properties of tissue samples. Following exactly the same experimental procedure as described above, two trials of measurements were made on each of several samples of human vocal-fold mucosa, separated by a duration of 30 min during which the tissue samples were kept in saline solution at room temperature. Figures 5 and 6 show the effect of this repeated testing on the elastic shear modulus and the dynamic viscosity, respectively, of a mucosal tissue sample from Subject 15. It is clear that the shear properties of the sample remained basically the same in the two trials, with differences between the trials falling within or close to the range of experimental error (10%) at all frequencies. Specifically, the magnitude of differences between test and retest ranged between 10%–15% at lower frequencies (<1 Hz), while at higher frequencies (>1 Hz) it was smaller than 10%. This finding was also consistently observed with other test samples. Hence, the

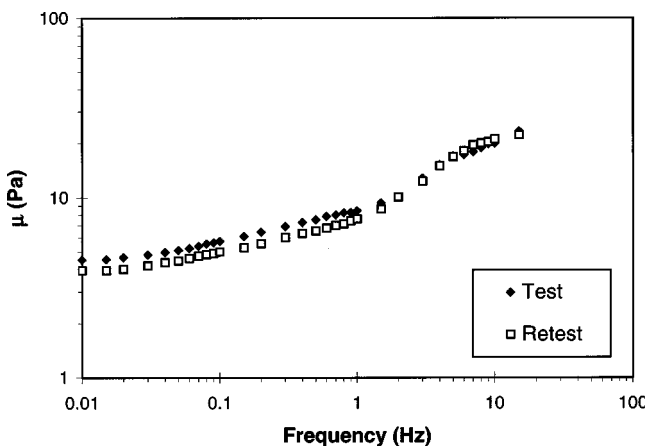


FIG. 5. Effect of rheological testing on elastic shear modulus  $\mu$  of human vocal-fold mucosa (86-year-old female).

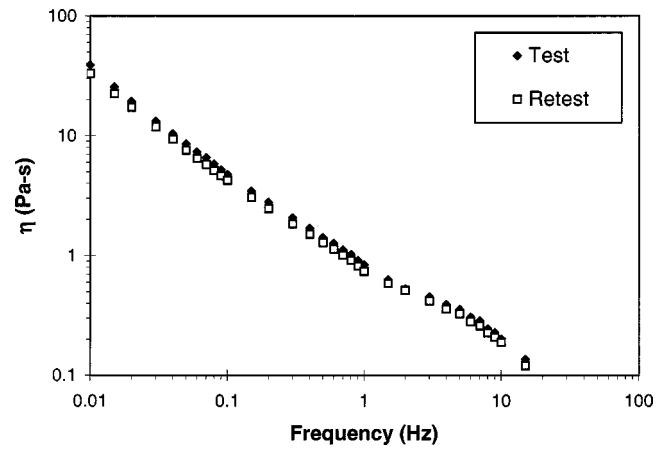


FIG. 6. Effect of rheological testing on dynamic viscosity  $\eta$  of human vocal-fold mucosa (86-year-old female).

experimental procedure itself did not significantly affect the shear properties of the tissue samples.

### III. RESULTS AND DISCUSSION

#### A. Symmetry in shear properties

Figures 7 and 8 show the elastic shear modulus  $\mu$  and the dynamic viscosity  $\eta$ , respectively, of human vocal-fold mucosal tissue samples obtained from the left and the right vocal folds of the same larynx (from Subject 14), as a function of oscillation frequency on a log–log scale. Logarithmic scales are used so that decades of changes in shear properties and in frequency could be clearly shown on the same graph. Results showed that the differences in  $\mu$  and  $\eta$  observed between the two vocal folds were on the order of the magnitude of experimental error (10%). Specifically, the magnitude of differences between the two vocal folds was smaller than 10% at lower (<0.2 Hz) and higher (>2 Hz) frequencies, while at other frequencies it ranged between 10%–15%. The same finding was also observed in other subjects, where measurements were made on both vocal folds. Hence, it was concluded that there were no indications of mechanical asymmetry in the shear properties between mucosal tissues of contralateral vocal folds. This finding was not surprising,

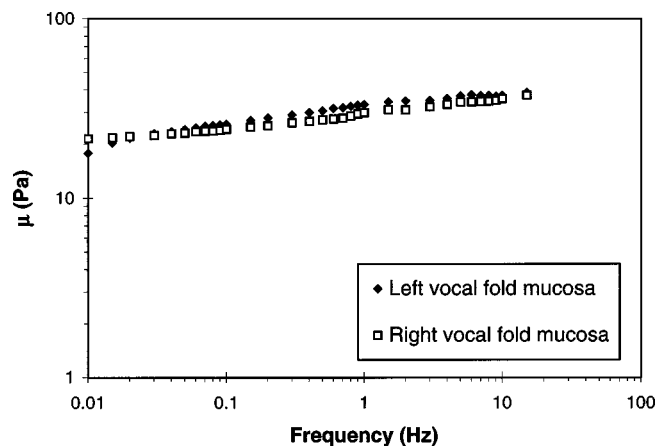


FIG. 7. Left–right symmetry in elastic shear modulus  $\mu$  of human vocal-fold mucosa (71-year-old female).

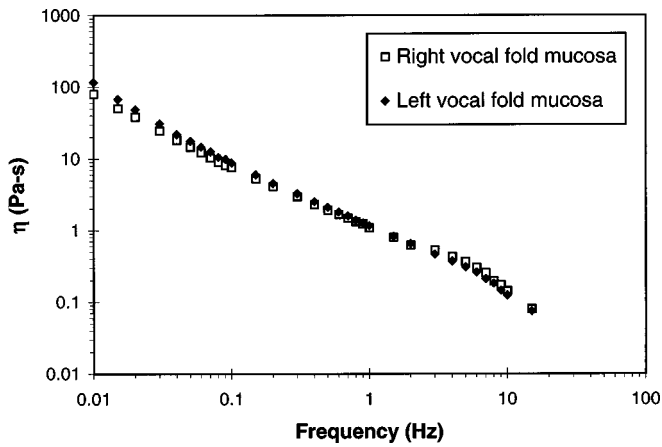


FIG. 8. Left–right symmetry in dynamic viscosity  $\eta$  of human vocal-fold mucosa (71-year-old female).

given that symmetric or near-symmetric vocal-fold vibration is commonly observed in normal human subjects, and that not much asymmetry in vocal-fold histology and molecular constituents has been reported in the literature.

### B. Shear properties of human vocal-fold mucosa

Figures 9 to 14 show the viscoelastic shear properties of human vocal-fold mucosal tissues from all subjects, including the elastic shear modulus  $\mu$  (Figs. 9 and 10), the dynamic viscosity  $\eta$  (Figs. 11 and 12), and the damping ratio  $\zeta$  (Figs. 13 and 14). Data for male and female subjects are shown in separate figures, and they are again plotted as a function of oscillation frequency on a log–log scale (the standard in rheology).

The elastic shear modulus  $\mu$  was a relatively flat function of frequency and showed a very similar pattern of slope changes across the frequency range observed (Figs. 9 and 10). For male subjects, it remained basically constant at low frequency but increased slowly with frequency at higher frequencies ( $>1$  Hz) (Fig. 9). An exception was the 72-year-old male, whose  $\mu$  was increasing with frequency across the entire frequency range and thus had a slightly greater overall slope. Another exception was the 30-year-old male, who did not show the slow increase in  $\mu$  at frequencies above 1 Hz.

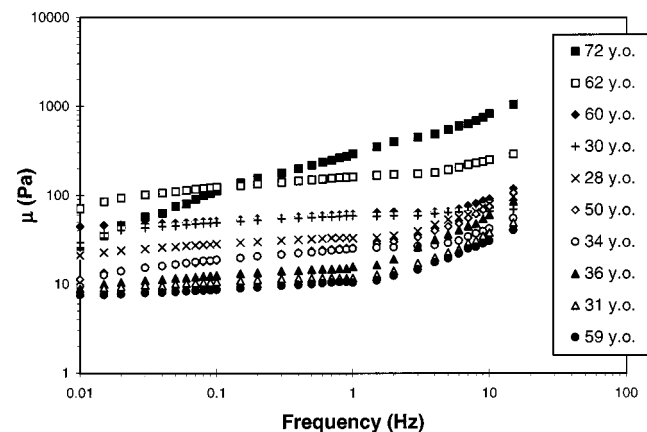


FIG. 9. Elastic shear modulus  $\mu$  of human vocal-fold mucosa as a function of frequency (male subjects).

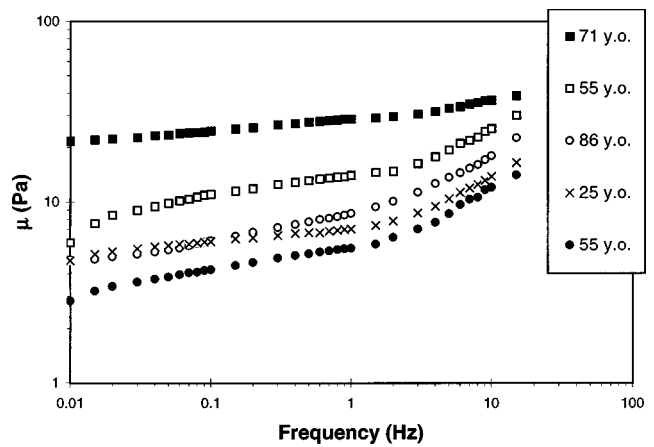


FIG. 10. Elastic shear modulus  $\mu$  of human vocal-fold mucosa as a function of frequency (female subjects).

For females,  $\mu$  was relatively flat at low frequency ( $<1$  Hz) (although it was less flat than the male data), but again it was increasing slowly with frequency at higher frequencies (Fig. 10). Despite these similarities across different subjects, there were huge intersubject differences in the magnitude or numerical value of the elastic shear modulus, where differences were sometimes as large as one or two orders of magnitude. For instance, the male data ranged from approximately 10 to 100 Pa at low frequency and from approximately 40 to 1000 Pa at relatively high frequency (15 Hz) (Fig. 9), while for female subjects they ranged from approximately 3 to 20 Pa at low frequency and from approximately 10 to 40 Pa at high frequency (15 Hz) (Fig. 10). A closer look at these data showed that *in general* the male vocal-fold mucosa was more elastic (stiffer) than the female vocal-fold mucosa, and that older subjects seemed to have a more elastic vocal-fold mucosa than younger subjects (see the next section for further discussion).

The dynamic viscosity  $\eta$  was a monotonically decreasing function of frequency for all subjects (Figs. 11 and 12), a *shear-thinning* effect commonly observed in biological soft tissues and polymeric materials (Ferry, 1980; Fung, 1993). The decrease with frequency was approximately linear on the log–log scale, which indicated that the relationship between

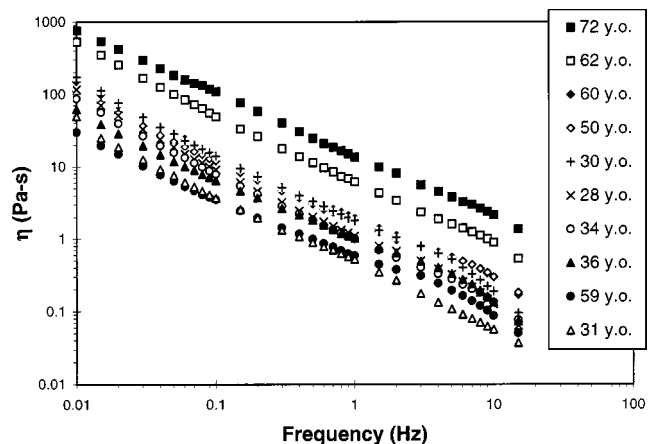


FIG. 11. Dynamic viscosity  $\eta$  of human vocal-fold mucosa as a function of frequency (male subjects).

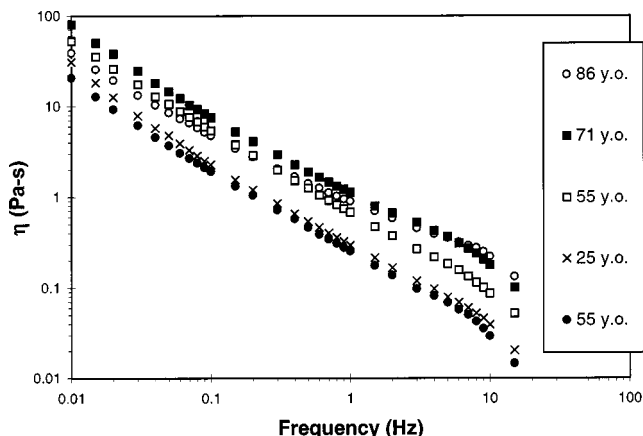


FIG. 12. Dynamic viscosity  $\eta$  of human vocal-fold mucosa as a function of frequency (female subjects).

dynamic viscosity and frequency could be modeled by a power law

$$\eta = kf^n, \quad (20)$$

or

$$\log \eta = \log k + n \log f, \quad (21)$$

where  $f$  is frequency in Hz,  $k$  and  $n$  are constants. Data were averaged across male and female subjects and then fitted to Eq. (21) by simple linear least-squares regression. Results are summarized in Table II, where values of the coefficient  $k$  and the exponential index  $n$  of the power law are given. The square of the correlation coefficient  $r$  (the coefficient of determination) is also given as an estimate of the goodness of fit. It can be seen that data are very well matched by the regression equation, yielding correlation coefficients above 0.99. Dynamic viscosity decreased with frequency at approximately the same rate for different subjects (with similar values of the slope  $n$ ), but there was considerable difference between the subjects in vertical separation between the curves (Figs. 11 and 12). Similar to  $\mu$ , huge intersubject differences in  $\eta$  as large as an order of magnitude were again evident for most subjects, with older and/or male subjects generally showing a more viscous vocal-fold mucosa than younger and/or female subjects.

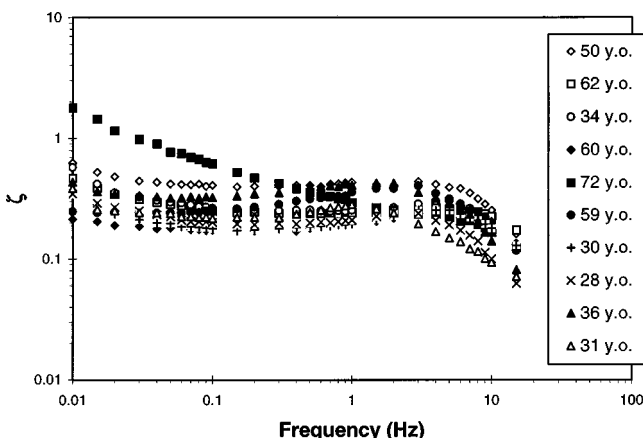


FIG. 13. Damping ratio  $\zeta$  of human vocal-fold mucosa as a function of frequency (male subjects).

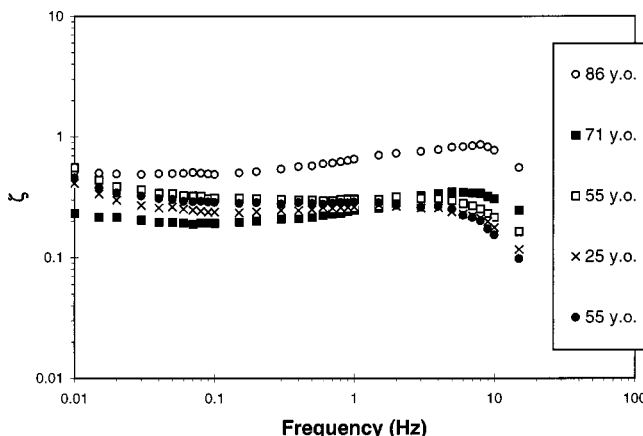


FIG. 14. Damping ratio  $\zeta$  of human vocal-fold mucosa as a function of frequency (female subjects).

The damping ratio  $\zeta$  was basically a relatively flat function of frequency for most subjects (Figs. 13 and 14), which has been commonly observed in other biological soft tissues as well (Fung, 1993). Most subjects showed a consistent pattern of changes in damping ratio across the frequency range, with  $\zeta$  decreasing very slightly with frequency at low frequencies ( $<0.1$  Hz), remaining constant or increasing very slightly with frequency at intermediate frequencies (0.1 to 3–5 Hz), and decreasing more rapidly at relatively high frequency ( $>3$ –5 Hz). Two exceptions to this pattern were the 72-year-old male, whose  $\zeta$  was basically a monotonically decreasing function with a slightly more negative overall slope (Fig. 13), and the 86-year-old female, whose  $\zeta$  was slightly increasing with frequency across most of the frequency range (Fig. 14). Regarding intersubject differences, it was interesting to note that there was an apparent “convergence” of the data to a relatively small range of  $\zeta$  for different subjects, despite huge (orders of magnitude) differences observed in the other shear properties ( $\mu$  and  $\eta$ ). Comparing to the data of  $\mu$  and  $\eta$ ,  $\zeta$  was within a very small range of values for most subjects, approximately 0.2–0.5 at low frequency and 0.1–0.3 at higher frequencies (10–15 Hz) (except the 72-year-old male and the 86-year-old female). These damping ratio values were close to those reported previously by Kaneko *et al.* (1972) and Isshiki (1977), whose measurements were based on whole vocal folds excited quite differently (by mechanical impulses) at quite different frequencies (30–40 Hz and 130 Hz). These results suggested that vocal-fold oscillation would remain underdamped ( $\zeta < 1.0$ ) for most subjects<sup>1</sup> despite the fact that their vocal-fold mucosal tissues were often very different in elasticity and viscosity.

### C. The effects of age and gender

Some of the large intersubject differences in the elastic shear modulus and the dynamic viscosity of human vocal-

TABLE II. Results of linear least-squares regression curve-fitting for the dynamic viscosity ( $\eta$ ) of human vocal-fold mucosa according to Eq. (21).

Tissue	$k$ (Pa-s <sup>2</sup> )	$n$	$r^2$
Male vocal-fold mucosa ( $N=10$ )	3.1440	-0.8755	0.999
Female vocal-fold mucosa ( $N=5$ )	0.7038	-0.8345	0.996

fold mucosa may be attributed to the effects of subject age and gender, although it is also possible that some of these differences may simply be normal intersubject variability, or may have been caused by the subjects' different states of health prior to death and other factors unknown to the authors (e.g., undocumented smoking or exposures to other laryngeal irritants, laryngeal tissue responses to medications). On any given figure, subjects at the upper end of the range often showed a shear property three or four times, or even an order of magnitude larger than those at the lower end of the range. For example, at higher frequencies (10–15 Hz) where extrapolations to frequencies of phonation might be possible, the 60-year-old male and the 50-year-old male had mucosal tissues approximately three times more elastic than those of the 31-year-old and the 59-year-old (Fig. 9), while  $\mu$  was an order of magnitude above for the 72-year-old and the 62-year-old. Comparably large intersubject differences were also observed for  $\eta$  and among female subjects.

Part of these intersubject differences may reflect an apparent age effect. For males, older subjects generally showed larger  $\mu$  and  $\eta$  than younger subjects at higher frequencies (10–15 Hz), especially when ages at or above 50 were compared with ages below 50 (Figs. 9 and 11). An obvious exception to this trend was the 59-year-old male, who often showed the smallest  $\mu$  and  $\eta$  among all male subjects. For females, there was also a rather clear age effect for  $\eta$  (except one of the 55-year-old female), but the trend was less clear for  $\mu$  (Figs. 10 and 12).

In summary, it appeared that the viscoelastic shear properties of human vocal-fold mucosa generally increased with age. Older subjects seemed to show a more elastic (stiffer) and more viscous vocal-fold mucosa, especially for male subjects. The few exceptions to this trend could reflect the dependence of tissue shear properties on *physiological age* rather than chronological age of the subjects. This is only speculative, however, as data on physiological fitness and general health of the subjects were not collected (e.g., cardiovascular and pulmonary data).

There was also an apparent gender effect. The *average* elasticity and viscosity of the male vocal-fold mucosa were around three to five times larger than those for females (Figs. 9–12). Sometimes the differences between female subjects and subjects at the upper end of the male range were as large as an order of magnitude. For example, at frequency  $>1$  Hz,  $\mu$  of the 72-year-old male and the 62-year-old male were an order of magnitude above those of the female subjects (Figs. 9 and 10). For  $\eta$  there was also a large difference across the entire frequency range of measurements.

An implication of these findings is that a similar pattern of intersubject differences in phonation threshold pressure ( $P_{th}$ ) may be expected in subjects of different age and gender.  $P_{th}$  has been shown to be directly related to the viscous properties (viscous shear modulus  $G''$  and dynamic viscosity  $\eta$ ) of the vocal-fold mucosa, according to previous analytical (Chan, 1998; Titze, 1988) and experimental studies (e.g., Titze *et al.*, 1995; Verdolini *et al.*, 1994). Because the present data suggested that older (and male) subjects may have a more viscous vocal-fold mucosa than younger (and female) ones, it is hypothesized that  $P_{th}$  may increase with

age, especially for males, and that it may be higher for males than females. That is, older and male subjects may need to expend more energy in phonation in order to overcome a larger amount of energy loss in their vocal-fold tissues. A careful literature search showed that a direct testing of this hypothesis is lacking, but the results of several studies demonstrated a pattern that was quite consistent with the hypothesis (Higgins and Saxman, 1991; Holmes *et al.*, 1994; Verdolini *et al.*, 1994).

Verdolini *et al.* (1994) measured the phonation threshold pressure in 12 normal subjects (3 males and 9 females) aged between 20–30.  $P_{th}$  was estimated from intraoral pressure measurements. Holmes *et al.* (1994) did not directly target  $P_{th}$  but measured the subglottal pressure (estimated from intraoral pressure) associated with “soft” phonation in 60 healthy subjects (30 men and 30 women) divided into three age groups (all subjects  $>55$  years old). Higgins and Saxman (1991) also estimated the subglottal pressure in soft voice in 41 healthy subjects, including 10 young females, 10 young males, 11 elderly females, and 10 elderly males. A collective analysis of the results of these studies showed that elderly subjects generally had phonation threshold pressure somewhat higher than that of younger subjects, particularly in males, which seemed to support the above hypothesis. On the other hand, regarding the effect of gender, results of these studies showed that  $P_{th}$  was similar for both male and female subjects, except in elderly subjects, where it was higher for males than females. Although this finding seemed to be inconsistent with our hypothesis, it should be noted that subjects in the above studies were phonating at a comfortable level of pitch, i.e., the female productions were made at a much higher fundamental frequency ( $F_0$ ) than the male ones. As it has been shown that  $P_{th}$  increases with  $F_0$  (Titze, 1992), it is likely that the results would have been different if the male and female productions were both made at a similar  $F_0$ . In order to test the above hypothesis directly, further studies should be designed in such a way that all the variables affecting  $P_{th}$ , particularly subject age and gender, can be carefully controlled and balanced.

Rheologists generally believe that viscoelastic shear properties are macroscopic reflections and realizations of microscopic and molecular events (Barnes *et al.*, 1989; Bird *et al.*, 1977; Ferry, 1980; Whorlow, 1980). Therefore, an interpretation of the above empirical findings with a molecular approach is given next.

#### D. Molecular interpretations of tissue shear properties

Recall that for a viscoelastic material, the elastic shear modulus  $\mu$  is a quantification of the energy storage in the material. A highly elastic material is capable of near-complete storage and recovery of internal energy during (shear) deformation. The dynamic viscosity  $\eta$ , on the other hand, is associated with dissipation of internal energy, typically as heat. A highly viscous material flows slowly and dissipates much energy in the shear process.

At the microscopic level, shear elasticity and viscosity of a material depend on the ease of relative displacement and slippage between (and within) molecules, which is deter-



mined by the number and strength of different kinds of *intramolecular* and *intermolecular* interactions. Some of the most significant molecular interactions include physical entanglement, electrostatic forces (e.g., hydrogen bond), hydrophilic and hydrophobic interactions (of polar and nonpolar chemical functional groups), and other physico-chemical interactions (e.g., covalent cross-link formation) (Barnes *et al.*, 1989; Bird *et al.*, 1977; Lehninger *et al.*, 1993). The more frequent and the stronger the molecular interactions, the more resistant the molecules are to relative displacement and slippage, and the higher the elasticity or viscosity.

These interactions are especially important in polymeric materials, where long chains of macromolecules occupy a large space relative to their atomic dimensions. For example, more densely packed polymeric chains have a higher probability of physical overlap and entanglements, chemical cross-link formation, and hence stronger electrostatic, hydrophilic, and hydrophobic interactions. Longer and more highly charged polymeric chains are also more likely to have physical entanglements, chemical cross-links, and more significant electrostatic interactions. Therefore, *concentration*, *molecular weight*, and *molecular structure* of the macromolecules are all important in dictating molecular interactions and subsequently material elasticity and viscosity (Barnes *et al.*, 1989; Ferry, 1980). As a variety of biomacromolecules is found in the vocal-fold lamina propria, possible differences at these molecular levels could account for the apparent age and gender effects in tissue shear properties found in the present study.

Vocal-fold molecular composition has been quite extensively studied previously. Cellular structures found in the lamina propria mainly include fibroblasts and macrophages, and some myofibroblasts (for tissue repair and wound healing), while the extracellular matrix of lamina propria consists of biomacromolecules like fibrous proteins (primarily collagen and elastin), glycosaminoglycans (e.g., hyaluronate, keratan sulfate, heparan sulfate, chondroitin sulfate, and dermatan sulfate), proteoglycans (glycosaminoglycans covalently attached to a protein core, e.g., versican, decorin, and fibromodulin), and structural glycoproteins (e.g., fibronectin) (Catten *et al.*, 1998; Gray *et al.*, 1993, 1999; Hammond *et al.*, 1997, 1998; Hirano, 1981; Matsuo *et al.*, 1984; Pawlak *et al.*, 1996). As cellular structures are relatively sparse in the lamina propria, including the superficial layer (vocal-fold mucosa), the extracellular matrix probably plays an important role in metabolism, as well as in dictating tissue shear properties. Among the many constituents found in the extracellular matrix, the fibrous proteins collagen and elastin are important mechanically because they form a scaffolding meshwork structure which supports interactions with other macromolecules in the lamina propria. For the superficial layer, nonfibrous or interstitial proteins like hyaluronate (physiological form of *hyaluronic acid* at a pH of 7.0–7.4), proteoglycans, and structural glycoproteins are probably more important as fibrous proteins are relatively sparse (Gray *et al.*, 1993, 1999; Hammond *et al.*, 1997, 1998; Hirano, 1981). They are important to tissue mechanical properties because of their huge molecular weights (often on the order of  $10^5$ – $10^6$ ), high (negative) charge densities and capabili-

ties of extensive molecular interactions (Gray *et al.*, 1999; Lehninger *et al.*, 1993; Pawlak *et al.*, 1996).

### 1. Elastin

Fragmentation, degeneration, and loss of elastin fibers have been common findings in aging of human aorta and skin, which are probably the results of a decrease in resistance to proteolysis as well as increased enzymatic (elastase) activity (Sell and Monnier, 1995). A reduction in the biosynthesis of tropoelastin molecules has also been reported in old-aged human skin fibroblasts, which would lead to decreased elastin content (Sephel and Davidson, 1986). In addition, there is also evidence for an age-related increase of covalent cross-links among elastin fibers in aorta, as well as cross-links between elastin and other proteins like collagen (Fujimoto, 1982; Powell *et al.*, 1992). Coupled with a simultaneous increase in relative collagen content (concentration) in old age, these findings suggest that the mechanical properties of these tissues tend to change towards more collagen-like and less elastin-like, including typically a decrease in extensibility (ultimate deformation) and an increase in elastic modulus or stiffness.

Some of these age-related changes of elastin have also been reported for human vocal-fold tissues. Similar abnormalities, atrophy, and loss of elastin fibers have been found in the lamina propria, especially for male subjects (Hirano *et al.*, 1989). Hence, the relative changes in collagen content versus elastin content described above are also possible in the aging vocal-fold mucosa. Such age-related molecular changes could be responsible for an increase in stiffness and viscosity of the vocal-fold mucosa with age, as they tend to transform the lamina propria mechanically into more collagen-like and less elastin-like. These molecular changes have been found to be more evident and pronounced in males than in females (Hirano *et al.*, 1989), agreeing with our viscoelastic data which showed a less clear age-related difference for female subjects, especially in tissue elasticity ( $\mu$ ).

Using elastin-van Gieson staining (EVG or Verhoeff's elastic tissue stain), light microscopy and electron microscopy, Hammond *et al.* (1997, 1998) quantified the relative amount of elastin across different layers of the lamina propria in male and female subjects in infant, adult, and geriatric age groups. Ultrastructural observations using electron microscopy revealed substantial quantities of relatively immature forms of elastin fibers in the superficial layer (mucosa), including oxytalan (a fibrillar form) and elaunin (an amorphous form mixed with fibrillar components). Results of quantitative comparisons showed that there was an increase of elastin content with age, but no clear gender-related differences were found. These findings suggest that the age-related differences (but not the gender-related differences) in tissue shear properties reported in the present study may be partly explained by a difference in elastin concentration.

### 2. Hyaluronic acid

Age-related changes of hyaluronate reported in the aging literature have been commonly obtained from connective tis-

sues like human articular cartilages and intervertebral disks. One consistent change of hyaluronate as part of an aggregating collagen–proteoglycan network structure has been the selective proteolytic breakdown of hyaluronate chains with higher molecular weights (on the order of  $10^6$ ). The consequences of such degradations include a gradual decrease of the average molecular weight of hyaluronate chains (from  $>10^6$  at age 20 to approximately  $5 \times 10^5$  at age 80), a slow accumulation of partially degraded hyaluronate (with lower molecular weights), and a progressive increase of hyaluronate content or concentration (approximately a twofold change from age 40 to age 80), the last of which may also be a consequence of increased biosynthesis (Holmes *et al.*, 1988; Sell and Monnier, 1995).

Possible effects of these hyaluronate changes on tissue shear properties depend on the balance between the effects of concentration and the effects of molecular size. An increase in hyaluronate content or concentration would facilitate molecular interactions and lead to an increase in elasticity and viscosity, while a decrease in the molecular weight of hyaluronate would tend to reduce molecular interactions and hence a decrease in shear properties. If the facilitative effects on molecular interactions due to higher concentrations do outweigh the effects of decreased molecular weights, these age-related changes of hyaluronic acid likely contribute significantly to an increase in tissue elasticity and viscosity with age, because the hyaluronate chains have relatively huge molecular weights, high charge densities, and thus much potential for molecular interactions.

Hammond *et al.* (1997) quantified the relative amount of hyaluronate in male versus female subjects by image analysis of histological sections of vocal-fold lamina propria stained with acid mucopolysaccharide (AMP), with and without hyaluronidase treatment. They found that male subjects had an *average* hyaluronate content three times higher than that of females, which suggested a higher probability of molecular interactions and hence an increase in tissue elasticity (stiffness) and viscosity. Recall that the average elasticity and viscosity of the male vocal-fold mucosa were found to be three to five times higher than those for females. This molecular finding of gender-related difference in hyaluronate content is therefore very well correlated with the apparent gender effect in tissue shear properties found in the present study.

In summary, the effects of age and gender on shear properties of vocal-fold mucosal tissues observed in the present study seem to be correlated with some of the age- and gender-related molecular differences in the vocal-fold lamina propria reported in the literature. However, more data on the effects of age and gender on major molecular constituents of the lamina propria (especially collagen and proteoglycans) and their molecular interactions are certainly needed in order to establish clearer relationships between molecular and viscoelastic data.

#### IV. SUMMARY AND CONCLUSION

Using a parallel-plate rotational rheometer, small-amplitude oscillatory shear experiments were performed to measure the viscoelastic shear properties of human vocal-

fold mucosal tissues (the superficial layer of lamina propria). Elastic shear modulus, dynamic viscosity, and damping ratio of tissue samples from 15 excised human larynges (10 male and 5 female) were quantified as a function of oscillation frequency (0.01 to 15 Hz). Intersubject differences in elastic shear modulus and dynamic viscosity as large as an order of magnitude were found. Part of this large intersubject variability may be attributed to age differences. Tissue samples of older subjects were apparently stiffer and more viscous than those of younger subjects, especially for males. There was also an apparent gender effect, with stiffer and more viscous mucosal tissues being observed for male subjects. These findings may be interpreted in terms of the age- and gender-related molecular differences in the vocal-fold lamina propria reported in the literature.

The major limitation of the present investigation was that rheological data were obtained at or below 15 Hz. Because of limitations of the rheometer's plate and sample inertia, the highest frequency at which linear small-amplitude oscillations were possible and meaningful measurements could be made was 15 Hz. As this was approximately an order of magnitude below typical frequencies of vocal-fold oscillation (usually  $>100$  Hz), extrapolations of the data to frequencies of phonation are needed in order that the shear properties become relevant to the context of vocal-fold oscillation. Theoretical modeling of tissue shear properties helpful to such data extrapolations is addressed in another paper (Chan and Titze, 1999b).

Another limitation was that measurements of tissue shear properties were based on the tissue samples of a relatively limited number of subjects (10 males and 5 females). Although age- and gender-related differences in shear properties were observed, further studies with an improved experimental design are needed to specifically assess age and gender effects. For example, age effect could be assessed by comparing the data of different age groups (e.g., with ages  $<40$ ,  $40-60$ ,  $>60$ ), each of which has a reasonable number of subjects of the same gender. On the other hand, gender effect may be assessed by comparing the data of two gender groups with age-matched subjects.

#### ACKNOWLEDGMENTS

This study was supported by Grant No. P60 DC00976 from the National Institute on Deafness and Other Communication Disorders, a division of the National Institutes of Health. The authors gratefully acknowledge the contributions of Steven D. Gray of the University of Utah, who kindly supplied us some of the larynges and useful dissection instruments. He also made valuable suggestions and comments on many aspects of the project, especially those related to laryngology and molecular biology. We also thank Brad H. Story of the Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts for his contributions in rheological measurements.

<sup>1</sup>Note that the vocal-fold mucosa of one subject (72-year-old male) was overdamped ( $\zeta > 1.0$ ) at very low frequency ( $< 0.03$  Hz) (Fig. 13). We think that the phonation of that subject was unlikely to have been overdamped, however, because Fig. 13 also shows that the damping ratio of his

- vocal-fold mucosa was decreasing monotonically with frequency, "converging" with the data of other subjects to a range of 0.1–0.3 at higher frequencies.
- Alipour-Haghighi, F., and Titze, I. R. (1988). "A finite element simulation of vocal fold vibrations," Proceedings of Northeast Bioengineering Conference, IEEE Report 88 CH 2666-6, pp. 186–189.
- Alipour-Haghighi, F., and Titze, I. R. (1991). "Elastic models of vocal fold tissues," *J. Acoust. Soc. Am.* **90**, 1326–1331.
- Arnold, G. E. (1963). "Alleviation of aphonia or dysphonia through intracordal injection of Teflon paste," *Ann. Otol. Rhinol. Laryngol.* **72**, 384–395.
- Baer, T. (1975). "Investigation of phonation using excised larynges," Unpublished Ph.D. dissertation, MIT, Cambridge, MA.
- Barnes, H. A., Hutton, J. F., and Walters, K. (1989). *An Introduction to Rheology* (Elsevier, Amsterdam), pp. 1–10; 37–54; 97–114.
- Benninger, M. S., Alessi, D., Archer, S., Bastian, R., Ford, C., Koufman, J., Sataloff, R. T., and Spiegel, J. R. (1996). "Vocal fold scarring: Current concepts and management," *Otolaryngol.-Head Neck Surg.* **115**, 474–482.
- Berry, D. A., and Titze, I. R. (1996). "Normal modes in a continuum model of vocal fold tissues," *J. Acoust. Soc. Am.* **100**, 3345–3354.
- Berry, D. A., Moon, J. B., and Kuehn, D. P. (1999). "A histologically-based finite element model of the soft palate," *Cleft Palate Craniofacial J.* **36**, 217–223.
- Bird, R. B., Armstrong, R. C., and Hassager, O. (1977). *Dynamics of Polymeric Liquids. Vol. 1: Fluid Mechanics* (Wiley, New York), pp. 129–204; 275–303.
- Catten, M., Gray, S. D., Hammond, T. H., Zhou, R., and Hammond, E. H. (1998). "An analysis of cellular location and concentration in vocal fold lamina propria," *Otolaryngol.-Head Neck Surg.* **118**, 663–667.
- Chan, R. W. (1998). "Shear properties of vocal fold mucosal tissues and their effect on vocal fold oscillation," Unpublished Ph.D. dissertation, The University of Iowa, Iowa City, IA.
- Chan, R. W., and Titze, I. R. (1998). "Viscosities of implantable biomaterials in vocal fold augmentation surgery," *Laryngoscope* **108**, 725–731.
- Chan, R. W., and Titze, I. R. (1999a). "Hyaluronic acid (with fibronectin) as a bioimplant for the vocal fold mucosa," *Laryngoscope* **109**, 1142–1149.
- Chan, R. W., and Titze, I. (1999b). "Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling," *J. Acoust. Soc. Am.* (in press).
- Chan, R. W., Titze, I. R., and Titze, M. R. (1997). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," *J. Acoust. Soc. Am.* **101**, 3722–3727.
- Courey, M. S., and Ossoff, R. H. (1995). "Surgical management of benign voice disorders," in *Diagnosis and Treatment of Voice Disorders*, edited by J. S. Rubin, R. T. Sataloff, G. S. Korovin, and W. J. Gould (Igaku-Shoin, New York), pp. 366–382.
- Ferry, J. D. (1980). *Viscoelastic Properties of Polymers*, 3rd ed. (Wiley, New York).
- Finkelhor, B. K., Titze, I. R., and Durham, P. L. (1987). "The effect of viscosity changes in the vocal folds on the range of oscillation," *J. Voice* **1**, 320–325.
- Ford, C. N., Bless, D. M., and Loftus, J. M. (1992). "Role of injectable collagen in the treatment of glottic insufficiency: a study of 119 patients," *Ann. Otol. Rhinol. Laryngol.* **101**, 237–247.
- Fujimoto, D. (1982). "Aging and crosslinking in human aorta," *Biochem. Biophys. Res. Commun.* **109**, 1264–1269.
- Fukuda, H., Kawasaki, Y., Kawaida, M., Shiotani, A., Oki, K., Tsuzuki, T., Fujioka, T., and Takayama, E. (1991). "Physiological properties and wave motion of the vocal fold membrane viewed from different directions," in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and G. Hammarberg (Singular, San Diego), pp. 7–14.
- Fung, Y. C. (1993). *Biomechanics. Mechanical Properties of Living Tissues*, 2nd ed. (Springer, New York), pp. 23–65; 242–320.
- Gray, S. D., Hirano, M., and Sato, K. (1993). "Molecular and cellular structure of vocal fold tissue," in *Vocal Fold Physiology: Frontiers in Basic Science*, edited by I. R. Titze (Singular, San Diego), pp. 1–35.
- Gray, S. D., Titze, I. R., Chan, R., and Hammond, T. H. (1999). "Vocal fold proteoglycans and their influence on biomechanics," *Laryngoscope* **109**, 845–854.
- Hammond, T. H., Gray, S. D., Butler, J., Zhou, R., and Hammond, E. (1998). "Age- and gender-related elastin distribution changes in human vocal folds," *Otolaryngol.-Head Neck Surg.* **119**, 314–322.
- Hammond, T. H., Zhou, R., Hammond, E. H., Pawlak, A., and Gray, S. D. (1997). "The intermediate layer: A morphologic study of the elastin and hyaluronic acid constituents of normal human vocal folds," *J. Voice* **11**, 59–66.
- Higgins, M. B., and Saxman, J. H. (1991). "A comparison of selected phonatory behaviors of healthy aged and young adults," *J. Speech Hear. Res.* **34**, 1000–1010.
- Hirano, M. (1975). "Phonosurgery: Basic and clinical investigations," *Otologia (Fukuoka)* **21**, 239–440.
- Hirano, M. (1981). "Structure of the vocal fold in normal and disease states anatomical and physical studies," in *Proceedings of the Conference on the Assessment of Vocal Pathology*, edited by C. L. Ludlow and M. O. Hard (American-Speech-Language-Hearing Association, Rockville, MD), pp. 11–30.
- Hirano, M., Kurita, S., and Sakaguchi, S. (1989). "Aging of the vibratory tissue of human vocal folds," *Acta Oto-Laryngol.* **107**, 428–433.
- Holmes, L. C., Leeper, H. A., and Nicholson, I. R. (1994). "Laryngeal airway resistance of older men and women as a function of vocal sound pressure level," *J. Speech Hear. Res.* **37**, 789–799.
- Holmes, M. W. A., Bayliss, M. T., and Muir, H. (1988). "Hyaluronic acid in human cartilage: Age-related changes in content and size," *Biochem. J.* **250**, 435–441.
- Ishizaka, K., and Flanagan, J. L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* **51**, 1233–1268.
- Isshiki, N. (1977). *Functional Surgery of the Larynx* (Kyoto University, Kyoto, Japan), pp. 62–67.
- Kaneko, T., Asano, H., Naito, J., Kobayashi, N., Hayashi, K., and Kitamura, T. (1972). "Biomechanics of the vocal cords-On damping ratio," *J. Jpn. Soc. Bronchoesophagol.* **25**, 133–138.
- Lehninger, A. L., Nelson, D. L., and Cox, M. M. (1993). *Principles of Biochemistry*, 2nd ed. (Worth, New York), pp. 298–321.
- Matsuo, K., Watanabe, S., Hirano, M., Kamimura, M., Tanaka, Y., and Takazono, I. (1984). "Acid mucopolysaccharide and glycoprotein in the vocal fold: Alterations with aging," *Pract. Otolaryngol. Kyoto* **77**, 817–822.
- Min, Y. B., Titze, I. R., and Alipour-Haghighi, F. (1995). "Stress-strain response of the human vocal ligament," *Ann. Otol. Rhinol. Laryngol.* **104**, 563–569.
- Pawlak, A. S., Hammond, T. H., Hammond, E. H., and Gray, S. D. (1996). "Immunocytochemical study of proteoglycans in vocal folds," *Ann. Otol. Rhinol. Laryngol.* **105**, 6–11.
- Perlman, A. L. (1985). "A technique for measuring the elastic properties of vocal fold tissue," Unpublished Ph.D. dissertation, The University of Iowa, Iowa City, IA.
- Powell, J. T., Vine, N., and Crossman, M. (1992). "On the accumulation of D-aspartate in elastin and other proteins of the ageing aorta," *Atherosclerosis (Berlin)* **97**, 201–208.
- Saito, S., Fukuda, H., Kitahara, S., Isogai, Y., Tsuzuki, T., Muta, H., Takayama, E., Fujioka, T., Kokawa, M., and Makino, K. (1985). "Pellet tracking in the vocal fold while phonating—Experimental study using canine larynges with muscle activity," in *Vocal Fold Physiology*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver), pp. 169–182.
- Sataloff, R. T., Spiegel, J. R., Hawkshaw, M. J., Rosen, D. C., and Heuer, R. J. (1997). "Autologous fat implantation for vocal fold scar: A preliminary report," *J. Voice* **11**, 238–246.
- Schramm, V. L., May, M., and Lavorato, A. S. (1978). "Gelfoam paste injection for vocal fold paralysis: Temporary rehabilitation of glottic incompetence," *Laryngoscope* **88**, 1268–1273.
- Sell, D. R., and Monnier, V. M. (1995). "Aging of long-lived proteins: extracellular matrix (collagens, elastins, proteoglycans) and lens crystallins," in *Handbook of Physiology. Section 11: Aging*, edited by E. J. Masoro (Oxford University Press, New York), pp. 235–305.
- Sephel, G. C., and Davidson, J. M. (1986). "Elastin production in human skin fibroblast cultures and its decline with age," *J. Invest. Dermatol.* **86**, 279–285.
- Shaw, G. Y., Szweczyk, M. A., Searle, J., and Woodroof, J. (1997). "Autologous fat injection into the vocal folds: Technical considerations and long-term follow-up," *Laryngoscope* **107**, 177–186.
- Titze, I. R. (1976). "On the mechanics of vocal-fold vibration," *J. Acoust. Soc. Am.* **60**, 1366–1380.
- Titze, I. R. (1988). "The physics of small-amplitude oscillation of the vocal folds," *J. Acoust. Soc. Am.* **83**, 1536–1552.

- Titze, I. R. (1992). "Phonation threshold pressure: A missing link in glottal aerodynamics," *J. Acoust. Soc. Am.* **91**, 2926–2935.
- Titze, I. R., Schmidt, S. S., and Titze, M. R. (1995). "Phonation threshold pressure in a physical model of the vocal fold mucosa," *J. Acoust. Soc. Am.* **97**, 3080–3084.
- Titze, I. R., and Talkin, D. T. (1979). "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation," *J. Acoust. Soc. Am.* **66**, 60–74.
- Verdolini, K., Titze, I. R., and Druker, D. G. (1990). "Changes in phonation threshold pressure with induced conditions of hydration," *J. Voice* **4**, 142–151.
- Verdolini, K., Titze, I. R., and Fennell, A. (1994). "Dependence of phonatory effort on hydration level," *J. Speech Hear. Res.* **37**, 1001–1007.
- Whorlow, R. W. (1980). *Rheological Techniques* (Ellis Horwood, Chichester, West Sussex, England), pp. 243–307.
- Wilhelms-Tricarico, R. (1995). "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," *J. Acoust. Soc. Am.* **97**, 3085–3098.



# Interarticulator phasing, locus equations, and degree of coarticulation

Anders Löfqvist<sup>a)</sup>

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511  
and Department of Logopedics and Phoniatrics, Lund University, Lund, Sweden

(Received 23 October 1998; revised 2 March 1999; accepted 24 May 1999)

A locus equation plots the frequency of the second formant at vowel onset against the target frequency of the same formant for the vowel in a consonant–vowel sequence, across different vowel contexts. It has generally been assumed that the slope of the locus equation reflects the degree of coarticulation between the consonant and the vowel, with a steeper slope showing more coarticulation. This study examined the articulatory basis for this assumption. Four subjects participated and produced VCV sequences of the consonants /b, d, g/ and the vowels /i, a, u/. The movements of the tongue and the lips were recorded using a magnetometer system. One articulatory measure was the temporal phasing between the onset of the lip closing movement for the bilabial consonant and the onset of the tongue movement from the first to the second vowel in a VCV sequence. A second measure was the magnitude of the tongue movement during the oral stop closure, averaged across four receivers on the tongue. A third measure was the magnitude of the tongue movement from the onset of the second vowel to the tongue position for that vowel. When compared with the corresponding locus equations, no measure showed any support for the assumption that the slope serves as an index of the degree of coarticulation between the consonant and the vowel. © 1999 Acoustical Society of America. [S0001-4966(99)02009-3]

PACS numbers: 43.70.Aj, 43.70.Fq [JH]

## INTRODUCTION

A locus equation is based on the onset frequency of the second formant and the steady-state value of the same formant during the vowel in a sequence of a consonant and a vowel. A linear regression is performed on these two values for productions of the same consonant in different vowel contexts (e.g., Lindblom, 1963; Sussman *et al.*, 1991). The consonants investigated using this paradigm have mostly been voiced stops or fricatives; it is hard to define the proper  $F_2$  onset frequency in a stop with a long voice onset time, where most of the transition occurs during the aspiration noise. Although there is an ongoing controversy about the proper interpretation of locus equations and their potential relevance for the perception of speech (e.g., Brancazio and Fowler, 1998; Sussman *et al.*, 1998), the present paper will not address these broader issues of locus equations. Its focus is rather on one property of the locus equation, i.e., its slope and the proper interpretation of slope differences.

Almost all the work on locus equations has been made using acoustic analysis. The first link between the slope of the locus equation and the degree of coarticulation between the consonant and the vowel in a CV sequence was made by Krull (1987, 1988). She noted that a locus equation slope of 1 would indicate maximum amount of coarticulation between a consonant and a vowel; in this case, the onset and target frequencies of the second formant would be identical. Conversely, a slope of 0 would indicate no coarticulation; here, the onset frequency of the second formant would remain the same when the consonant occurs in different vowel

contexts (cf. Figure 10 in Sussman *et al.*, 1993). Although this interpretation was made in terms of acoustic results, the conceptual link between locus equation slope and articulatory kinematics was also made by Krull (1987), who wrote ‘‘Earlier results referred to above and the fact that the tongue is not involved in their production led us to expect more coarticulation with labial consonants, and such was also the case here’’ (Krull, 1987, pp. 50–51). More recent work on locus equations has also made such a link explicit. For example, Sussman *et al.* (1993) state ‘‘The relatively steep regression functions for both labials and velars indicate two stop places where the following vowel greatly influences articulation of the preceding stop closure’’ (Sussman *et al.*, 1993, p. 1267). In a study of locus equations for consonants preceding and following a vowel, the failure to find the normal relationship for a consonant following a vowel has also been rationalized in articulatory terms: ‘‘The ‘vowel-dominant’ view, if taken to a logical conclusion, would also claim that the V- to C<sub>1</sub> anticipatory interaction is more likely a planned articulatory sequence both demanding and realizing greater articulatory precision compared to a VC<sub>2</sub> articulation which is conceived more as a defaultlike, mechanical contextual effect’’ (Sussman *et al.*, 1997, p. 2834). Finally, the studies of Fowler (1994) and Brancazio and Fowler (1998) relate slope differences between consonants with different places of articulation to their coarticulatory resistance defined in articulatory terms.

In extending the acoustically based measure of coarticulation using locus equation slope, it has thus been assumed that this measure can reflect the articulatory coarticulation between the consonant and the vowel. The evidence in support of this extension is so far almost entirely indirect and

<sup>a)</sup>Electronic-mail: lofqvist@haskins.yale.edu

based on inferred articulatory interpretations of slope differences. The notion of “degree of coarticulation” is interesting within the wider context of speech motor control, but there is little empirical evidence on which it can be based and evaluated. The present paper is an attempt to address this issue using records of articulatory kinematics and acoustic analyses.

The most common argument for a relationship between slope and degree of coarticulation is that labial stop consonants tend to have steeper slopes than alveolar and velar consonants (cf. Sussman *et al.*, 1991; Fowler, 1994; Brancazio and Fowler, 1998). The specific argument made is that in a sequence of a labial consonant and a vowel, the consonant does not require any specific tongue movements for its production, while the vowel does not require any lip movements for its production, unless it is produced with lip rounding. Since the analysis of locus equations is usually made using consonant–vowel syllables placed in a carrier sentence, it would appear, however, that any tongue movement during the labial consonant closure/constriction and/or during the vowel following the consonant would mostly reflect the influence of the vowel preceding the consonant rather than that of the consonant itself. This is so because the tongue movement from the first to the second vowel in a vowel–bilabial stop–vowel sequence will commonly start before the lip closure for the consonant has been made. Its exact timing varies, among other things, with the magnitude of the tongue movement from the first to the second vowel (Löfqvist and Gracco, 1999). The other evidence has been obtained in studies that have manipulated the linguistic material in ways that have been assumed to artificially affect the coarticulation, such as spontaneous versus read speech (Duez, 1992; Krull, 1989), stressed versus unstressed syllables (Bakran and Mildner, 1995; Duez, 1992; Krull, 1989), and by comparing sequences where the vowels preceding and following the consonant are identical or different (Brancazio and Fowler, 1998; Crowther, 1994). The expected slope differences have usually been found. One additional thing to keep in mind in this context is that for the velar stop consonant /g/, two locus equations have often been computed, one when it is paired with front vowels and another one when it is paired with back vowels. This has been done because a single regression across all vowel contexts usually results in a worse fit. It is, however, hard to see why the degree of coarticulation for the velar stop /g/ should differ as a function of vowel place. Both the tongue configuration and the point of tongue–palate contact for velar stops are heavily influenced by vowel context (e.g., Mooshammer *et al.*, 1995; Löfqvist and Gracco, 1994; Dembowski *et al.*, 1998). In fact, when a single regression is computed for /g/ across vowel contexts, the slope tends to be as steep for velars as for labials (Brancazio and Fowler, 1998). During the oral closure for a velar stop consonant, the tongue continues to move (Mooshammer *et al.*, 1995; Löfqvist and Gracco, submitted). For stop consonants produced with the tongue tip or the tongue blade, the location of the point of contact between the tongue tip and the palate, or alveolar ridge, is influenced by phonetic context (Kent and Moll, 1972; Butcher and Weiher, 1976). Also for these stops, the tongue continues to move

during the oral closure. A comparison between the movement of the tongue body for a velar stop and the tongue tip for an alveolar stop during the oral closure has shown that the tongue tip movement is smaller than that of the tongue body (Löfqvist and Gracco, submitted).

Recently, Chennoukh *et al.* (1997) used an articulatory model to examine the relationship between speech articulation and locus equations. With the model, known as the Distinctive Region Model (cf. Mrayati *et al.*, 1988), they varied the amount of articulatory overlap between a stop consonant and a vowel in a vowel–stop–vowel sequence to see if the amount of overlap changed the slope of the locus equation. In one condition, the tongue movement from the first to the second vowel and the oral closing movement for the consonant started at the same time. In a second condition, the vowel-related tongue movement did not start until the closure for the consonant had occurred. In the third condition, the vowel movement started at the release of the consonant. The slope of the locus equation decreased from condition 1 to condition 3.<sup>1</sup>

One purpose of the present experiment was to examine the influence of consonant–vowel overlap on locus equation slope, following the paradigm suggested by Chennoukh *et al.* (1997). However, one limitation on real speech has to be acknowledged. In their modeling work, Chennoukh *et al.* (1997) could vary the overlap between the consonant and vowel gestures for three places of articulation for the consonant, labial, alveolar, and velar. Using real speech, it is not possible to perform a similar analysis, however. This is so because, in contrast to the model used by Chennoukh *et al.*, it is not possible to observe separate consonant and vowel gestures in a human vocal tract when the consonant is articulated with the tongue. The model does not have a tongue, and the movements are simulated by changing constriction size and location. In humans, the movements for the consonant and the surrounding vowels are blended when the consonant is alveolar or velar. Only in a sequence where the consonant is bilabial is it theoretically possible to examine the consonant and vowel gestures separately; in practice, such an analysis rests on a number of assumptions that will be discussed in more detail below. In a sequence of an unrounded vowel, a bilabial stop, and an unrounded vowel, the stop closure is made with the lips and the jaw, while the vowel gesture is made with the tongue and the jaw. Thus, in one part of the present study, articulatory movements in VCV sequences were analyzed only for the bilabial stop /b/ and in asymmetrical vowel contexts, i.e., the two vowels were not identical. In a sequence with the same vowel occurring before and after the consonant, the tongue movement is very small and it is impossible to reliably identify the onset and offset of the movement. The acoustic analysis also used symmetrical vowel contexts, however.

What is the proper quantification of “degree of coarticulation?” One possible index has already been discussed above, i.e., the phasing between articulatory movements. In the case of a bilabial stop consonant occurring between two vowels, one can thus view the temporal phasing between the tongue movement from the first to the second vowel and the lip closing movement as providing a metric. When the

tongue movement starts well before the consonant movement, there is, in a sense, more overlap than if it starts late. Such temporal relationships have been the focus of several studies of anticipatory coarticulation (e.g., Perkell and Matthies, 1992; Abry and Lallouache, 1995). These studies have mostly focused on how far in advance one articulatory movement for a segment starts before the “segment itself.” A typical case is lip rounding for a vowel, where the lip movements have been shown to start well before the acoustic onset of the vowel itself. Note that this definition of a segment presupposes that it is possible to define a point in time where a segment properly “begins” in the acoustic and/or articulatory record. This is far from a trivial task, however, since the movements for successive segments blend seamlessly with each other. Thus, drawing boundaries between segments in acoustic or articulatory records requires that these boundaries be properly justified. In some cases, drawing such boundaries may, in fact, be impossible. Although they have provided a wealth of evidence on the temporal aspects of speech production, these studies of coarticulation do not provide much information about the shape of the vocal tract during speech production and how it is influenced by phonetic context. A proper metric of coarticulation would need to supply, or be based on, such information.

This study examined the relationship between locus equation slope and measures of articulatory overlap, or coarticulation. Three different measurements of coarticulation were used. The following specific hypotheses were examined.

The first measure of articulatory overlap was essentially the one used by Chennoukh *et al.* (1997). That is, it consisted of the temporal interval between the onset of the lip closing movement for a bilabial stop and the onset of the tongue movement from the first to the second vowel in a VCV sequence. The predicted outcome here would be that as the overlap between the tongue and lip movements increased, the slope of the locus equation should become steeper. An increased articulatory overlap would be indexed by an earlier onset of the tongue movement relative to the onset of the lip movement. The second metric chosen for degree of coarticulation was the magnitude of the tongue movement during the oral closure for a stop consonant. The prediction would be that a large tongue movement during the closure would signify a high degree of coarticulation between the stop consonant and the following vowel, and hence be associated with a steeper slope, because the consonant would allow rather than resist such coarticulatory influences. A third metric chosen was the magnitude of the tongue movement during the vowel following the consonant. A small tongue movement between the consonant release and the vowel target would imply a large degree of articulatory overlap, and thus a steeper slope, since most of the tongue movement would have occurred during the consonant. Since the movement of the whole tongue will affect the vocal tract shape, for both these measures the movement magnitude was taken as the average across four receivers placed on the tongue. As a first approximation, a large change in vocal tract shape would lead to a large acoustic difference, while a smaller articulatory change would lead to a smaller acoustic change. Here, we immedi-

ately have to acknowledge a limiting factor for making direct comparisons between articulation and acoustics. There is a lawful relation between articulation (vocal tract shape) and the acoustic signal, but it is not linear. A further limitation in the present context is that we are tracking only receivers at the midline of the tongue and not the whole vocal tract. Nevertheless, it appears to be fruitful to explore the possibility of such a link between articulation and acoustics within the framework of locus equations.

## I. METHOD

### A. Subjects

Two female (LK, DR) and two male subjects (VG, AL) participated. All subjects had normal speech and hearing and no history of speech or hearing disorders. Three of the subjects (LK, DR, VG) are native speakers of American English. Subjects LK and DR grew up in the Mid-West, while subject VG grew up in Florida; they all currently live in the Northeast. Speaker AL is a native speaker of Swedish who is also fluent in English. Subject AL is the author.

### B. Linguistic material

The linguistic material consisted of  $V_1CV_2$  sequences, where the first and second vowels ( $V_1$  and  $V_2$ ) were always one of /i, a, u/, and the consonant (C) was one of /b, d, g/. The sequences were placed in the carrier phrase “Say... again” with sentential stress occurring on the second vowel ( $V_2$ ) of the sequence. Ten repetitions of each sequence were recorded.

### C. Procedure

The movements of the lips, the jaw, and the tongue were recorded using a three-transmitter magnetometer system (Perkell *et al.*, 1992). Receivers were placed on the upper and lower lips, on the lower incisors, and on four positions on the tongue. The tongue receivers will be referred to as tongue tip, tongue blade, tongue body, and tongue rear. The lip receivers were placed below and above the vermilion border of the upper and lower lip, respectively, with a vertical separation of approximately 1 cm when the lips were in a closed position. For the tongue, the first receiver was placed as far back as the subject could tolerate, and the next one close to the tongue tip; next, an attempt was made to space the other two receivers evenly between the first and the second. Two additional receivers placed on the nose and the upper incisors were used for the correction of head movements. The receivers on the lips, the incisors, and the nose were attached using Iso-Dent (Ellman International). For the tongue receivers, Ketac-Bond (ESPE) was used. Care was taken during each receiver placement to ensure that it was positioned at the midline with its long axis perpendicular to the sagittal plane. Two receivers attached to a plate were used to record the occlusal plane by having the subject bite on the plate during recording. All data were subsequently corrected for head movements and rotated to bring the oc-



clusal plane into coincidence with the  $x$  axis. This rotation was performed to obtain a uniform coordinate system for all subjects (cf. Westbury, 1994).

The articulatory movement signals (induced voltages from the receiver coils) were sampled at 625 Hz after low-pass filtering at 200 Hz. The resolution for all signals was 12 bits. After voltage-to-distance conversion, the movement signals were low-pass filtered using a 25-point triangular window with a 3-dB cutoff at 17 Hz. To obtain instantaneous velocity, the first derivative of the position signals was calculated using a three-point central difference algorithm. The velocity signals were smoothed using the same triangular window. A measure of lip opening was obtained by subtracting the vertical position of the lower lip receiver from that of the upper lip receiver. Only the vertical distance was used, since the lip movements for a bilabial stop closure primarily occur in the vertical dimension (Löfqvist and Gracco, 1997). All the signal processing was made using the Haskins Analysis Display and Experiment System (HADES) (Rubin and Löfqvist, 1996). The acoustic signal was preemphasized, low-pass filtered at 9.5 kHz, and sampled at 20 kHz.

The definition of the onset of the closing movement of the lips for the bilabial stop presented some problems. Usually, movement onsets are defined at zero crossings in velocity signals. Here, the second derivative of the derived lip opening signal was used. Using a zero crossing in the first derivative of the lip opening signal was difficult when the first vowel was /u/ that included lip rounding. In this case, the rounding gesture made the lip opening change continuously and a zero crossing would not appear in the first derivative at a point in time close to the oral closure. Thus, the onset of the lip closure for the stop consonant was defined as the minimum in the second derivative of the lip opening signal prior to the oral closure, cf. Fig. 1. This point was defined algorithmically. In a strict sense, this point is not the ‘onset’ of the lip closing movement, but it is related to it.<sup>2</sup> Figure 1 presents the acoustic, lip opening, and tongue body signals for one production of the sequence ‘abu’ by subject VG. Since the interpretation of a second derivative is not always straightforward, the lip opening signal and its first derivative are also included in Fig. 1. We should add that the actual lip opening is at zero throughout the oral closure. The change in the lip opening signal during the closure is due to the fact that it represents the vertical distance between the receivers on the upper and lower lips, and these receivers move during the closure (cf. Löfqvist and Gracco, 1997; Westbury and Hashi, 1997). The tongue body receiver was used for defining measurement points for the tongue movements. Its speed [ $v = \sqrt{\dot{x}^2 + \dot{y}^2}$ ] was calculated. Tongue movement onsets and offsets were identified algorithmically from the speed signal as minima during the first and second vowels. Their identification is also shown in Fig. 1. The onset is shown by the vertical line labeled ‘First vowel’ in Fig. 1, while the offset of the movement is shown by the vertical line labeled ‘Second vowel’ in Fig. 1. We should note that at these points in time, the horizontal and vertical velocity of the tongue is not necessarily zero. However, at these points in time the tongue movement is minimal.

The onset and release of the oral closure were identified

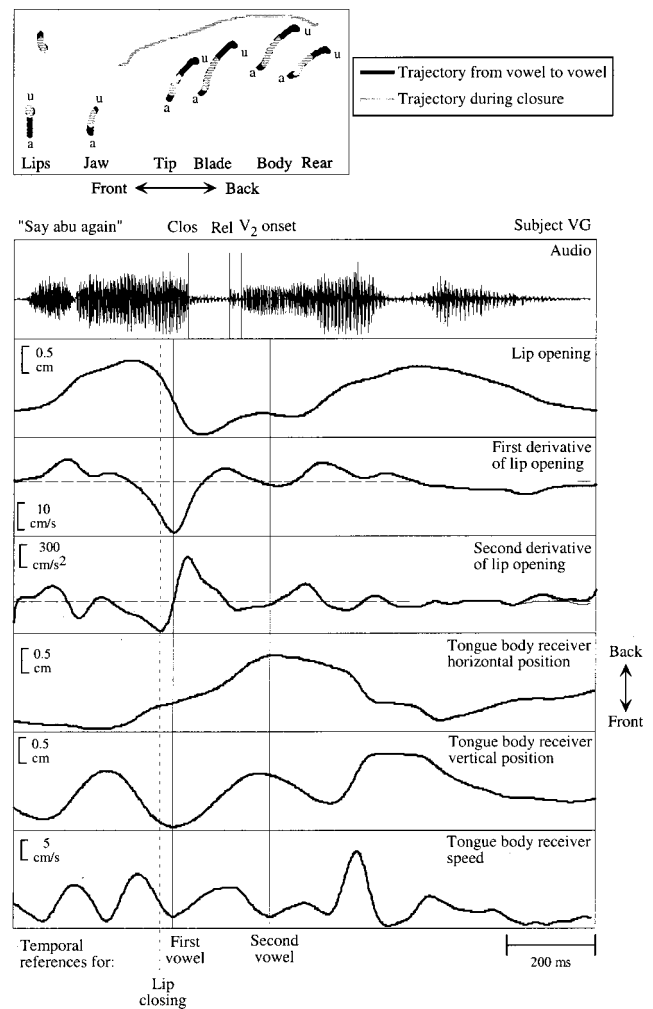


FIG. 1. Acoustic and movement signals during the production of the utterance ‘Say abu again’ by subject VG. In addition to the derived lip opening signal, its first and second derivatives are shown. The labels in the audio signal correspond to the onset and release of the oral closure for the /b/, and the onset of the second vowel in /abu/. The label in the second derivative of the lip opening signal was used to mark the ‘onset’ of the lip closing movement for the consonant. The labels in the speed signal of the tongue body identify the onset and offset of the tongue movement from the first to the second vowel in /abu/. The top panel shows the receiver trajectories from the first to the second vowel and also the part of the trajectory made during the oral closure for the movement.

in waveform and spectrogram displays of the acoustic signal. The onset of the closure was identified by the decrease in the amplitude of the acoustic waveform, and by the disappearance of spectral energy at higher frequencies. The release was identified by its burst. The onset of regular glottal vibrations for the second vowel was also marked. The second formant frequency were measured from both DFT and LPC spectra with 28 coefficients, using a Hamming window, a 51-ms window size, and a window separation of 6 ms. Measurements were made at the onset and at the midpoint of the vowel following the criteria described by Sussman *et al.* (1991). All labeling and acoustic measurements were made by the author.

As a measure of the temporal overlap between the bilabial stop consonant and the vowel, the interval between the onset of the tongue movement from the first to the second vowel and the onset of the lip closing movement was used.



## Bilabial stops in asymmetrical vowel contexts

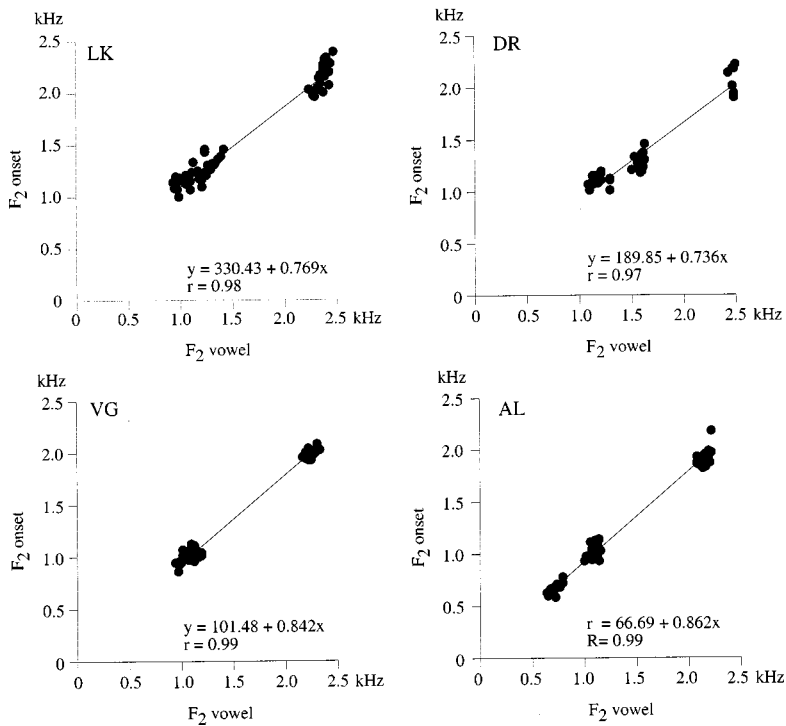


FIG. 2. Scatter plots, locus equations, and correlation coefficients for the consonant /b/ produced in asymmetrical vowel contexts.

The magnitude of the tongue movement trajectory during the oral closure was obtained by summing the Euclidean distances between successive samples between the acoustically defined onset and offset of the closure for each receiver. A measure of overall tongue movement was then obtained as the average of the four receiver movements during the oral closure. The same procedure was used to measure the average tongue movement from the acoustically defined onset of the second vowel to the end of the tongue movement from the first to the second vowel.

The kinematic signals represent the movements of receivers placed at the midline of the lips, the jaw, and the tongue. When presenting the results, we will use the terms “tongue body receiver” and “tongue body” interchangeably, while acknowledging that we are only examining the movements of a single point. Thus, we make no claims about asymmetrical movements of the left and right sides of the lips, or the tongue. The tongue and lower lip signals contain the contribution of the jaw. These are the appropriate movements to examine when the focus of the analysis is on the lower lip and the tongue as end effectors. It is reasonable to assume that a speaker has joint control of different articulators during speech production to produce the desired results.

## II. RESULTS

Figure 2 shows plots of the second formant at vowel onset and vowel steady state together with the locus equations for the sequences with a voiced bilabial stop /b/ and asymmetrical vowel contexts. The derived locus equations show a high degree of linearity and the slope values are similar to the ones reported previously in the literature. To see if there was a difference in the slope of the locus equations between symmetrical and asymmetrical vowel contexts,

separate regressions were calculated for the symmetrical contexts. The slopes for the symmetrical vowel contexts were 0.889, 0.712, 0.878, and 0.948 for subjects LK, DR, VG, and AL, with all  $r$  values 0.99. Thus, with the exception of subject DR, the slopes were higher in the symmetrical vowel context.

The analysis of locus equation slopes and the measure of articulatory overlap between a labial stop consonant and the following vowel will focus on differences between speakers. Chennoukh *et al.* (1997) interpreted their results in terms of differences between speakers. In addition, within-speaker comparisons are difficult, since the calculation of the locus equation requires that several different vowel contexts be used. Figure 3 plots the slope of the locus equation for each

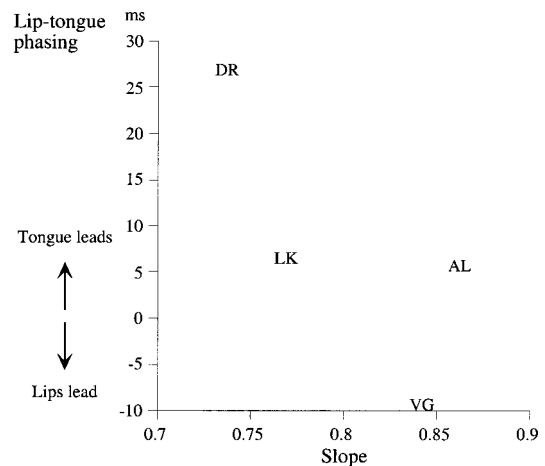


FIG. 3. The slope of the locus equation for each subject plotted against the interval between onset of tongue movement from the first to the second vowel and the onset of the lip closing movement for the consonant. The letters refer to the four subjects.

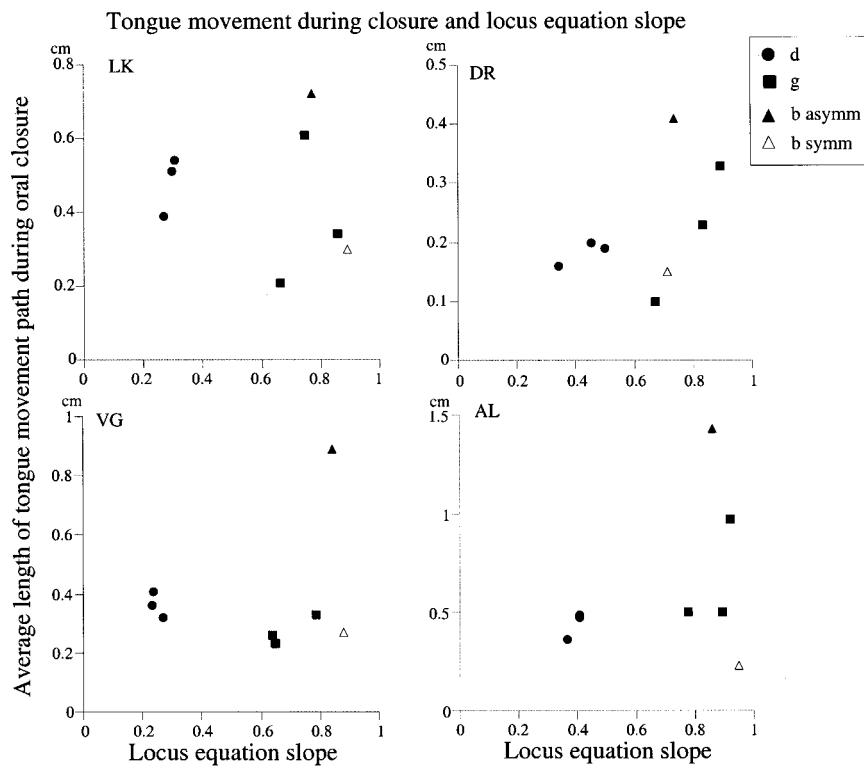


FIG. 4. Plot of the slope of the locus equation and the magnitude of the tongue movement during the stop closure.

subject against the interval between the onset of the tongue movement from the first to the second vowel and the onset of the lip closing movement for the consonant. The slope of the locus equation is that for the asymmetrical vowel contexts. The measure of lip-tongue phasing has been averaged across all sequences with a voiced bilabial stop /b/ and asymmetrical vowel contexts. The subjects are represented by their initials. According to the hypothesis that a higher slope is related to a greater degree of articulatory overlap, we would expect the following relationship between the two variables. In those instances where the tongue movement starts well before the lip movement, thus signifying a large degree of articulatory overlap, the slope of the locus equation should be high. That is, as the tongue movement leads the lip movement, the slope of the locus equation should increase. However, that is not what is shown in Fig. 3. The results show rather the opposite. We should note here that even though the lip movement may lead the tongue movement, the tongue movement does not necessarily start after the oral closure. The reason is that the onset of the lip movement, as defined here, always precedes the oral closure, cf. Fig. 1.

Figure 4 plots the slope of the locus equation against the magnitude of the tongue movement during the oral closure for the stop consonant. The results for the velar and alveolar consonants are shown separately for the three different first vowel contexts. In addition, the results for the bilabial stops in symmetrical and asymmetrical vowel contexts are shown. The plots show that for all four subjects, the slope for the velar stop is higher than that for the alveolar consonant. The working assumption here is that a large tongue movement during the closure is associated with more coarticulation, and hence a higher slope. The relationship between the two variables is not very strong and quite different between subjects, however. Note that the labial stop in the asymmetric vowel

environment is associated with the largest tongue movements during the closure for all subjects. It does not necessarily have the highest slope, however. In comparison, the tongue movement is much smaller for the bilabial stop in a symmetric vowel environment, but the slope difference between the labial stops in these two vowel environments is rather small. Although the slope is always higher for the velar than for the alveolar stop, the magnitude of the tongue movement during the oral closure tends to be larger for the alveolar stop for subjects LK and VG, and about the same for subjects DR and AL. In some instances, there is almost a linear relationship between the two variables, e.g., the alveolar stops for subjects LK and AL, and the velar stops and the (symmetrical) labial stops for subject DR. Here, the relationship between tongue movement and locus equation slope is, in fact, the expected one.

Figure 5 plots the slope of the locus equation for each subject against the magnitude of the tongue movement during the vowel following the consonant. According to the hypothesis, one might expect that when there is a small amount of tongue movement during the second vowel, the slope of the locus equation should be high. This is so because most of the tongue movement would have been anticipated and occurred during the consonant, since it would allow it. A smaller movement during the second vowel would correspond to greater articulatory overlap and thus a steeper slope. One thing to note in Fig. 5 is that the bilabial stop in the asymmetric vowel context has the least amount of tongue movement during the second vowel. The locus equation slope is not necessarily the highest for this consonant, however. Only subject AL shows an overall negative relationship between the two variables plotted in Fig. 5, which is the predicted result. For the other three subjects, there is no clear overall relationship. In two instances, the velar stops for sub-

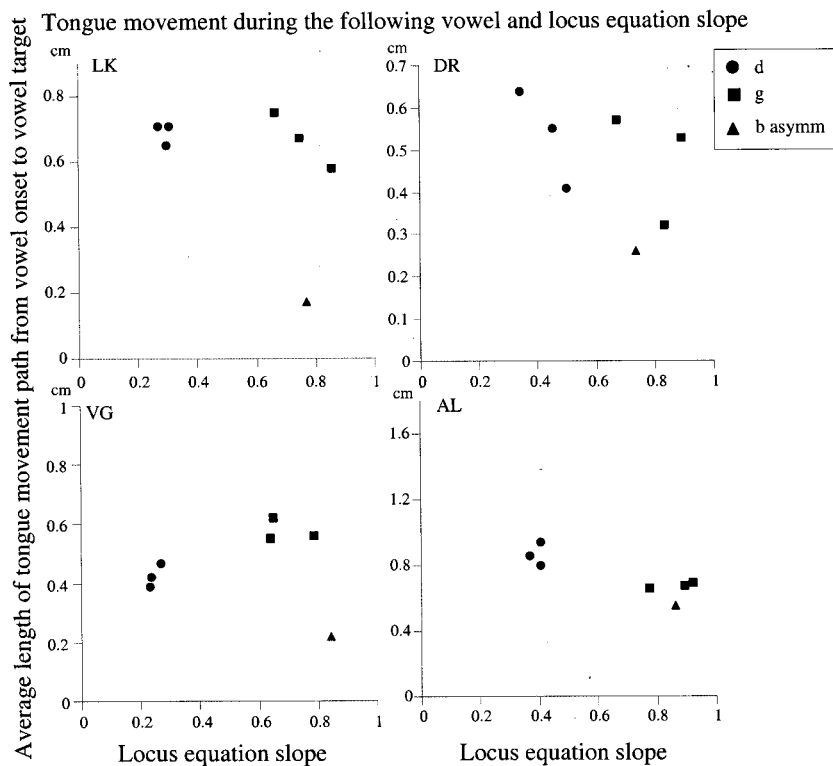


FIG. 5. Plot of the slope of the locus equation and the magnitude of the tongue movement during the vowel following the stop consonant.

ject LK and the alveolar stops for subject DR, there is an almost linear negative relationship as predicted.

### III. DISCUSSION

The locus equations obtained in the present study are in good agreement with those presented in other studies. Alveolar consonants have a less steep slope than labial and velar consonants. When the velar consonants are analyzed across front and back vowel contexts, their slope is as high as that for the labial consonants. Before discussing the present results, it is important to note that the articulatory analyses presented here are, by necessity, coarse grained. This is so because the nature of locus equations requires that they be obtained across a variety of vowel contexts. A single consonant production does not have a locus equation slope, so comparisons cannot be made at a finer grain of analysis. Another limitation is that only four receivers placed on the midline of the tongue have been recorded. This restriction on receiver placement along the midline is an inherent limitation on the recording technique used. The movements of the rearmost part of the tongue have not been recorded and analyzed due to the difficulty of placing receivers in this area, and also because grooving of the tongue may make the recordings unreliable due to receiver tilt. In spite of these limitations, the present results form a first examination of locus equation slope and articulatory overlap, or coarticulation.

The present results do not show any real support for the idea that the slope of the locus equations is related to the degree of coarticulation between the consonant and the vowel. At least, this is not the case for the three measures of coarticulatory overlap used here. The first one is essentially adapted from Chennoukh *et al.* (1997), who found the expected relationship in their simulations. One reason for these

conflicting results appears to be that their simulations included some degrees of overlap that are not found in human productions. In particular, the tongue movement hardly ever starts after the oral closure for the consonant and never at the release of the consonant (cf. Löfqvist and Gracco, 1999). The latter was one condition used in the simulations, however. Another reason is most likely that human subjects vary the amplitude, velocity, and duration of their tongue movements in different ways. Again, this is in contrast to the simulations, where these parameters were held constant.

The second measure of coarticulation showed a substantial difference in the amount of tongue movement during the oral closure for the bilabial stops in symmetrical and asymmetrical vowel contexts, but a very small difference in locus equation slope. Similarly, there was a consistent difference in locus equation slope between alveolar and velar stops for all subjects, with the slope being higher for the velar consonants. However, the amount of tongue movement during the closure did not show any consistent related difference with a larger movement for the velar stops.

Although it may be generally true that a labial stop consonant will allow quite substantial movements of the tongue during the oral closure [more than 50% of the vowel-to-vowel trajectory in a VCV sequence (Löfqvist and Gracco, 1999)], this assumes that the context in fact requires movements of the tongue. That is not the case in a VCV sequence where the two vowels are identical. Hence, Fig. 5 shows a substantial difference in the amount of tongue movement for a bilabial stop in symmetrical and asymmetrical vowel contexts, but a very small difference in locus equation slope. Possibly, the slope is so high in these two cases that the slope measure is unable to capture the difference in tongue movement. It is more troublesome for the hypothesized relationship between locus equation slope and coarticulation that the

articulatory difference between alveolar and velar stops are not really related to the great difference in slope.

The third measure was based on the assumption that a large degree of articulatory overlap between a consonant and a following vowel would result in a small movement during the vowel, since most of the movement has already been made during the consonant. Again, this measure did not show the overall expected relationship to the slope of the locus equation. One reason for concern here is that the relationship between articulation and acoustics is lawful but not linear.

It thus appears that the hypothesis about a link between locus equation slope and degree of articulatory overlap (coarticulation) as operationally defined here should be questioned. The hypothesis is not supported because of the finding that labial stops in different vowel contexts have similar slopes but differ in degree of coarticulation, and also by the finding that alveolar and velar stops differ in slope but show the same degree of coarticulation. Although there were a few cases for some subjects where the expected relationship between slope and coarticulation was found, the overall results suggest otherwise. Hence, it appears that caution is necessary in making claims that locus equation slope can be used to assess differences in coarticulation and applied to studies of developmental issues (e.g., Sussman *et al.*, 1996), or deviant speech production, as suggested by Sussman *et al.* (1998).

The technique for recording articulatory movement used here only provides information on receivers placed at the midline of the articulators under investigation. Hence, the overall vocal tract shape cannot be captured. This is a limitation of the technique that can only be alleviated by trying to get estimates of the tongue shape by curve fitting procedures. Another way to address the relationship between degree of coarticulation and locus equation slope would be to use an analysis-by-synthesis approach with articulatory models that allow selective manipulations of articulators (cf. Lindblom, 1998). One simulation briefly reported by Lindblom (1998) does, in fact, show some support for a relationship between slope and coarticulation. In this simulation, a /dV/ syllable was modeled using two conditions. In one of them, the tongue body had the shape for the vowel during the oral closure for the stop. In the other condition, the tongue shape was more neutral. The derived locus equation slope was 0.94 in the first condition and 0.07 in the second. Further simulation work using realistic variations in tongue shape would obviously be useful for understanding coarticulatory overlap and acoustics.

The arguments for a relationship between the slope of the locus equation and the degree of coarticulation between the consonant and the vowel in a CV sequence have been based on inferences about articulatory organization based on phonological and/or phonetic descriptions of consonants and vowels. It appears from the present results, however, that making predictions about the detailed characteristics of articulatory movements in connected speech on such a basis is at best difficult and at worst impossible, as has also been suggested by Mooshammer *et al.* (1995) and by Dembowski *et al.* (1998).

## ACKNOWLEDGMENTS

I am grateful to René Carré for discussions at the early stages of this work and to Carol Fowler for deconstructing the first version of this study. Comments made by James Hillenbrand, Harvey Sussman, and John Westbury on an earlier version of the manuscript have also been very helpful. This work was supported by Grant Nos. DC-00865 and DC-02717 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

<sup>1</sup>Although these simulations do provide support for a relationship between degree of articulatory overlap and locus equation slope, there are a few things that should be considered before these results are accepted. All the movements used in the simulations were of constant duration, amplitude, and velocity. Perhaps more importantly, the variation in the amount of overlap between the three conditions is quite extreme compared to human articulatory data. In fact, a recent study by Löfqvist and Gracco (1999) indicated that the onset of the tongue movement from the first to the second vowel in a sequence of a vowel, a bilabial stop, and a vowel almost always occur before the oral closure for the consonant has occurred. Although there is variability in the temporal phasing of the lip closing movement and the tongue movement, both the lip and the tongue movements start before the oral closure for the consonant.

<sup>2</sup>The temporal interval between movement onset and peak velocity does not appear to change very much across variations in movement amplitude for some speech movements (e.g., Löfqvist and Yoshioka, 1981; Munhall and Ostry, 1985) and also for arm movements (Freund and Büdingen, 1978). Hence, the same might well be true for acceleration.

Aby, C., and Lallouache, T. (1995). "Le MEM: Un modèle d'anticipation paramétrable par locuteur, Données sur l'arrondissement en français," *Bull. Comm. Parlée* 3, 85–99.

Bakran, J., and Mildner, V. (1995). "Effect of speaking rate and coarticulation strategies on the locus equation determination," in *Proceedings of the XIII International Congress of Phonetic Sciences*, edited by K. Elenius and P. Branderud (KTH and Stockholm University, Stockholm), Vol. I, pp. 26–29.

Brancazio, L., and Fowler, C. (1998). "The relevance of locus equations for production and perception of stop consonants," *Percept. Psychophys.* 60, 24–50.

Butcher, A., and Weiher, E. (1976). "An electropalatographic investigation of coarticulation in VCV sequences," *J. Phonetics* 4, 59–74.

Chennoukh, S., Carré, R., and Lindblom, B. (1997). "Locus equations in the light of articulatory modeling," *J. Acoust. Soc. Am.* 102, 2380–2389.

Crowther, C. S. (1994). "Modeling coarticulation and place of articulation using locus equations," *UCLA Working Papers in Phonetics* 88, 127–148.

Dembowski, J., Lindstrom, M., and Westbury, J. (1998). "Articulator point variability in the production of stop consonants," in *Neuromotor Speech Disorders: Nature, Assessment, and Management*, edited by M. Cannito, K. Yorkston, and D. Beukelman (Brookes, Baltimore, MD), pp. 27–46.

Duez, D. (1992). "Second-formant locus patterns: An investigation of spontaneous French speech," *Speech Commun.* 11, 417–427.

Fowler, C. A. (1994). "Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation," *Percept. Psychophys.* 55, 597–610.

Freund, H.-J., and Büdingen, H. (1978). "The relationship between speed and amplitude of the fastest voluntary contractions of human arm muscles," *Exp. Brain Res.* 31, 1–12.

Kent, R., and Moll, K. (1972). "Cinefluorographic analyses of selected lingual consonants," *J. Speech Hear. Res.* 15, 453–473.

Krull, D. (1987). "Second formant locus patterns as a measure of consonant-vowel coarticulation," *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm, PERILUS V*, 43–61.

Krull, D. (1988). *Acoustic Properties as Predictors of Perceptual Responses: A Study of Swedish Voiced Stops* (Institute of Linguistics, University of Stockholm, Stockholm).

Krull, D. (1989). "Consonant-vowel coarticulation in spontaneous speech and reference words," *Phonetic Experimental Research, Institute of Linguistics, Stockholm University, PERILUS X*, 101–105.



- Lindblom, B. (1963). *On vowel reduction* (Royal Institute of Technology, Stockholm), Report 29.
- Lindblom, B. (1998). "An articulatory perspective on the locus equation," *Behav. Brain Sci.* **21**, 274–275.
- Löfqvist, A., and Gracco, V. L. (1994). "Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context," *Phonetica* **51**, 52–67.
- Löfqvist, A., and Gracco, V. L. (1997). "Lip and jaw kinematics in bilabial stop consonant production," *J. Speech Hear. Lang. Res.* **40**, 877–893.
- Löfqvist, A., and Gracco, V. (1999). "Interarticulator programming in VCV sequences: Lip and tongue kinematics," *J. Acoust. Soc. Am.* **105**, 1864–1876.
- Löfqvist, A., and Gracco, V. (submitted). "Control of oral closure in lingual stop consonant production," manuscript submitted for publication.
- Löfqvist, A., and Yoshioka, H. (1981). "Interarticulator programming in obstruent production," *Phonetica* **38**, 21–34.
- Mooshammer, C., Hoole, P., and Kühnert, B. (1995). "On loops," *J. Phonetics* **23**, 3–21.
- Mrayati, M., Carré, R., and Guérin, B. (1988). "Distinctive region and modes: A new theory of speech production," *Speech Commun.* **7**, 257–286.
- Munhall, K., and Ostry, D. (1985). "Ultrasonic measurements of laryngeal kinematics," in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I. Titze and R. Scherer (The Denver Center for the Performing Arts, Denver), pp. 141–161.
- Perkell, J., and Matthies, M. (1992). "Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability," *J. Acoust. Soc. Am.* **91**, 2911–2925.
- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabeta, I., and Jackson, M. (1992). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," *J. Acoust. Soc. Am.* **92**, 3078–3096.
- Rubin, P. E. R., and Löfqvist, A. (1996). "HADES (Haskins Analysis Display and Experiment System)," Haskins Labs. Status Rep. Speech Res. (available at [www.haskins.yale.edu/haskins/SR/SRweb/SRwebl.html](http://www.haskins.yale.edu/haskins/SR/SRweb/SRwebl.html)).
- Sussman, H., Hoemke, K., and Ahmed, F. (1993). "A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation," *J. Acoust. Soc. Am.* **94**, 1256–1268.
- Sussman, H., Bessell, N., Dalston, E., and Majors, T. (1997). "An investigation of stop place of articulation as a function of syllable position: A locus equation perspective," *J. Acoust. Soc. Am.* **101**, 2826–2838.
- Sussman, H. M., Fruchter, D., Hilbert, J., and Sirosh, J. (1998). "Linear correlates in the speech signal: The orderly output constraint," *Behav. Brain Sci.* **21**, 241–299.
- Sussman, H., McCaffrey, H., and Matthews, S. (1991). "An investigation of locus equations as a source of relational invariance for stop place categorization," *J. Acoust. Soc. Am.* **90**, 1309–1325.
- Sussman, H., Minifie, F., Buder, E., Stoel-Gammon, C., and Smith, J. (1996). "Consonant-vowel interdependencies on babbling and early words: Preliminary examination of a locus equation approach," *J. Speech Hear. Res.* **39**, 424–433.
- Westbury, J. (1994). "On coordinate systems and the representation of articulatory movements," *J. Acoust. Soc. Am.* **95**, 2271–2273.
- Westbury, J., and Hashi, M. (1997). "Lip-pellet positions during vowels and labial consonants," *J. Phonetics* **25**, 405–419.

# Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words

J. Sean Allen<sup>a)</sup> and Joanne L. Miller<sup>b)</sup>

Department of Psychology, Northeastern University, Boston, Massachusetts 02115

(Received 22 February 1999; revised 18 June 1999; accepted 21 June 1999)

Two speech production experiments tested the validity of the traditional method of creating voice-onset-time (VOT) continua for perceptual studies in which the systematic increase in VOT across the continuum is accompanied by a concomitant decrease in the duration of the following vowel. In experiment 1, segmental durations were measured for matched monosyllabic words beginning with either a voiced stop (e.g., *big, duck, gap*) or a voiceless stop (e.g., *pig, tuck, cap*). Results from four talkers showed that the change from voiced to voiceless stop produced not only an increase in VOT, but also a decrease in vowel duration. However, the decrease in vowel duration was consistently less than the increase in VOT. In experiment 2, results from four new talkers replicated these findings at two rates of speech, as well as highlighted the contrasting temporal effects on vowel duration of an increase in VOT due to a change in syllable-initial voicing versus a change in speaking rate. It was concluded that the traditional method of creating VOT continua for perceptual experiments, although not perfect, approximates natural speech by capturing the basic trade-off between VOT and vowel duration in syllable-initial voiced versus voiceless stop consonants. © 1999 Acoustical Society of America. [S0001-4966(99)01710-5]

PACS numbers: 43.70.Fq, 43.70.Hs, 43.71.Es [AL]

## INTRODUCTION

An important acoustic correlate that signals the voicing of an initial stop consonant is voice-onset-time (VOT), defined acoustically as the time from the release burst of the stop consonant to the onset of high-amplitude, quasiperiodic energy associated with the following vowel. The role of VOT in speech perception is often studied by creating speech continua in which VOT is systematically varied from short VOT values appropriate for a voiced stop consonant to longer VOT values appropriate for a voiceless stop consonant. Typically (though not universally, e.g., Borsky *et al.*, 1998; Summerfield, 1981), when such VOT continua are created, the increase in VOT across the continuum is accompanied by a concomitant decrease in the duration of the following vowel,<sup>1</sup> such that overall syllable duration is held constant across all tokens in the continuum (e.g., Diehl *et al.*, 1980; Ganong, 1980; Kluender, 1991; Kuhl, 1981; Lisker and Abramson, 1970; Pind, 1995; Pisoni and Lazarus, 1974; Samuel, 1982). Thus in the traditional method of creating VOT series, VOT and vowel duration trade off with one another in a perfect one-to-one manner across the change in syllable-initial voicing. As a consequence, the vowel at the voiced endpoint of the series is longer than the vowel at the voiceless endpoint of the series and, moreover, the difference in vowel duration equals the difference in VOT.

One purpose of the present investigation was to determine how closely this traditional method of creating VOT continua reflects the way in which VOT and vowel duration covary in speech production with a change in syllable-initial voicing. Specifically, the traditional method tacitly assumes

that two conditions obtain in natural speech: First, the increase in VOT resulting from a change from voiced to voiceless consonant is accompanied by a decrease in the duration of the following vowel, such that vowels following voiced stops are longer than vowels following voiceless stops; and second, the decrease in vowel duration is equal in magnitude to the increase in VOT, resulting in a constant overall syllable duration.

The speech production literature is relatively sparse concerning the effect of syllable-initial voicing on the duration of the following vowel. However, a number of studies have reported data, typically in the context of investigations not specifically focused on this issue, that provide a basis for anticipating the results of the present study. Peterson and Lehiste (1960) report that for words in a constant sentence frame, vowels following voiced stops tend to be longer in duration than vowels following voiceless stops. Crystal and House (1988) report a similar pattern in connected speech, and at least two early studies also report tendencies in the same direction (Heffner, 1940; Rositzke, 1939). In addition, Pind (1995) and Fischer-Jørgensen (1964) find a comparable pattern for words spoken in citation form in Icelandic and Danish, respectively. Taken together, these findings suggest that the first condition outlined above, that vowels following voiced stops are longer in duration than vowels following voiceless stops, may be true. However, Fischer-Jørgensen (1964; see also Heffner, 1940; Pind, 1995) points out that the magnitude of the decrease in vowel duration following voiceless compared to voiced stops is insufficient to compensate fully for the increase in VOT. These results call into question the second condition outlined above, namely, that the decrease in vowel duration should be equal in magnitude to the increase in VOT for voiced versus voiceless stops.

<sup>a)</sup>E-mail: jallen@lynx.neu.edu

<sup>b)</sup>E-mail: jlmiller@neu.edu

Instead, it appears that the difference in vowel duration may be smaller than the difference in VOT, so that a constant syllable duration is not maintained across a change in syllable-initial voicing. Experiment 1 was designed to test directly the way in which VOT and vowel duration covary with a change from voiced to voiceless initial stop consonant and, in so doing, evaluate the extent to which the traditional method of creating VOT continua for perceptual experiments approximates the temporal patterns found in natural speech.<sup>2</sup>

A second and closely related purpose of the investigation, which forms the basis of experiment 2, was to contrast the effects on vowel duration of a change in syllable-initial voicing with those of a change in speaking rate. This experiment was motivated in part by a recent paper by Kessinger and Blumstein (1998) challenging the validity of the traditional method of creating VOT continua. Their challenge is based on findings from production studies that investigate how VOT and vowel duration covary as a function of speaking rate. In their own production study, Kessinger and Blumstein (1998) replicate previous results (e.g., Kessinger and Blumstein, 1997; Miller *et al.*, 1986; Pind, 1995; Volaitis and Miller, 1992) showing that as speaking rate slows, both VOT and vowel duration increase for words beginning with a voiceless stop consonant. On the basis of this finding, Kessinger and Blumstein argue that (at least for those studies that examine speaking rate effects) the traditional method of creating VOT continua, in which vowel duration decreases as VOT increases along the continuum, is invalid. However, the way in which VOT and vowel duration covary for a given voiceless stop consonant across changes in speaking rate is a separate issue from the way in which VOT and vowel duration covary for voiced versus voiceless stop consonants within a given speaking rate. Because VOT continua are intended to range from a voiced to a voiceless stop consonant at a given speaking rate, the positive relationship between VOT and vowel duration across speaking rates does not directly bear on the validity of the traditional method of creating VOT continua. The purpose of experiment 2, which introduced a speaking rate manipulation, was to clarify this issue by directly comparing the way in which VOT and vowel duration vary as a function of a change in speaking rate to the way in which they vary as a function of a change in syllable-initial voicing within a given speaking rate.

## I. EXPERIMENT 1

The purpose of the first experiment was to test the way in which VOT and vowel duration covary with a change from syllable-initial voiced to voiceless consonant for matched word pairs (e.g., *big-pig*, *duck-tuck*, *gap-cap*). If the traditional method of creating VOT continua for perceptual experiments successfully captures the temporal characteristics of such spoken syllables, then two findings will emerge. First, a change from syllable-initial voiced to voiceless consonant (e.g., from *big* to *pig*) will result in both an increase in VOT and a decrease in vowel duration. Second, the decrease in vowel duration will be equal in magnitude to the increase in VOT, such that overall syllable duration will remain constant for the two words in a given pair.

TABLE I. Critical word pairs for experiment 1, for three places of articulation of the initial stop consonant. The critical word pairs used in experiment 2 are designated by asterisks.

Labial	Place of articulation	
	Alveolar	Velar
back - pack*	dab - tab*	goat - coat*
big - pig*	dug - tug*	good - could*
bat - pat	dip - tip	gap - cap
bet - pet	duck - tuck	gut - cut
bub - pub	dub - tub	gape - cape

## A. Method

### 1. Subjects

The speech of four talkers, two female and two male, was analyzed. All talkers were native speakers of American English with no reported speech or hearing disorders. Talkers were paid for their participation.

### 2. Materials

For each syllable-initial place of articulation (labial, alveolar, velar), five stop-vowel-stop monosyllabic word pairs were selected, such that each pair consisted of words that differed phonetically from one another only in the voicing of the initial stop consonant. The 15 selected word pairs (30 critical words), which contain a variety of vowels and final stop consonants, are shown in Table I.

These 30 critical words, along with 60 filler words, were used to create an experimental word list. The filler words were all monosyllabic words beginning with a variety of consonants, none of which was a stop consonant. The list was created as follows. In order to ensure that no two critical words from a given pair appeared adjacently or close to one another in the list, the 30 critical words were divided equally into five blocks, such that no two words from the same critical pair occurred in the same or in adjacent blocks. The 60 filler words were equally distributed among the five blocks. The order of the words within each block was then randomized, and the individual blocks were concatenated to form a single sublist of 90 words. This sublist was repeated five times, with a different within-blocks randomization each time, to yield the final experimental list of 450 words.

### 3. Procedure

Each talker was recorded individually. The words were presented to the subject visually one at a time on a computer screen in a sound-treated booth. Each word was displayed for 1500 ms, followed by a pause of 1500 ms before the next word was displayed. Talkers were instructed to say each word as it appeared on the screen. They were asked to speak at a comfortable rate in a clear but natural manner. Their speech was recorded via a microphone (AKG C 460B) onto digital audio tape (Aiwa HHB 1 Pro DAT recorder). The microphone was positioned approximately four inches from the talker's mouth and input recording levels were set so as to obtain a high-amplitude signal with no clipping.

TABLE II. Mean VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration, in milliseconds, for words beginning with voiced versus voiceless stop consonants. The means are collapsed across talkers and words, and standard deviations (across talkers) are shown in parentheses. Experiment 1.

Initial voicing	VOT	Vowel	Closure	Release burst
Voiced	16 (1.0)	188 (15.7)	104 (21.7)	122 (26.7)
Voiceless	70 (13.0)	170 (19.3)	103 (20.2)	121 (17.1)

#### 4. Acoustic analysis

The recording of each critical word was transferred to a PC at a sampling rate of 20 kHz using the CSL system (Kay Elemetrics Corp.). All acoustic measurements were made from the CSL waveform display, with supporting reference to spectrographic analysis and auditory output. VOT was measured from the onset of the initial release burst to the onset of periodicity. Vowel duration was measured from the onset of periodicity to the offset of periodicity associated with vocalic energy, marked by an abrupt decrease in amplitude. For completeness, the closure duration and the release burst duration for the final consonant were also measured, although we did not expect these to differ as a function of the voicing of the initial consonant. The closure duration of the final stop was measured from the offset of vocalic energy to the onset of the final release burst. The duration of the release burst was measured from the onset of the burst to the offset of measurable energy in the signal.

A small number of tokens exhibited prevoicing, in which periodicity began before the release of the initial stop consonant. Prevoiced words were not included in the analyses. In addition, mispronunciations and hesitations were not included in the analyses. In all, less than 2% of the data were excluded for these reasons. Mean durations were calculated across the five tokens (less excluded tokens) of each word for each talker. These means served as the basis for all further analyses.

All reported measurements were made by a single highly trained experimenter. To assess the reliability of the measurements, a subset of 20% of the tokens from each talker was also measured by a second experimenter. Measurements across the two experimenters were highly correlated ( $r=0.99, 0.99, 0.98,$  and  $0.88$  for VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration, respectively) and tightly clustered (mean differences were less than 1 ms for VOT and vowel duration, and less than 2 ms for final-consonant closure duration and final-consonant release burst duration).

#### B. Results

The mean results for the four acoustic measures (VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration) are given in Table II and are displayed in Fig. 1. The means are collapsed across the 15 word pairs and across the four talkers. As expected, initial voiceless stops had a longer VOT than initial voiced stops. In addition, initial voiceless stops were associated with a shorter vowel duration than initial voiced stops. However,

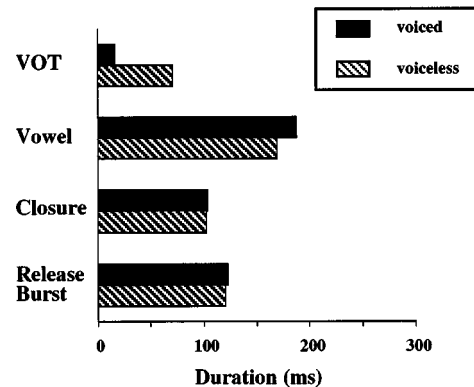


FIG. 1. Mean duration measurements (averaged across talkers and words) for VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration, for words beginning with voiced versus voiceless consonants. Experiment 1.

the change in vowel duration was smaller in magnitude than the change in VOT. There was little effect of syllable-initial voicing on the two acoustic measures for the final consonant.

To provide statistical support for these observations, we conducted separate analyses of variance for the effect of syllable-initial voicing (voiced versus voiceless) on each of the four acoustic measures, VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration. Furthermore, for each acoustic measure, we performed two analyses, one using subjects as the random variable ( $F1$ ) and the other using items as the random variable ( $F2$ ). As expected, the effect of syllable-initial voicing on VOT was highly reliable [ $F1(1,3)=78.37, p<0.01; F2(1,14)=1656.77, p<0.001$ ]. In addition, the effect of syllable-initial voicing on vowel duration was also highly reliable [ $F1(1,3)=73.07, p<0.01; F2(1,14)=67.36, p<0.001$ ]. The effect of syllable-initial voicing on the two final consonant acoustic measures was, as expected, nonsignificant ( $p>0.10$  in each case).

The trade-off between VOT and vowel duration was robust, in that the increase in VOT with a change in syllable-initial voicing was accompanied with rare exception by a decrease in vowel duration in each of the word pairs for each of the four individual talkers. However, VOT and vowel duration did not trade off in a one-to-one fashion. Instead, as the initial consonant changed from voiced to voiceless, the increase in VOT was greater than the decrease in vowel duration for each of the individual word pairs (collapsed across talkers) and for each of the talkers (collapsed across word pairs). This resulted in a net increase in the duration of the CV portion of words (defined as VOT plus vowel duration) beginning with voiceless compared to voiced stop consonants. This pattern can be seen in Table III, where means (collapsed across the four talkers) for VOT, vowel duration, and CV duration are shown for each of the critical word pairs. The change in each of these measures as a function of initial-consonant voicing (voiceless–voiced) is also shown. For each word pair, VOT increased as the initial consonant changed from voiced to voiceless, indicated by a positive VOT difference in the table, whereas vowel duration de-



TABLE III. Mean VOT, vowel duration, and CV duration (VOT+vowel duration), in milliseconds, for words beginning with voiced versus voiceless stop consonants. The means are collapsed across talkers. Change in duration for each measure as a function of syllable-initial voicing (voiceless-voiced), as well as the absolute proportion of change in vowel duration to change in VOT, are also shown. Experiment 1.

Voiced initial stop				Voiceless initial stop				Change (voiceless-voiced)			
Word	VOT	Vowel	CV	Word	VOT	Vowel	CV	VOT	Vowel	CV	Vowel/VOT
<i>back</i>	9	206	215	<i>pack</i>	64	190	254	55	-16	39	0.29
<i>big</i>	10	195	205	<i>pig</i>	61	186	247	51	-9	42	0.18
<i>bat</i>	9	212	221	<i>pat</i>	70	177	247	61	-35	26	0.57
<i>bet</i>	9	154	163	<i>pet</i>	62	148	210	53	-6	47	0.11
<i>bub</i>	9	177	186	<i>pub</i>	68	169	237	59	-8	51	0.14
<i>dab</i>	17	283	300	<i>tab</i>	79	255	334	62	-28	34	0.45
<i>dug</i>	17	211	228	<i>tug</i>	76	191	267	59	-20	39	0.34
<i>dip</i>	18	123	141	<i>tip</i>	64	108	172	46	-15	31	0.33
<i>duck</i>	17	147	164	<i>tuck</i>	67	126	193	50	-21	29	0.42
<i>dub</i>	15	192	207	<i>tub</i>	75	180	255	60	-12	48	0.20
<i>goat</i>	23	181	204	<i>coat</i>	74	163	237	51	-18	33	0.35
<i>good</i>	25	211	236	<i>could</i>	80	198	278	55	-13	42	0.24
<i>gap</i>	20	206	226	<i>cap</i>	72	177	249	52	-29	23	0.56
<i>gut</i>	23	152	175	<i>cut</i>	77	126	203	54	-26	28	0.48
<i>gape</i>	20	169	189	<i>cape</i>	66	155	221	46	-14	32	0.30

creased as the initial consonant changed from voiced to voiceless, indicated by a negative vowel duration difference in the table. In each case the change in vowel duration was less than the change in VOT, resulting in a net increase in CV duration as the initial consonant changed from voiced to voiceless; this is indicated by a positive CV duration difference in the table. In order to determine if the magnitude of the trade-off between VOT and vowel duration was stable across phonetic context (e.g., place of articulation of the initial stop, vowel identity), we calculated the absolute proportion of change in vowel duration to change in VOT for each of the word pairs, as shown in the rightmost column of the table. This proportion was not constant, ranging from 0.11 to 0.57 across the word pairs, and moreover, there was no apparent patterning of this proportion as a function of phonetic context. However, because our materials were not specifically designed to control for phonetic context, we cannot rule out the possibility that it determines, at least in part, the precise relation between the change in VOT and the change in vowel duration.

### C. Discussion

If the traditional method of creating VOT continua for perceptual experiments fully captures the temporal characteristics of spoken syllables, then two conditions should be met in the acoustic analyses. The first is that a change from syllable-initial voiced to voiceless consonant (e.g., from *big* to *pig*) should result not only in an increase in VOT, but also in a decrease in vowel duration. Our results are fully in accord with this prediction, in that for every word pair (collapsed across talkers) and for every talker (collapsed across word pairs), vowels following voiceless stops were shorter than vowels following voiced stops. The second prediction, that the decrease in vowel duration should be equal in magnitude to the increase in VOT (thereby maintaining overall syllable duration) was not upheld. For every word pair (collapsed across talkers) and for every talker (collapsed across word pairs), the change in vowel duration was less than the

change in VOT, with the consequence that the CV portion of the syllable was longer for words beginning with voiceless compared to voiced consonants. Based on these findings, which are in accord with earlier reports in the literature (e.g., Fischer-Jørgensen, 1964; Pind, 1995), we conclude that the traditional method of creating VOT continua approximates the temporal pattern found in speech production by trading off an increase in VOT with a decrease in vowel duration across the continuum, but it does not capture the one-to-one nature of the trade-off.

As we noted in the Introduction, Kessinger and Blumstein (1998) have recently challenged the assumption of a trade-off between VOT and vowel duration that underlies the traditional method of creating VOT continua. Their challenge is based on findings from production studies (including their own) showing that a change in speaking rate for voiceless stops produces an increase in both VOT and vowel duration. However, as we argued earlier, because VOT continua are intended to range from a voiced to a voiceless stop consonant at a given speaking rate, the positive relationship between VOT and vowel duration across speaking rates does not directly bear on the appropriateness of the traditional method of creating VOT continua. The purpose of experiment 2 was to further clarify this issue by directly comparing the effect of speaking rate and the effect of syllable-initial voicing within a given speaking rate on VOT and vowel duration in speech production.<sup>3</sup>

## II. EXPERIMENT 2

Experiment 2, like experiment 1, tested the assumptions of the traditional method of creating VOT continua regarding the covariation of VOT and vowel duration. In addition, experiment 2 introduced a speaking rate manipulation. We expected that at a given speaking rate (fast or slow), the change from voiced to voiceless initial stop consonant would result in an increase in VOT that was accompanied by a smaller decrease in vowel duration, replicating experiment 1. In contrast, we expected that for any given voiceless stop conso-

nant, the change from fast to slow speaking rate would result in an increase in VOT that was accompanied by an increase in vowel duration, replicating previous findings (e.g., Kessinger and Blumstein, 1997; Kessinger and Blumstein, 1998; Miller *et al.*, 1986; Volaitis and Miller, 1992; see also Pind, 1995). For voiced stop consonants, we expected only a minimal (if any) change in VOT with a change in speaking rate from fast to slow, given that prior results have shown that VOT values for voiced stops are relatively insensitive to changes in speaking rate (e.g., Kessinger and Blumstein, 1997; Miller *et al.*, 1986; Volaitis and Miller, 1992; see also Pind, 1995); we did, nevertheless, expect that for voiced stops the change from fast to slow would produce an increase in vowel duration.<sup>4</sup> Finally, in line with experiment 1, we expected that the change in syllable-initial voicing would have no effect on either final-consonant closure or final-consonant release burst duration at either rate of speech; however, we expected both of these durations to increase as speaking rate changed from fast to slow.

## A. Method

### 1. Subjects

Four new talkers, two female and two male, participated in experiment 2. All were native speakers of American English with no reported speech or hearing disorders. Talkers were paid for their participation.

### 2. Materials

A subset of stimuli from experiment 1 was selected for use in experiment 2. Two word pairs were selected for each place of articulation of the initial stop consonant, yielding six critical word pairs (12 critical words). The selected word pairs are indicated by asterisks in Table I. These 12 critical words, together with 24 filler words also drawn from experiment 1, were used to create the experimental word list, maintaining the proportion of critical words to filler words that was used in the first experiment. The new experimental word list was created in an analogous manner to that in experiment 1, such that no two critical words from a given pair appeared close to one another in the final list. There were five repetitions of each of the 12 critical words and 24 filler words, yielding a final experimental list of 180 words. Two different (within-block) randomizations of this 180-word list were created, one for the SLOW speaking rate condition and one for the FAST speaking rate condition.

### 3. Procedure

All subjects participated in two speaking rate conditions, FAST and SLOW, with the SLOW condition run first (this condition was run first in order to most closely replicate the conditions of experiment 1). The procedures and equipment for the SLOW condition were identical to those used in experiment 1. Each word was again displayed for 1500 ms, followed by a pause of 1500 ms before the next word was displayed, and subjects were instructed to say each word at a comfortable speaking rate. The procedure for the FAST condition was the same as that for the SLOW condition, except that the subjects were instructed to say the words at a faster

TABLE IV. Mean VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration, in milliseconds, for words beginning with voiced versus voiceless stop consonants, in both the FAST and SLOW speaking rate conditions. The means are collapsed across talkers and words, and standard deviations (across talkers) are shown in parentheses. Experiment 2.

Initial voicing	VOT	Vowel	Closure	Release burst
FAST speaking rate:				
Voiced	14 (1.0)	160 (29.4)	72 (9.6)	67 (18.0)
Voiceless	49 (7.9)	146 (29.8)	77 (7.8)	62 (20.8)
SLOW speaking rate:				
Voiced	13 (1.2)	208 (13.9)	105 (23.4)	106 (25.7)
Voiceless	78 (4.7)	182 (16.7)	107 (22.6)	101 (21.4)

rate of speech, and the words were displayed for 750 rather than 1500 ms in order to encourage talkers to speak at the faster rate. Subjects were given a short break between conditions.

## 4. Acoustic analysis

The procedures for acoustic analysis were the same as those used in experiment 1. Prevoiced words, hesitations, and mispronunciations were again eliminated from the analyses: 6.5% of the data were excluded for these reasons. More data were excluded in this experiment than in experiment 1 (where less than 2% of the data were excluded) primarily because of an increased tendency toward prevoicing (across both speaking rates) for the talkers in experiment 2. As in experiment 1, means were calculated across the five tokens (less excluded tokens) of each word for each talker, and these means served as the basis for all further analyses. The same experimenter as in experiment 1 made all reported measurements, and reliability was again assessed by a second experimenter who also measured 20% of the tokens from each talker. Measurements across the two experimenters were highly correlated ( $r=0.99, 0.98, 0.94,$  and  $0.76$  for VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration, respectively) and tightly clustered (mean differences were less than 1 ms for VOT, 3 ms for vowel duration and final-consonant closure duration, and 7 ms for final-consonant release burst duration).

## B. Results

The mean results for the four acoustic measures (VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration), collapsed across the six word pairs and across the four talkers, are presented in Table IV and displayed in Fig. 2. As in experiment 1, both subjects ( $F1$ ) and items ( $F2$ ) analyses of variance were conducted on the data for each of the four acoustic measures. In this experiment, the analyses for each measure involved two factors, syllable-initial voicing (voiced versus voiceless) and speaking rate (fast versus slow).

Consider first the factor of syllable-initial voicing. The left portion of Fig. 2 (panels A and B) shows the results of a change in syllable-initial voicing separately for the FAST and SLOW speaking rate conditions. The pattern of results

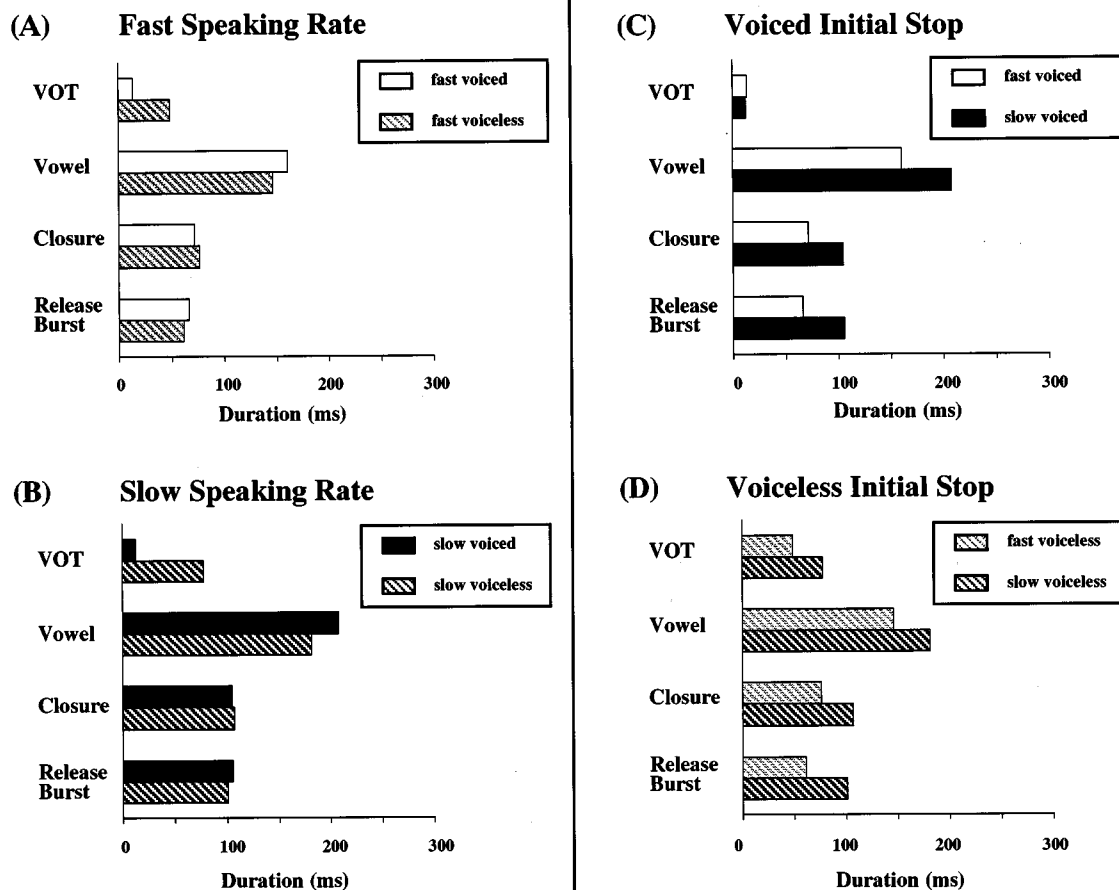


FIG. 2. Mean duration measurements (averaged across talkers and words) for VOT, vowel duration, final-consonant closure duration, and final-consonant release burst duration, for words beginning with voiced versus voiceless consonants, at fast and slow speaking rates. The same data are grouped according to speaking rate in panels A and B and according to syllable-initial voicing in panels C and D (see text for explanation). Experiment 2.

for both speaking rate conditions replicated experiment 1. As the initial consonant changed from voiced to voiceless, there was a highly reliable overall increase in VOT [ $F1(1,3) = 379.65, p < 0.001$ ;  $F2(1,5) = 476.84, p < 0.001$ ]. Although there was a reliable interaction between syllable-initial voicing and speaking rate for VOT [ $F1(1,3) = 50.73, p < 0.01$ ;  $F2(1,5) = 133.97, p < 0.001$ ], the increase in VOT due to a change from voiced to voiceless initial consonant was reliable for both the fast and slow speaking rates ( $p < 0.01$  in each case). The change in syllable-initial voicing also produced a highly reliable overall decrease in vowel duration [ $F1(1,3) = 503.87, p < 0.001$ ;  $F2(1,5) = 169.78, p < 0.001$ ]. Although there was a reliable interaction between syllable-initial voicing and speaking rate for vowel duration [ $F1(1,3) = 39.75, p < 0.01$ ;  $F2(1,5) = 14.31, p < 0.025$ ], the decrease in vowel duration as a function of syllable-initial voicing was reliable for both speaking rates ( $p < 0.01$  in each case). There were no reliable effects of syllable-initial voicing on the duration of either the final-consonant closure or release burst ( $p > 0.10$  in each case).

Consider next the factor of speaking rate. The right portion of Fig. 2 (panels C and D) displays the same data as in the left portion, but now arranged separately for the voiced and voiceless initial consonants in order to highlight the ef-

fect of speaking rate on each of the four acoustic measures. As speaking rate changed from fast to slow, there was a highly reliable overall increase in VOT [ $F1(1,3) = 39.05, p < 0.01$ ;  $F2(1,5) = 65.38, p < 0.001$ ]. The interaction noted above between syllable-initial voicing and speaking rate for VOT is apparent in that the change in rate produced a reliable increase in VOT for voiceless consonants ( $p < 0.01$ ) but (not surprisingly, given the existing literature) not voiced consonants ( $p > 0.10$ ). There was also a highly reliable overall effect of speaking rate on vowel duration [ $F1(1,3) = 27.70, p < 0.025$ ;  $F2(1,5) = 34.12, p < 0.01$ ]. Although as noted above there was a reliable interaction between syllable-initial voicing and speaking rate for vowel duration, the speaking rate effect held for both the voiced and voiceless initial consonants ( $p < 0.01$  in each case). Finally, both the duration of the final-consonant closure [ $F1(1,3) = 18.41, p < 0.025$ ;  $F2(1,5) = 17.24, p < 0.01$ ] and the duration of the final-consonant release burst [ $F1(1,3) = 25.03, p < 0.025$ ;  $F2(1,5) = 50.87, p < 0.001$ ] reliably increased with the change in speaking rate from fast to slow; speaking rate did not interact with syllable-initial voicing for either measure ( $p > 0.10$  in each case).

Table V shows the group means (collapsed across the four talkers) for VOT, vowel duration, and CV duration, for

TABLE V. Mean VOT, vowel duration, and CV duration (VOT+vowel duration), in milliseconds, for words beginning with voiced versus voiceless stop consonants, in both the FAST and SLOW speaking rate conditions. The means are collapsed across talkers. Change in duration for each measure as a function of syllable-initial voicing (voiceless-voiced), as well as the absolute proportion of change in vowel duration to change in VOT, are also shown. Experiment 2.

Word	Voiced initial stop			Word	Voiceless initial stop			Change (voiceless-voiced)			
	VOT	Vowel	CV		VOT	Vowel	CV	VOT	Vowel	CV	Vowel/VOT
FAST speaking rate:											
<i>back</i>	11	148	159	<i>pack</i>	42	142	184	31	-6	25	0.19
<i>big</i>	9	149	158	<i>pig</i>	39	141	180	30	-8	22	0.27
<i>dab</i>	11	211	222	<i>tab</i>	45	193	238	34	-18	16	0.53
<i>dug</i>	13	162	175	<i>tug</i>	53	143	196	40	-19	21	0.48
<i>goat</i>	17	149	166	<i>coat</i>	53	133	186	36	-16	20	0.44
<i>good</i>	20	142	162	<i>could</i>	59	124	183	39	-18	21	0.46
SLOW speaking rate:											
<i>back</i>	9	176	185	<i>pack</i>	67	149	216	58	-27	31	0.47
<i>big</i>	6	201	207	<i>pig</i>	61	173	234	55	-28	27	0.51
<i>dab</i>	11	282	293	<i>tab</i>	86	251	337	75	-31	44	0.41
<i>dug</i>	12	214	226	<i>tug</i>	81	181	262	69	-33	36	0.48
<i>goat</i>	18	181	199	<i>coat</i>	79	155	234	61	-26	35	0.43
<i>good</i>	22	196	218	<i>could</i>	94	180	274	72	-16	56	0.22

each of the critical word pairs in the FAST and SLOW speaking rate conditions. The change in each of these measures as a function of initial-consonant voicing (voiceless-voiced), as well as the absolute proportion of change in vowel duration to change in VOT for each of the word pairs, is also shown. The pattern of results for the individual word pairs was highly consistent and comparable to the results of experiment 1. For each of the word pairs in both the FAST and SLOW speaking rate conditions, as the initial consonant changed from voiced to voiceless, VOT increased, whereas vowel duration decreased to a lesser extent. This resulted in a net increase in the CV duration for words beginning with voiceless as opposed to voiced stops. As in experiment 1, the absolute proportion of change in vowel duration to change in VOT was not constant across the word pairs, ranging from 0.19 to 0.53 in the FAST speaking rate condition, and from 0.22 to 0.51 in the SLOW speaking rate condition. Also as in experiment 1, there was no clear effect of phonetic context on the proportion, nor was there consistency in the rank ordering of the proportions for the word pairs across speaking rate conditions or for those pairs that were tested in both experiments.

### C. Discussion

Within both the FAST and SLOW speaking rate conditions, experiment 2 replicated the findings of experiment 1: There was a trade-off between VOT and vowel duration as a function of a change in syllable-initial voicing, in which the increase in VOT is accompanied by a decrease in vowel duration that is smaller in magnitude. In addition, experiment 2 highlights a distinction between two different factors that produce an increase in VOT: a change from a voiced to a voiceless initial stop consonant within a given rate of speech, and a change from a voiceless stop produced at a fast speaking rate to a voiceless stop produced at a slow speaking rate. Although both factors cause VOT to increase, they have opposite effects on the duration of the following vowel. Specifically, as the initial consonant changes from voiced to

voiceless, vowel duration decreases, but as speaking rate changes from fast to slow, vowel duration increases along with other portions of the syllable.

Thus although the traditional method of creating VOT continua overestimates the magnitude of the decrease in vowel duration as VOT increases, the present results provide clear evidence that vowel duration does decrease as VOT increases when the voicing of an initial stop consonant changes from voiced to voiceless. Kessinger and Blumstein's (1998) observation that vowel duration increases as VOT increases was based on data from voiceless stops across changes in speaking rate and, as noted above, this is distinct from what happens when a consonant changes from voiced to voiceless within a particular rate. Because VOT continua are intended to vary from VOT values appropriate for a voiced to a voiceless stop within a particular rate, it would appear that Kessinger and Blumstein's argument against the validity of the traditional method for creating VOT continua is unfounded.<sup>5</sup>

### III. GENERAL DISCUSSION

Taken together, the results of the current investigation indicate the presence of a robust, though imperfect, trade-off between VOT and vowel duration as a function of the voicing of the initial stop consonant: A change in syllable-initial voicing from voiced to voiceless produces an increase in VOT that is accompanied by a decrease in vowel duration, although the decrease in vowel duration is smaller than the increase in VOT. The results also replicate previous findings showing that (for voiceless stops) an increase in VOT due to a slowing of the speaking rate is accompanied by an increase in vowel duration (e.g., Kessinger and Blumstein, 1998; Volaitis and Miller, 1992). These findings illustrate that increases in VOT due to syllable-initial voicing and increases in VOT due to speaking rate are accompanied by opposite effects on the duration of the following vowel. Thus despite recent arguments to the contrary (Kessinger and Blumstein, 1998), which are based on speaking rate manipulations rather



than manipulations in syllable-initial voicing, the tacit assumption of a basic trade-off between VOT and vowel duration that underlies the traditional method of creating VOT continua is confirmed.

However, the traditional method of creating VOT continua does not perfectly reflect the temporal patterns found in natural speech. In the traditional method, increases in VOT are accompanied by equal decreases in vowel duration across the continuum (thereby keeping overall syllable duration constant), whereas in our production data the decrease in vowel duration was consistently smaller than the increase in VOT. This resulted in a net increase in the CV (VOT + vowel duration) portion of the syllable for words produced with voiceless as opposed to voiced stops, and the traditional method of creating VOT continua fails to capture this increase in CV duration. Interestingly, there is an alternative method of creating VOT continua that has occasionally been reported in the literature, in which there is an increase in CV duration along the continuum (e.g., Borsky *et al.*, 1998; Summerfield, 1981). This increase occurs because as VOT is systematically lengthened, vowel duration is kept constant. However, precisely because vowel duration is kept constant, the increase in CV duration is larger in magnitude than that found in natural speech and, of course, this method fails to capture the basic trade-off between VOT and vowel duration that occurs with the change in syllable-initial voicing in natural speech. Thus neither the traditional method nor the alternative method perfectly mimics the observed temporal patterns in the production data, although both capture elements of these patterns.

Future research, which directly compares current methods of generating VOT continua with methods that even more closely approximate the temporal patterns found in speech production, will be required to assess any differential effects of the various methods on the measurement of perceived voicing. On the basis of the current data, however, we conclude that the traditional method of creating VOT continua for perceptual experiments, although not perfect, would appear to be a reasonable approximation of the pattern found in natural speech.

## ACKNOWLEDGMENTS

This research was supported by NIH Grant DC 00130 from the National Institute on Deafness and Other Communication Disorders to the second author and by a Minority Fellowship from the Acoustical Society of America to the first author. The authors thank Peter Eimas and Larry Brancazio for helpful comments on an earlier version of the paper, and thank Sharon Clayman and Jon Pierce for their assistance in conducting the research. Some of the data reported in this paper were presented at the 136th meeting of the Acoustical Society of America, Norfolk, VA, 1998.

<sup>1</sup>“Vowel duration” is operationally defined in the present study, as it is in much of the literature, as the time between the onset of periodicity in the signal (i.e., from the end of the VOT interval) to the offset of vocalic energy. It is important to distinguish this definition of vowel duration from another that is occasionally used in the literature, whereby vowel duration is measured from the release of the initial consonant, subsuming VOT in its measure.

<sup>2</sup>In contrast to the sparse literature on the relation between syllable-initial voicing and following vowel duration, there is considerable literature on vowel duration and postvocalic voicing. Many studies have established that vowels before voiced consonants are generally longer in duration than vowels before voiceless consonants (e.g., Chen, 1970; House, 1961; House and Fairbanks, 1953; Peterson and Lehiste, 1960). Moreover, it is also known that the closure duration of postvocalic voiced stops is shorter than the closure duration of postvocalic voiceless stops (e.g., Lisker, 1957; Port, 1981). Thus the change from a voiced to a voiceless postvocalic stop entails both an increase in closure duration and a decrease in preceding vowel duration. This effect, as well as the trade-off we describe between VOT and following vowel duration, can be seen as part of a larger class of “compensatory” effects whereby the durations of acoustic intervals primarily associated with consonants and vowels vary in a complementary manner—although, as in our data, such compensatory trade-offs are typically not one-to-one. As a result, overall syllable durations are more constant than would be the case without such covariation. Although such compensatory effects have been well documented, the underlying mechanisms responsible for these effects are currently a matter of considerable debate (for discussion see Munhall *et al.*, 1992).

<sup>3</sup>In the Kessinger and Blumstein (1998) study, a positive relation between VOT and vowel duration for syllables beginning with voiceless stops was found not only across their two speaking rate conditions (fast versus slow), but also within a given speaking rate condition. However, the positive relation within a given condition can be understood as being due to small variations in speaking rate for a particular token (e.g., /pi/), analogous to the larger variations that occur across the change from fast to slow speaking rate conditions. We argue that the change in VOT due to speaking rate, whether large or small, should be considered separately from the change in VOT due to a change in syllable-initial voicing (e.g., /bi/ versus /pi/) at a given rate of speech.

<sup>4</sup>Our predictions concerning the effect of speaking rate on VOT refer to voiced versus voiceless stop consonants in English. For a discussion of how the effects of speaking rate on voicing categories pattern across languages, see Kessinger and Blumstein (1997).

<sup>5</sup>Beyond their main argument, Kessinger and Blumstein (1998) point out another potential problem with the traditional method of creating VOT continua. The problem concerns those perceptual studies that include stimuli deliberately created to be poor exemplars, specifically, stimuli whose VOT values are unnaturally long and whose vowel durations are correspondingly unnaturally short (e.g., Flege *et al.*, 1996; Miller and Volaitis, 1989; Volaitis and Miller, 1992). Kessinger and Blumstein argue that these stimuli are not only unnatural, but that the perceived identity of the vowel might be altered due to the very short vowel. For example, /pi/ might sometimes be perceived as /pI/, since the vowel /I/ has a shorter intrinsic duration than the vowel /i/, and at least under some circumstances these vowels can be distinguished on the basis of durational differences alone (Gottfried *et al.*, 1990). However, these two vowels are distinguished not only on the basis of temporal information, but also—and primarily—on the basis of spectral information (Peterson and Barney, 1952). This critical spectral information, i.e., the formant structure of the vowel, exists not only during voicing (during the “vowel” as defined in this study) but also prior to voicing (i.e., during the VOT interval), when the formants are excited by a noise source rather than a periodic source. As a consequence, even in cases where vowel duration (defined as the duration of periodicity) is extremely short, the stimulus is still a /pi/, albeit a poor exemplar of /pi/.

Borsky, S., Tuller, B., and Shapiro, L. P. (1998). “‘How to milk a coat:’ The effects of semantic and acoustic information on phoneme categorization,” *J. Acoust. Soc. Am.* **103**, 2670–2676.

Chen, M. (1970). “Vowel length variation as a function of the voicing of the consonant environment,” *Phonetica* **22**, 129–159.

Crystal, T. H., and House, A. S. (1988). “Segmental durations in connected-speech signals: Current results,” *J. Acoust. Soc. Am.* **83**, 1553–1573.

Diehl, R. L., Souther, A. F., and Convis, C. L. (1980). “Conditions on rate normalization in speech perception,” *Percept. Psychophys.* **27**, 435–443.

Fischer-Jørgensen, E. (1964). “Sound duration and place of articulation,” *Z. Phonet. Sprachwiss. Kommunikationsforsch.* **17**, 175–207.

Flege, J. E., Schmidt, A. M., and Wharton, G. (1996). “Age of learning affects rate-dependent processing of stops in a second language,” *Phonetica* **53**, 143–161.

Ganong, W. F. (1980). “Phonetic categorization in auditory word perception,” *J. Exp. Psychol. Human Percept. Perform.* **6**, 110–125.

- Gottfried, T. L., Miller, J. L., and Payton, P. E. (1990). "Effect of speaking rate on the perception of vowels," *Phonetica* **47**, 155–172.
- Heffner, R-M. S. (1940). "A note on vowel length in American speech," *Language* **16**, 33–47.
- House, A. S. (1961). "On vowel duration in English," *J. Acoust. Soc. Am.* **33**, 1174–1178.
- House, A. S., and Fairbanks, G. (1953). "The influence of consonantal environment upon the secondary acoustical characteristics of vowels," *J. Acoust. Soc. Am.* **25**, 105–113.
- Kessinger, R. H., and Blumstein, S. E. (1997). "Effects of speaking rate on voice-onset time in Thai, French, and English," *J. Phonetics* **25**, 143–168.
- Kessinger, R. H., and Blumstein, S. E. (1998). "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," *J. Phonetics* **26**, 117–128.
- Kluender, K. R. (1991). "Effects of first formant onset properties on voicing judgments result from processes not specific to humans," *J. Acoust. Soc. Am.* **90**, 83–96.
- Kuhl, P. K. (1981). "Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories," *J. Acoust. Soc. Am.* **70**, 340–349.
- Lisker, L. (1957). "Closure duration and the intervocalic voiced-voiceless distinction in English," *Language* **33**, 42–49.
- Lisker, L., and Abramson, A. (1970). "The voicing dimension: Some experiments in comparative phonetics," in *Proceedings of the 6th International Congress of Phonetic Sciences* (Academia, Prague), pp. 563–567.
- Miller, J. L., Green, K. P., and Reeves, A. (1986). "Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast," *Phonetica* **43**, 106–115.
- Miller, J. L., and Volaitis, L. E. (1989). "Effect of speaking rate on the perceptual structure of a phonetic category," *Percept. Psychophys.* **46**, 505–512.
- Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E. (1992). "'Compensatory shortening' in monosyllables of spoken English," *J. Phonetics* **20**, 225–239.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Peterson, G., and Lehiste, I. (1960). "Duration of syllabic nuclei in English," *J. Acoust. Soc. Am.* **32**, 693–703.
- Pind, J. (1995). "Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues," *Percept. Psychophys.* **57**, 291–304.
- Pisoni, D. B., and Lazarus, J. H. (1974). "Categorical and noncategorical modes of speech perception along the voicing continuum," *J. Acoust. Soc. Am.* **55**, 328–333.
- Port, R. F. (1981). "Linguistic timing factors in combination," *J. Acoust. Soc. Am.* **69**, 262–274.
- Rositzke, H. A. (1939). "Vowel-length in general American speech," *Language* **15**, 99–109.
- Samuel, A. G. (1982). "Phonetic prototypes," *Percept. Psychophys.* **31**, 307–314.
- Summerfield, Q. (1981). "Articulatory rate and perceptual constancy in phonetic perception," *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 1074–1095.
- Volaitis, L. E., and Miller, J. L. (1992). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," *J. Acoust. Soc. Am.* **92**, 723–735.

# A model of auditory perception as front end for automatic speech recognition

Jürgen Tchorz and Birger Kollmeier

*Carl von Ossietzky Universität Oldenburg, AG Medizinische Physik, Fachbereich Physik,  
D-26111 Oldenburg, Germany*

(Received 4 November 1998; accepted for publication 22 June 1999)

A front end for automatic speech recognizers is proposed and evaluated which is based on a quantitative model of the “effective” peripheral auditory processing. The model simulates both spectral and temporal properties of sound processing in the auditory system which were found in psychoacoustical and physiological experiments. The robustness of the auditory-based representation of speech was evaluated in speaker-independent, isolated word recognition experiments in different types of additive noise. The results show a higher robustness of the auditory front end in noise, compared to common mel-scale cepstral feature extraction. In a second set of experiments, different processing stages of the auditory front end were modified to study their contribution to robust speech signal representation in detail. The adaptive compression stage which enhances temporal changes of the input signal appeared to be the most important processing stage towards robust speech representation in noise. Low-pass filtering of the fast fluctuating envelope in each frequency band further reduces the influence of noise in the auditory-based representation of speech. © 1999 Acoustical Society of America. [S0001-4966(99)03410-4]

PACS numbers: 43.71.Cq, 43.72.Ne, 43.72.Ar, 43.66.Ba [DO]

## INTRODUCTION

Front ends for automatic speech recognition (ASR) systems are designed to transform the incoming speech signal into a representation which serves as input for later pattern recognition stages. The representation should extract and highlight important features from the speech signal which are relatively independent from speaker variability and channel conditions. It should suppress irrelevant redundancies contained in the speech waveform, thus reducing the data rate at subsequent processing stages of the recognition system. In addition, the representation of speech should be influenced as little as possible by both additive background noise and convolutive distortions (e.g., a change of the transmission channel) to allow for robust recognition in realistic environments outside the laboratory. Unfortunately, the desired robustness against noise is far away from being realized in present speech recognition systems. Even slightly disturbed speech often leads to a distinct decrease in the performance of ASR systems and makes the usefulness of the recognition system questionable.

The human auditory system, on the other hand, performs speech processing which is very robust against noise and allows us to understand speech even under poor conditions. Human speech recognition and understanding is made possible by the interplay between the auditory periphery, which transforms the incoming sound signal into its “internal representation,” and the higher auditory processing stages in the brain, which performs the recognition task based on the internal representation [see Kollmeier (1990) for a review]. While comparatively little is known about the neural mechanisms of the central auditory processing stages in the brain, much more is known about the peripheral auditory processing stages. (Here we use the term periphery to characterize the first stages of auditory processing including filtering of

the basilar membrane, half-wave rectification of the hair cells, and encoding of hair cell movement into auditory nerve firing.) The knowledge about these mechanisms is primarily based on physiological measurements and psychoacoustical experiments on frequency selectivity, loudness perception, short-term adaptation, temporal masking, and other topics.

Despite the progress in understanding auditory processing mechanisms, only a few aspects of sound processing in the auditory periphery are modeled and simulated in common front ends for ASR systems. One example is that most current front ends perform auditory frequency filtering proportional to a perceptually based frequency scale rather than a linear scale, which in general improves recognition robustness in noise (Jankowsky *et al.*, 1995). Another example is the use of the logarithm of certain speech features that approximate the nonlinear dynamic compression in the auditory system which allows us to cover the huge dynamical range between hearing threshold and uncomfortable loudness level.

Ideally, the automatic speech recognizer should operate as a pattern recognizer on the same input pattern that is available to the central auditory system after preprocessing in the peripheral auditory system, as the central auditory system can be regarded as a very powerful recognizer which bases on and makes use of the peripheral representation of sounds. This peripheral representation contains all information that is needed for robust speech recognition. In reality, of course, a technical pattern recognizer works differently than our brain does, and it is questionable whether the technical recognizer can exploit auditorylike features for more robust recognition. Thus, the interactions between feature characteristics and recognizer properties have to be considered carefully. In addition, it is not clear which parts and details of peripheral

auditory processing contribute to robust speech recognition (and thus might be worth being simulated in a technical front end), and which parts are important for other skills of the auditory system (and thus would “whiten” the features with information that is not needed). A more detailed discussion on the simulation of auditory processing in ASR is given by Hermansky (1998).

Having the above-mentioned restrictions in mind, an auditory-motivated representation of speech would include certain properties of auditory processing that have been characterized by physiological and psychoacoustical experiments and have not yet been incorporated in common ASR systems. One example of these properties is short-term adaptation. With short-term adaptation, we denote the combination of two effects: (i) Given a constant stimulus, the auditory nerve response decreases monotonically with increasing stimulus duration, i.e., it adapts and asymptotically approaches a steady-state rate (Kiang *et al.*, 1965). (ii) After stimulus offset, a period of recovery in auditory nerve activity with a firing rate below spontaneous emission in quiet and reduced response to a new stimulus can be observed (Smith, 1979; Delegutte and Kiang, 1984). Another example is the processing of amplitude modulations in the auditory system. During recent years, more insight has been gained about the coding of amplitude modulations in the auditory system and their contribution to pitch representation and speech perception. Langner and Schreiner (1988), for example, found neurons in the inferior colliculus of the cat that were tuned to certain modulation frequencies. Furthermore, a tonotopical organization of units with respect to their modulation tuning was found that appeared to be perpendicular to the tonotopical encoding of the carrier frequencies (Langner, 1992). An enhancement of speech intelligibility could be achieved when these properties were exploited (Kollmeier and Koch, 1994). The importance of temporal amplitude modulations for speech perception was demonstrated by Shannon *et al.* (1995). They observed nearly perfect speech recognition under conditions of greatly reduced spectral information. Their results indicate that amplitude modulations below 50 Hz in particular are important for speech perception.

Several researchers have proposed algorithms to model different psychoacoustical aspects or physiological processing stages of the auditory periphery. Only a few of these models were tested in speech recognition systems, though. Ghitza (1988) introduced a model of temporal discharge patterns of auditory-nerve fibers (ensemble interval histogram—EIH) as feature extraction for automatic speech recognition. In recognition experiments with the TIMIT database, EIH feature extraction showed little robustness improvement when compared to common mel-scale cepstral coefficient front ends (Sandhu and Ghitza, 1995).

Seneff (1988) proposed a bank of filters model of auditory perception which simulates the transformation of basilar membrane motion into auditory nerve firing patterns. In a subsequent processing stage, a physiologically motivated “generalized synchrony detector” measures to which extent a frequency channels firing rate is periodic with the characteristic period  $1/f_c$  of that certain frequency channel. Jan-

kowsky *et al.* (1995) evaluated the robustness of Seneff’s and Ghitza’s auditory front ends in additive and convolutive noise and compared them with a mel frequency filter bank (MFB)-based cepstral front end as control feature extraction. In speaker-dependent, isolated word recognition experiments, almost no difference could be observed between the auditory models and the control front end when speech was degraded by convolutive distortions (e.g., by telephone filtering). In additive noise, the auditory-based front ends appeared to be slightly superior to the control front end. This small advantage had to be paid for with much higher computational effort compared to MFB cepstra. The authors emphasized the necessity of choosing an appropriate control front end when evaluating the robustness of auditory-based front ends. In their experiments, MBF cepstra significantly outperformed linear predictive coding (LPC)-based cepstra used in other investigations which indicated a decent benefit from using auditory preprocessing in noisy environment.

Strope and Alwan (1997) presented a computationally efficient model of dynamic perception which augments common mel frequency cepstrum front ends. An additive logarithmic adaptation stage simulates short-term adaptation. The model’s parameters were fitted to predict psychoacoustical forward masking experiments. A subsequent peak isolation mechanism is intended to further enhance the dynamic spectral cues. First evaluations in speaker-independent, isolated digit recognition experiments in static additive noise yielded better results compared to common front ends.

An earlier approach to simulate short-term adaptation and to implement it in an ASR front end was presented by Cohen (1989). He combined critical band filtering with loudness scaling and a reservoir-type adaptation equation for simulating the hair cell action (Schroeder and Hall, 1974), which relates stimulus intensity to auditory-nerve firing rate. He reported lower error rates and reduced computational costs in large-vocabulary, connected word recognition experiments when compared to a standard bank of filters front end. Noise immunity was not addressed in his experiments.

In recent years, growing attention is paid to low-frequency amplitude modulations. RASTA processing of speech (Hermansky and Morgan, 1994) filters the time trajectories of speech. The RASTA filter passes amplitude modulation components between about 1 and 12 Hz in a relatively flat pass band and with rather steep slopes. Ideally, by appropriate feature transformation, the disturbing components in the input signal should combine linearly with the components that origin from speech, so that these components can be separated by RASTA filtering. The log-RASTA approach uses static logarithmic compression prior to modulation filtering. It is primarily intended to remove convolutive distortions (e.g., due to changes in the transmission channel), as these distortions are additive in the log-domain. The adaptive J-RASTA approach is intended to suppress the influence of additive noise. Here, the compression prior to modulation filtering depends on an estimation of the present noise energy, which is measured in speech-free intervals. When applied as front end for ASR systems, considerable increase of robustness compared to standard front ends could be observed (see, e.g., Kasper *et al.*, 1997).



The outline of this paper is to describe the application of a psychoacoustically motivated model of auditory perception to automatic speech recognition. A set of experiments in different types of noise evaluates the robustness of the auditory feature extraction compared to a standard front end. A subsequent analysis of single auditory model processing steps is intended to explore their contribution to robust recognition and to answer the question whether the parameters of the psychoacoustical model are optimal for the new task of ASR feature extraction.

## I. AUDITORY MODEL-BASED PREPROCESSING

The intention of the quantitative model of auditory processing is to transform an incoming sound waveform into its "internal" representation. Rather than trying to model each physiological detail of auditory processing, the approach is to focus on the "effective" signal processing in the auditory system which uses as little physiological assumptions and physical parameters as necessary, but to predict as many psychoacoustical aspects and effects as possible.

The model was originally developed for describing human performance in typical psychoacoustical spectral and temporal masking experiments, e.g., predicting the thresholds in backward, simultaneous, and forward-masking experiments (Dau *et al.*, 1996a, b). The parameters of the model were chosen to fit these experiments. Gap detection and modulation detection experiments (Dau *et al.*, 1997a, b) were simulated with a combination of the model with a subsequent modulation filterbank. In the field of speech processing, the auditory model was applied to objective speech quality measurement (Hansen and Kollmeier, 1997, 1998), speech intelligibility prediction in noise (Wesselkamp, 1994) and in hearing impaired listeners (Holube and Kollmeier, 1996).

### A. Processing steps

A block diagram of the auditory model and its processing stages is shown in Fig. 1. The first processing step of the auditory model is a preemphasis of the input signal with a first-order differentiation. This flattens the typical spectral tilt of speech signals and reflects the transfer function of the outer ear (Djupesland and Zwislocki, 1972). The preemphasis was introduced for feature extraction in speech recognition systems and was not used in previous applications of the auditory model. In the next processing stage, the preemphasized signal is filtered by a gammatone filterbank (Patterson *et al.*, 1987) using 19 frequency channels equally spaced on the equivalent regular bandwidth (ERB) scale with center frequencies ranging from 300 to 4000 Hz. The impulse responses of the gammatone filterbank are similar to the impulse responses of the auditory system found in physiological measurements (de Boer and Kruidenier, 1990). The implemented gammatone filterbank is linear. It does not consider nonlinear effects such as level-dependent upward spread of masking and combination tones. After gammatone filtering, each frequency channel is half-wave rectified and first-order low-pass filtered with a cutoff frequency of 1000 Hz for envelope extraction. Information about the fine struc-

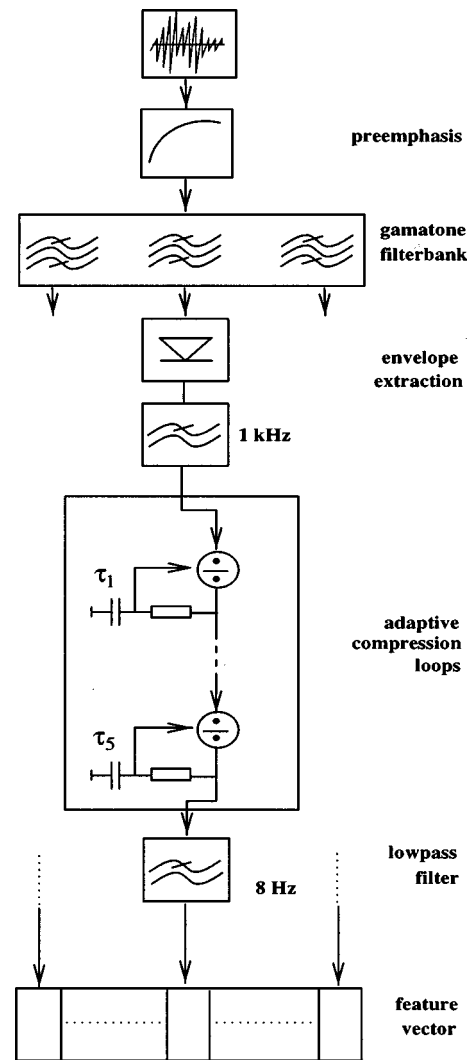


FIG. 1. Processing stages of the auditory model.

ture of the signal at high frequencies gets lost; this reflects the limiting phase locking for auditory nerve fibers above 1000 Hz. At this stage of processing, each frequency channel contains information about the amplitude magnitude of the input signal within the channel. Amplitude compression is performed in a following processing step. In contrast to a conventional bank of filters front ends, the amplitude compression of the auditory model is not static (e.g., logarithmic) but adaptive, which is realized by an adaptation circuit consisting of five consecutive nonlinear adaptation loops (Püschel, 1988). Each of these loops consists of a divider and a low-pass filter with time constants  $\tau_1 = 5$  ms,  $\tau_2 = 50$  ms,  $\tau_3 = 129$  ms,  $\tau_4 = 253$  ms, and  $\tau_5 = 500$  ms (Dau *et al.*, 1996a, 1997a). For each adaptation loop, the input signal is divided by the output signal of the low-pass filter. Thus, a stationary input signal  $X$  is transformed to an output signal  $Y = \sqrt{X}$ . The output of five consecutive adaptation loops is then  $Y = \sqrt[32]{X}$ , which approximates the logarithm of the input signal. Fluctuations of the input signal that are very fast compared to the time constants of the adaptation loops, on the other hand, are transformed linearly because of time-delayed denominator changes. Due to this transformation characteristic, changes in the input signal like onsets and

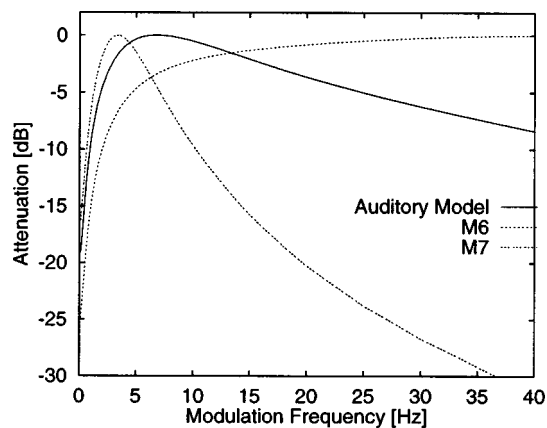


FIG. 2. Amplitude modulation transfer function of the auditory model and of modifications M6 and M7 as described in Sec. III A. The modulations in the output of the channel with a center frequency of 1 kHz were compared with the input modulations for a sinusoidally amplitude-modulated sinusoid at 1 kHz as a function of modulation frequency.

offsets are emphasized, whereas steady-state portions are compressed. Thus, the adaptation loops can be regarded as an inherent memory of the system. Given an input signal, the output signal of the auditory model at time  $t$  does not only depend on the properties of the input signal within a narrow time frame around  $t$ , but also on the characteristics of the input signal during a preceding time interval of about 200 ms. Due to the inherent memory of the auditory model, the dynamical structure of the input signal is taken into account over a relatively long period of time. Short-term adaptation including enhancement of changes and temporal integration is simulated and allows a quantitative prediction of important temporal effects in auditory perception, such as backward and forward masking (Dau *et al.*, 1996b). The last processing step of the auditory model is a first-order low-pass filter with a cutoff frequency of 8 Hz. The filter was introduced by Dau *et al.* (1996b) to optimize predictions of psychoacoustical masking experiments. It attenuates fast envelope fluctuations of the signal in each frequency channel.

The output of the auditory preprocessing is down-sampled to a rate of 100 feature vectors per second to serve as input for the subsequent recognition device. The auditory model is implemented in C code. Processing takes about one-third real time on a current standard personal computer.

## B. Modulation filtering of the auditory model

Suppression of very slow envelope fluctuations by the adaptation loops and attenuation of fast fluctuations by the low-pass filter results in a bandpass characteristic of the amplitude modulation transfer function of the auditory model. We measured the modulation transfer function using a constant carrier (i.e., a sinusoid at 1 kHz) which was sinusoidal amplitude modulated at different modulation frequencies with a modulation depth of 20%. The attenuation of the modulated signal in the corresponding frequency channel after processing with the auditory model was measured for each modulation frequency. The resulting amplitude modulation transfer function is plotted in Fig. 2 (solid curve).<sup>1</sup> The maximum amplitude modulation transmission of the model

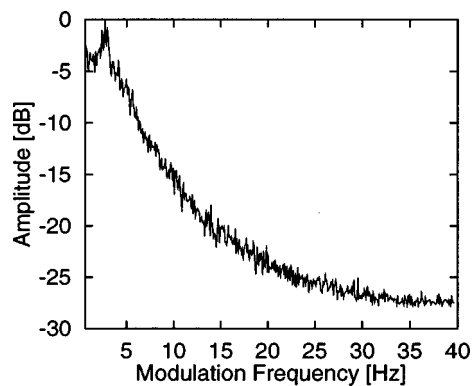


FIG. 3. Broadband modulation spectrum of a speech sample.

can be found at modulation frequencies around 6 Hz. There is a strong attenuation of modulation frequencies below 2 Hz due to the steady-state compression of the adaptation loops. In the high-frequency part, the transfer function declines with approximately 3 dB/oct.

A broadband modulation spectrum of speech is shown in Fig. 3. It was generated from 380 s of speech from 140 different American speakers from the TIMIT database (70 male, 70 female). The envelope was extracted from the waveform using half-wave rectification and low-pass filtering using a second-order Butterworth filter with a cutoff frequency of 100 Hz. The peak around 3 Hz in the modulation spectrum is a characteristic feature of speech which originates from the average syllable rate (Houtgast and Steeneken, 1985). It can be seen that the modulation transfer function of the auditory model resembles, but does not exactly fit the average modulation spectrum of speech in terms of most prominent modulation frequency and decline towards higher modulation frequencies.

## C. Examples of sound and speech processing

As a first example, the response of the auditory preprocessing when stimulated with a pure tone is demonstrated. Figure 4 (first panel, left) shows the envelope of the stimulus (a 1000-Hz tone with a duration of 500 ms). In the second panel (left), the output of the corresponding frequency channel after processing is shown. The figure illustrates the simulation of short-term adaptation, including initial linear response at signal onset, transition to steady-state compression, “undershoot” after signal offset due to discharged adaptation loops, and recovery time.

An example of speech processing is shown in Fig. 5. The waveforms of an undisturbed word and that of the same word disturbed by additive speech simulating noise at 10 dB SNR are plotted on top. Their corresponding representations after processing with the auditory front end can be seen in the two panels in the second row. Peaks in the internal representations are indicated by bright spots, low feature values by dark areas. The two panels in the third row show one single frequency channel (with center frequency  $f_c = 780$  Hz) after processing for a more detailed study. The relatively stationary noise preceding and following the speech signal is suppressed due to the steady-state compression performed by the adaptation loops. The onsets and off-

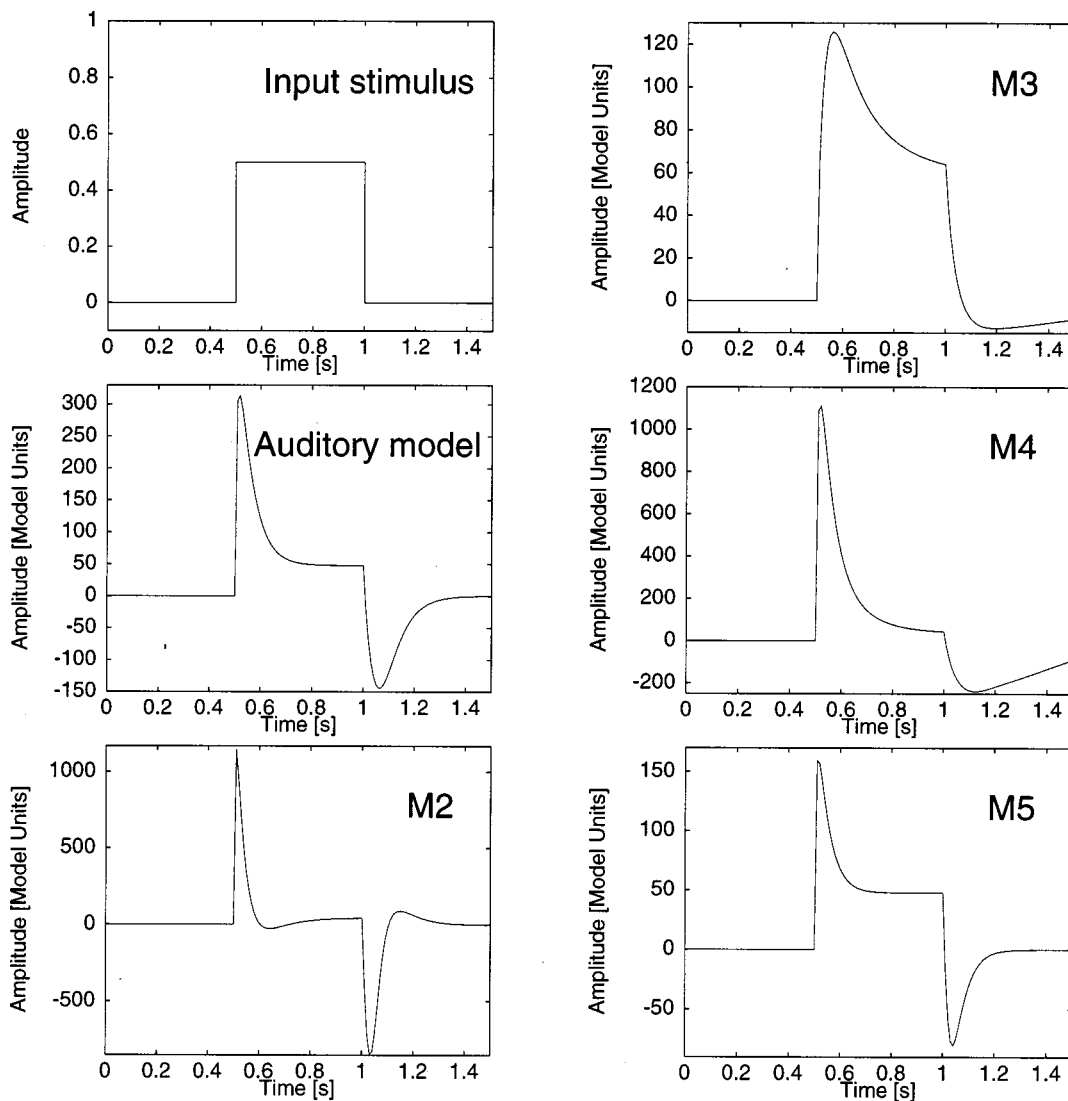


FIG. 4. Simulation of short term adaptation in the auditory model. First panel (left): envelope of the input stimulus, a 1000-Hz 500-ms tone pulse. Second panel (left): response of the auditory model in the frequency channel corresponding to the stimulus with initial linear response, transition to steady-state compression, and period of recovery after stimulus offset. Other panels: responses of modifications M2–M5 of the auditory model on the stimulus. See Sec. III A. (Note that the scaling of the y axis is different for each figure.)

sets of the speech signal are enhanced in both quiet and in noise. Due to the adaptive amplitude compression which emphasizes changes and suppresses constant portions, speech encoding of the model can be described as sparse and distinct.

To visualize the difference between adaptive compression and static compression, the two panels at the bottom of Fig. 5 show the representation for the same input signals and the same frequency channel with the adaptive compression stage replaced by a static log-compression. Here, the most prominent parts of the speech representation are maintained, but the “floor” of the representation during nonspeech portions is shifted due to the background noise.

## II. RECOGNITION EXPERIMENTS WITH THE AUDITORY MODEL

A number of speaker-independent, isolated digit recognition experiments in different types of additive noise were performed to evaluate the robustness of the auditory-based representation of speech quantitatively.

### A. Experimental setup

The speech material for training of the word models and scoring was taken from the ZIFKOM database of Deutsche Telekom AG. Each German digit was spoken once by 200 different speakers (100 males, 100 females). The recording sessions took place in soundproof booths or quiet offices. The speech material was sampled at 16 kHz.

To cover a certain range of possible additive distortions of speech in actual ASR systems, three different types of noise were added to the speech material at different signal-to-noise ratios before feature extraction: white noise (WN), speech-simulating noise (SN), which was generated from a random superposition of words spoken by a male speaker, and background noise recorded on a construction site (CS). The first two noise types (WN and SN) are stationary, i.e., their spectral shape and energy do not change over time. The last noise (CS) exhibits fluctuations in both spectral shape and energy. To prepare the additive distortions, the rms value of each word utterance including short pauses before and

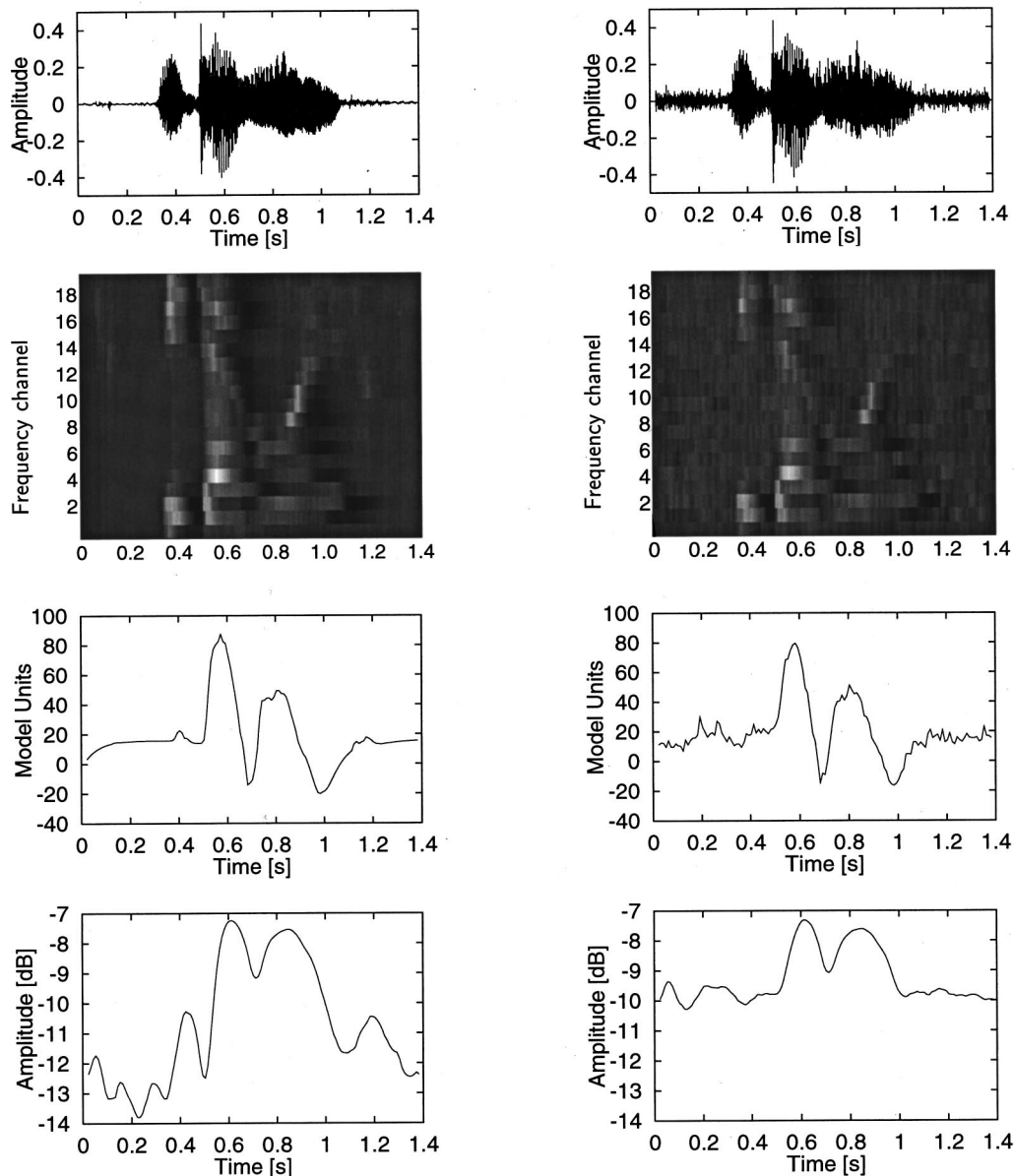


FIG. 5. An example for speech processing performed by the auditory model. On top, the waveforms of an utterance of the German word “wiederholen” from a female speaker in quiet (left) and in speech simulating noise at 10 dB SNR (right) are shown. In the second row, the corresponding “internal” representations after preprocessing are visualized. High and low amplitudes are indicated by bright and dark areas, respectively. In the third row, the output of one single-frequency channel (center frequency  $c_f = 780$  Hz) is shown. It can be seen how onsets and offsets are contrasted by the adaptive compression stage. The relatively stationary background noise is compressed and causes only minor fluctuations in the representation. The last two pictures, in contrast, show the respective representations for the same input signals and the same frequency channel when the adaptive compression stage was replaced by a static log-compression. Here, the post-prominent parts of the speech representation are well maintained, but the “floor” of the representation during nonspeech portions is shifted due to the background noise.

after the utterance was calculated separately. The background noises were scaled and added to the utterances with signal-to-noise ratios of 20, 15, 10, and 5 dB.

For training and testing, we used a standard continuous-density Hidden Markov model (HMM) recognizer with five Gaussian mixtures per state, diagonal covariance matrices, and six emitting states per word model. The word models were trained with features from 100 undisturbed utterances of each digit. Features for testing were calculated from another 100 utterances of each digit which were distorted by additive noise before preprocessing. As control front end we used mel frequency cepstral coefficients (MFCC, Davis and Mermelstein, 1980), which are widely used in common ASR

systems. The coefficients were calculated from Hamming-windowed, preemphasized 32-ms segments of the input signal with a frame period of 10 ms. In our experiments, each mel cepstrum feature vector contained 26 features (12 coefficients, log energy, and the respective first temporal derivatives as additional delta-features).

## B. Results

The speaker-independent digit recognition rates in clean speech and in additive noise obtained with the auditory preprocessing and the control front end are shown in Fig. 6. The



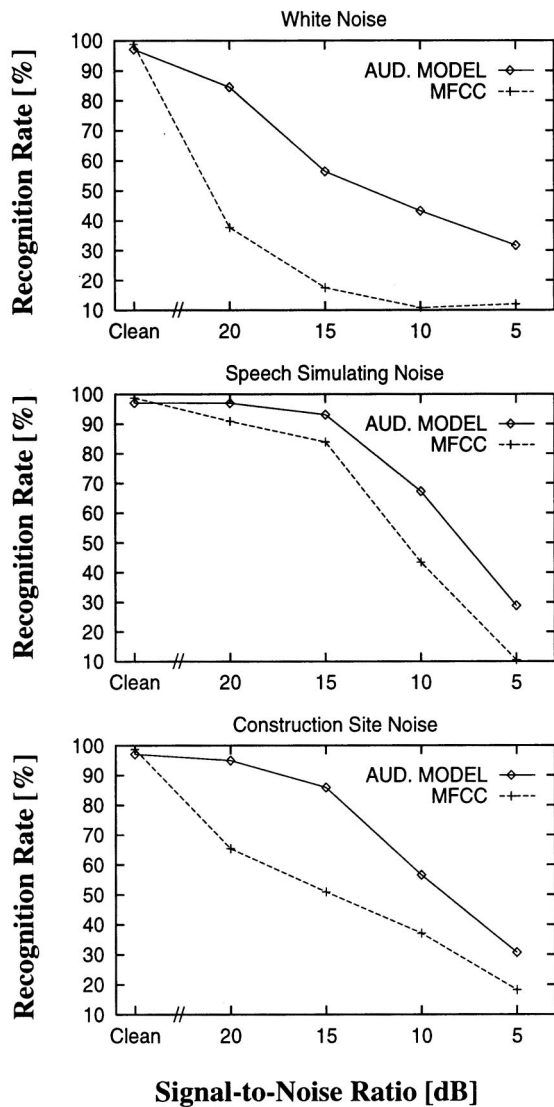


FIG. 6. Recognition rates in % in different types of additive noise yielded with the auditory model and with the control front end (mel-scale cepstral coefficients, MFCC) as function of signal-to-noise ratio in dB.

recognition rates in percent are plotted as a function of the signal-to-noise ratio in dB. The three panels show the results for white noise, speech simulating noise, and construction site noise, respectively. In undisturbed speech, the control front end yields a higher recognition rate (98.8%) than the auditory-based front end (97.1%). In additive noise, however, the auditory features are more robust than those of the control front end. Even in only slightly disturbed speech (20 dB SNR), the recognition rates obtained with the auditory model are significantly higher in all tested types of noise. The largest difference between the two front ends occur in white noise, where the error rate is decreased by a factor of 3, approximately. In construction site noise, small additive distortions lead to severely decreased recognition rates with the control front end, whereas the auditory model allows more robust recognition. In speech simulating noise, the difference in performance between the two front ends is smaller, but still distinct.

### III. RECOGNITION EXPERIMENTS WITH MODIFICATIONS OF THE AUDITORY MODEL

The preceding section showed that the auditory front end allows promising recognition rates in different types of noise. Two questions arise at this point: First, how do the different signal processing stages of the auditory model contribute to robust recognition? Second, are the parameters of the model, which were previously determined to fit psychoacoustical experiments, optimal for the new task in the field of speech processing? To answer these questions, we performed a number of recognition experiments with modifications of the auditory model.

#### A. Modifications

There are two processing steps which dominate the representation of speech performed by the auditory model compared to standard bank of filters front ends. First, the nonlinear adaptation loops realize effects of short-term adaptation on the incoming signal. Steady-state portions are suppressed, whereas fast fluctuations are transformed without attenuation. Second, the 8-Hz low-pass filter smoothes the fast fluctuating envelope and leads to a bandpass modulation transfer function. These two processing steps were altered in different ways to analyze their respective contribution to robust representation of speech. In total, seven modified versions (M1–M7) of the auditory front end were implemented and tested.

In modifications M1–M4, the adaptive amplitude compression stage of the model was changed. To determine whether adaptive compression contributes to robust recognition at all, the five adaptation loops were replaced by static logarithmic compression (M1). Dynamic temporal properties of speech such as onsets and offsets are then no longer emphasized in the resulting feature vectors (see the bottom panels of Fig. 5). In modification M2, the number of adaptation loops was increased from five to eight. The time constants  $\tau_1, \dots, \tau_8$  were linearly equispaced between 5 and 500 ms. Thus, temporal contrasts in the input signal are further enhanced in the representation. On the other hand, the amplitude  $X$  of a steady-state portion is compressed to  $^{25}\sqrt{X}$ , i.e., there is almost no contribution of constant signal portions left in the representation.

For modification M3, the number of adaptation loops was decreased from 5 to 2, i.e., onsets and offsets are only moderately emphasized. Steady-state portions are compressed to  $\sqrt[4]{x}$ , and the time constants are  $\tau_1 = 5$  ms and  $\tau_2 = 500$  ms.

In modification M4, the number of adaptation loops was five, as in the original model. Their time constants were increased and set to  $\tau_1 = 50$  ms,  $\tau_2 = 300$  ms,  $\tau_3 = 700$  ms,  $\tau_4 = 1100$  ms, and  $\tau_5 = 1600$  ms. This leads to a slower transition to steady-state compression. Only rather long steady-state portions are fully compressed to  $\sqrt[3]{x}$ ; the attenuation of faster modulation frequencies is weaker than in the original model.

Shorter time constants as in the original model were chosen in modification M5. They were set to  $\tau_1 = 2$  ms,  $\tau_2 = 20$  ms,  $\tau_3 = 50$  ms,  $\tau_4 = 100$  ms, and  $\tau_5 = 200$  ms. In this case, the transition to full steady-state compression is faster,

TABLE I. Modifications of the auditory model at a glance.

M1	logarithmic compression
M2	eight adaptation loops
M3	two adaptation loops
M4	loops with longer time constants
M5	loops with shorter time constants
M6	no modulation low-pass filtering
M7	enhanced modulation low-pass filtering

and the peaks in the representation indicating onsets and offsets are narrower.

The different compression schemes of the modified auditory front ends M2–M5 are illustrated in Fig. 4, where their respective tone pulse responses are shown.

The modified versions M6 and M7 concern the 8-Hz low-pass filter at the end of the auditory preprocessing. In modification M6, the filter was simply left out. This leads to a high-pass amplitude modulation transfer function. High modulation frequencies are no longer attenuated; low modulation frequencies are damped by the steady-state compression of the adaptation loops. In M7, the 8-Hz low-pass filter was replaced by a second-order low pass with a cutoff frequency of 4 Hz. Fast envelope fluctuations are almost fully compressed.

The modulation transfer functions of the two modified auditory front ends M6 and M7, compared to the original preprocessing, are shown in Fig. 2. Without low-pass filtering of the envelope (M6), the modulation transfer function has a high-pass characteristic due to the adaptive compression stage of the model. With the original 8-Hz first-order low-pass filter, the decline in the high-frequency part is not 6 dB/oct, as one could expect, but less, because the filter has to compensate for the increase in the transfer function towards high modulation frequencies without filtering. With increased low-pass filtering of the envelope (M7), the decline in the modulation transfer function approximates the average modulation spectrum of speech in the low modulation frequency part (see Fig. 3). The modifications of the auditory model at a glance are listed in Table I.

Speaker-independent, isolated-digit recognition experiments for clean speech and in additive noise were performed with all modifications. The setup of the experiments was the same as described in Sec. II A.

## B. Results

The recognition rates obtained with modifications M1–M7 of the auditory model are plotted in Fig. 7. Replacing the adaptation loops by static logarithmic compression (M1) yields satisfactory results in clean speech, but the recognition rate drops very fast even if the signal-to-noise ratio is as high as 20 dB SNR. This indicates that adaptive compression of the envelope in each frequency channel plays an essential role for robust speech recognition with the auditory model.

Changing the parameters of the adaptation loops appears to be disadvantageous in most situations. Taking eight adaptation loops instead of five (M2) and thus suppressing steady-state information almost always leads to increased er-

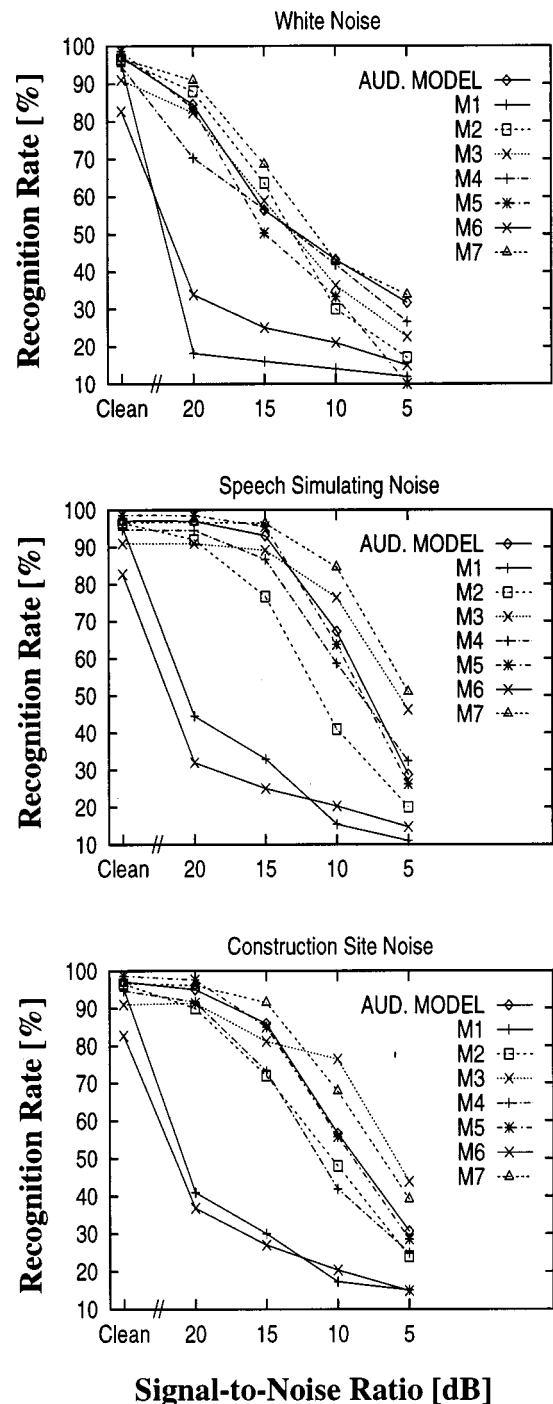


FIG. 7. Recognition rates in % in different types of additive noise as function of signal-to-noise ratio in dB obtained with the original auditory model and with its modifications M1–M7 (see text).

ror rates, especially in low signal-to-noise ratios. With only two adaptation loops (M3), performance in moderate noise is worse compared to the original model. In low signal-to-noise ratios, however, higher recognition rates can be observed in speech simulating noise and construction site noise. The effect of taking longer or shorter time constants for the low-pass filters within the adaptation loops (M4 and M5, respectively) is limited. Longer time constants lead to slightly increased error rates in most situations, whereas shorter time constants yield similar results compared to the original model, except for the case of white noise, where the perfor-

mance of the original model is a bit better in low signal-to-noise ratios.

In summary, the original parameters of the adaptive compression stage, which were optimized for predicting psychoacoustical experiments, appear to be well suited for application in ASR systems. They seem to represent a broad optimum in terms of robust representation of speech in noise, and no distinct improvement of performance in all signal-to-noise ratios could be found when modifying them. The results indicate that there has to be some adaptive compression *at all* to improve robustness in noise, as can be seen from the performance of M1. Without adaptive compression, recognition rates are satisfactory in clean speech, but drop quite fast in noise.

Low-pass filtering of the envelope after compression in each frequency channel also plays an important role for robust representation of speech, as the recognition rates obtained for modifications M6 and M7 of the auditory front end show. Without filtering (M6), only poor recognition rates in noise are yielded. Stronger attenuation of fast amplitude modulations (M7), on the other hand, leads to decent improvement of robustness in construction site noise and speech simulating noise. In clean speech, however, M7 results in a slightly degraded recognition rate (96.7%), compared to the original processing (97.1%). In contrast to the adaptive compression stage, the psychoacoustically motivated low-pass filter parameter at the end of the auditory preprocessing seems not to be optimal for speech processing. In our experiments, stronger attenuation of higher modulation frequencies (which better reflects the modulation spectrum of speech) enhanced robustness of speech recognition in noise, i.e., ASR performance benefits from variation of the original psychoacoustically motivated low-pass filter.

#### IV. DISCUSSION

The main findings in this study can be summarized as follows:

- (1) The presented auditory-based ASR front end allows more robust speaker-independent digit recognition compared to standard feature extraction in additive noise, even when the parameters of the psychoacoustically motivated model were not refitted for the new task in the field of speech processing.
- (2) An important processing step of the auditory model for robust representation of speech in additive noise is the adaptive compression stage, which encodes the dynamic evolution of the input signal and allows simulation of temporal aspects of processing found in the auditory system.
- (3) Modulation bandpass filtering centered at low modulation frequencies plays an important role for robust speech representation. Changing the original psychoacoustical filter parameters to better reflect the average modulation spectrum of speech further enhances the robustness of isolated digit recognition in noise.

Encoding the dynamic evolution of the input signal as realized by the auditory model emphasizes changes in the

input signal relative to constant portions. The signal representation of the auditory model is characterized by sparse and distinct peaks. These peaks are well maintained in additive noise and seem to serve as quite robust cues for the recognizer. A study on the interplay between auditory-based features and a recurrent neural network as recognizer (Tchorz *et al.*, 1997) supports this suggestion. It was shown that the sparse and distinct peaks contain sufficient information for robust recognition with neural networks. Even if 80% of all feature values (i.e., those which did not exceed a certain threshold) were set to zero, a digit recognition rate of more than 90% was reached on these strongly reduced features. Thus, the experiments presented here demonstrate that the consideration of short-term adaptation as in the auditory model might be fruitful for reducing the influence of additive noise in future ASR front ends. Findings from other researchers support this hypothesis (Strope and Alwan, 1997).

In human speech perception, analysis of low modulation frequencies plays an important role. In a study on the intelligibility of temporally smeared speech, Drullman *et al.* (1994) found that modulation frequencies below 8 Hz are the most important ones for speech intelligibility. Components between 8 and 16 Hz were found useful, too, whereas rates above 16 Hz are not required for speech intelligibility.

The 8-Hz low-pass filter at the end of the auditory model preprocessing, which leads to a bandpass modulation transfer function of the model, was introduced by Dau *et al.* (1996b) to optimize predictions of psychoacoustical masking experiments. When applied as ASR front end, the bandpass characteristic attenuates nonspeech sources in disturbed speech. Further reduction of the filter's cutoff frequency allows for higher digit recognition rates in noise (which might be explained by the better correspondence with the average modulation spectrum of speech), but also slightly degraded the performance in clean speech.

Kanedera *et al.* (1997) measured the effects of bandpass filtering of the time trajectories of spectral envelopes on speech recognition. Their results showed that most of the useful linguistic information for ASR is in modulation frequency components in the range from 1 to 16 Hz, with the dominant component around 4 Hz. In noisy environment, the range below 2 Hz and above 16 Hz sometimes degraded the recognition accuracy.

In isolated word recognition, there seems to be a tradeoff concerning the modulation frequencies above about 4 Hz. In clean speech, they carry useful information which can be exploited by the recognizer and thus should be passed through. In noise, however, the restriction to the most prominent modulation frequencies of speech helps to enhance robustness.

In the field of subword unit recognition (e.g., phonemes), however, strong attenuation of higher modulation frequencies might lead to problems even in clean speech. Transitions between short segments are blurred by low-pass filtering of the envelope. The feature representation of a single phoneme strongly depends on its temporal context then, and might make it difficult to train and recognize subword units.

The auditory model described in this paper exhibits



similar features as RASTA processing of speech (Hermansky *et al.*, 1994) with respect to representation of temporal properties of the incoming signal. Both techniques perform some kind of envelope bandpass filtering around 4 Hz and hence take roughly 200 ms of “signal history” into account for feature calculation. This is in clear contrast to common short-term acoustic features, which represent independent 10–20-ms frames. Temporal processing in the auditory model is essential to quantitatively predict psychoacoustical masking experiments. While the auditory model used here was primarily developed to model a variety of temporal psychoacoustical effects, RASTA was also shown to model certain aspects of forward masking experiments (Hermansky and Pavel, 1998).

Kasper *et al.* (1997) compared the robustness of ASR systems with four different types of feature vectors, namely cepstral coefficients, features from the auditory model as described in this paper, and log-RASTA and J-RASTA coefficients (Hermansky and Morgan, 1994) in isolated-digit recognition experiments. In a first set of experiments, a HMM recognizer was used. The poorest recognition rates in noise (either additive noise or convolutive distortions of the test material) were measured with cepstral coefficients. Log-RASTA processing performed slightly better than the auditory model front end. Adaptive J-RASTA processing allowed for the highest recognition rates in almost all conditions. In a second set of experiments, a locally recurrent neural network (Kasper *et al.*, 1997) was used as recognizer instead of the HMM recognizer. This led to a further decrease in robustness with cepstral coefficients as features. The results for RASTA processing were almost not affected by the different recognizer. The recognition rates for auditory model features, however, significantly increased with the neural network recognizer and were comparable to those obtained with adaptive J-RASTA processing (without requiring explicit speech-free portions for noise power estimation, as J-RASTA). Obviously, the neural network better exploits the characteristics of the auditory model features for robust recognition than the HMM recognizer does.

In a recent study, Kasper and Reininger (1999) further investigated these recognizer dependencies. They transformed the original feature vectors of the auditory model into their cepstra prior to HMM training and testing. With these modified features, the authors reported much improved recognition rates in both additive noise and convolutive distortions, compared to the original features. These results demonstrate that the original auditory model features are only partly suited for a HMM recognition framework with diagonal covariance matrices, as the feature values are partly correlated across frequency channels. For optimal performance in actual recognition systems, a careful adjustment between the front end and the recognizer is necessary. This point holds for most nonstandard approaches to feature extraction and was discussed for different types of front ends by Bourlard *et al.* (1996).

## V. CONCLUSION

The psychoacoustically motivated auditory model which was originally developed to describe human performance in

typical psychoacoustical spectral and temporal masking experiments yields promising results when applied as front end to ASR systems, especially in noisy environment. In the model, an appropriate temporal processing in each frequency channel of the auditory model plays an important role for robust representation of speech.

To further evaluate the potential of the auditory model in speech recognition systems, experiments with large word vocabularies as well as recognition experiments basing on subword units are necessary.

## ACKNOWLEDGMENTS

This work was supported by Deutsche Forschungsgemeinschaft (Ko 942). Many thanks to Torsten Dau, Martin Hansen, Matthias Wesselkamp, and all other members of the AG Medizinische Physik for support and fruitful discussions, and to Klaus Kasper and Herbert Reininger from Frankfurt University for close cooperation and experiments with the auditory model. We also thank three anonymous reviewers for their comments and suggestions, which definitely helped to improve the paper.

<sup>1</sup>Note that due to the nonlinearity of the adaptation loops, the effective modulation transfer function (MTF) obtained from the model depends strongly on the modulation spectrum of the input signal. Hence, the characteristic plotted in Fig. 2 is only an approximation of the MTF that may be obtained with speech signals.

- Bourlard, H., Hermansky, H., and Morgan, N. (1996). “Towards increasing speech recognition error rates,” *Speech Commun.* **18**, 205–231.
- Cohen, J. R. (1989). “Application of an auditory model to speech recognition,” *J. Acoust. Soc. Am.* **85**, 2623–2629.
- Dau, T., Püschel, D., and Kohlrausch, A. (1996a). “A quantitative model of the effective signal processing in the auditory system: I. Model structure,” *J. Acoust. Soc. Am.* **99**, 3615–3622.
- Dau, T., Püschel, D., and Kohlrausch, A. (1996b). “A quantitative model of the ‘effective’ signal processing in the auditory system: II. Simulations and measurements,” *J. Acoust. Soc. Am.* **99**, 3633–3631.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). “Modeling auditory processing of amplitude modulation: I. Modulation 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.
- Davis, K., and Mermelstein, P. (1980). “Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences,” *IEEE Trans. Acoust., Speech, Signal Process.* **28**(4), 357–366.
- de Boer, E., and Kruidenier, C. (1990). “On ringing limits of the auditory periphery,” *Biol. Cybern.* **63**, 433–442.
- Delgutte, B., and Kiang, N. Y. S. (1984). “Speech coding in the auditory nerve: IV. Sounds with consonantlike dynamic characteristics,” *J. Acoust. Soc. Am.* **75**, 897–907.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). “Effect of temporal envelope smearing on speech reception,” *J. Acoust. Soc. Am.* **95**, 1053–1064.
- Djupesland, G., and Zwislocki, J. J. (1972). “Sound pressure distribution in the outer ear,” *Scand. Audiol.* **1**, 197–203.
- Ghitza, O. (1988). “Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment,” *J. Phonetics* **16**, 109–123.
- Hansen, M., and Kollmeier, B. (1998). “Objective modeling of speech quality with a psychoacoustically validated auditory model,” *J. Audio Eng. Soc.* (submitted).
- Hansen, M., and Kollmeier, B. (1997). “Using a quantitative psychoacoustical signal representation for objective speech quality measurement,” *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* **2**, 1387–1391.



- Hermansky, H. (1998). "Should recognizers have ears?" *Speech Commun.* **25**, 3–24.
- Hermansky, H., and Morgan, N. (1994). "RASTA processing of speech," *IEEE Trans. Speech Audio Process.* **2**, 578–589.
- Hermansky, H., and Pavel, M. (1998). "RASTA model and forward masking," *Proc. NATO/ASI Conference on Computational Hearing, II Ciocco, Italy*, pp. 157–162.
- Holube, I., and Kollmeier, B. (1996). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," *J. Acoust. Soc. Am.* **100**, 1703–1716.
- Houtgast, T., and Steeneken, H. J. M. (1985). "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.
- Jankowski, C. R., Vo, H-D. H., and Lippmann, R. P. (1995). "A comparison of signal processing front ends for automatic word recognition," *IEEE Trans. Speech Audio Process.* **3**, 286–293.
- Kanedera, N., Arai, T., Hermansky, H., and Pavel, M., (1997). "On the importance of various modulation frequencies for speech recognition," *Proc. EUROSPEECH '97* **3**, 1079–1082.
- Kasper, K., and Reininger, R. (1999). "Evaluation of PEMO in robust speech recognition," paper presented at the ASA/EAA/DEGA Joint Meeting on Acoustics, Berlin, Germany.
- Kasper, K., Reininger, R., and Wolf, D. (1997). "Exploiting the potential of auditory preprocessing for robust speech recognition by locally recurrent neural networks," *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* **2**, 1223–1227.
- Kiang, N. Y. S., Watanabe, T., Thomas, E. C., Clark, L. F. (1965). "Discharge patterns of single fibers in the cat's auditory nerve" *Res. Monogr.* **35** (MIT, Cambridge, MA).
- Kollmeier, B. (1990). "Messmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache," *Habilitationsschrift, Universität Göttingen*.
- Kollmeier, B., and Koch, R. (1994). "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," *J. Acoust. Soc. Am.* **95**, 1593–1602.
- Langner, G. (1992). "Periodicity coding in the auditory system," *Hearing Res.* **60**, 115–142.
- Langner, G., and Schreiner, C. E. (1988). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," *J. Neurophysiol.* **60**, 1799–1822.
- Patterson, R. D., Nimmo-Smith, J., Holdsworth, J., and Rice, P. (1987). "An efficiency auditory filterbank based on the gammatone function," paper presented at a meeting of the IOC Speech Group on Auditory Modeling at RSRE, 14–15 December.
- Püschel, D. (1988). "Prinzipien der zeitlichen Analyse beim Hören," *Doctoral thesis, Universität Göttingen*.
- Sandhu, S., and Ghitza, O. (1995). "A comparative study of mel cepstra and EIH for phone classification under adverse conditions," *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* **1**, 401–405.
- Schroeder, M. R., and Hall, J. L. (1974). "A model for mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.* **55**, 1055–1060.
- Seneff, S. (1988). "A joint synchrony/mean-rate model of auditory speech processing," *J. Phonetics* **16**, 55–76.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonsky, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Smith, R. L. (1979). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," *J. Acoust. Soc. Am.* **65**, 166–178.
- Strope, B., and Alwan, A (1997). "A model of dynamic auditory perception and its application to robust word recognition," *IEEE Trans. Speech Audio Process.* **5**(5), 451–464.
- Tchorz, J., Kasper, K., Reininger, H., and Kollmeier, B. (1997). "On the interplay between auditory-based features and locally recurrent neural networks," *Proc. EUROSPEECH '97* **4**, 2075–2078.
- Wesselkamp, M. (1994). "Messung und Modellierung der Verständlichkeit von Sprache," *Doctoral thesis, Universität Göttingen*.

# Acoustical and perceptual study of gemination in Italian stops

Anna Esposito

*International Institute for Advanced Scientific Studies "E. R. Caianiello" (IIASS), Via G. Pellegrino 19, 184019 Vietri sul mare (SA), Italy, INFN Salerno University*

Maria Gabriella Di Benedetto

*INFOCOM Department, Rome University "La Sapienza," Via Eudossiana 18, 100184 Rome, Italy*

(Received 10 October 1996; accepted for publication 13 May 1999)

On the basis of theoretical considerations and results from acoustic and perceptual analyses, it is hypothesized that closure duration is the primary cue for gemination in Italian. Results of an acoustic analysis of a large number of single and geminate Italian utterances show two acoustic correlates of gemination: the length of the closure and the length of the vowel preceding the consonant. Other acoustic parameters were not systematically related to gemination. These results were validated perceptually. At the perceptual level, the above cues were used by the listeners in the geminate/nongeminate discrimination; however, closure duration played a major role. Moreover, it was found that the significant lengthening of consonant was only partially compensated by the shortening of the previous vowel and by a small lengthening of the geminate utterance with respect to the nongeminate one. This result suggests that speakers follow a sort of timing (rhythm) which is fixed in duration and depends on the number of syllables in the word: words with equal numbers of syllables do not change in utterance length, an elongated segment being partly compensated by the shortening of another. This process seems to be applied also perceptually suggesting that the timing (rhythm) of a language is also an auditory attitude. © 1999 Acoustical Society of America. [S0001-4966(99)04309-X]

PACS numbers: 43.71.Es, 43.71.Hw, 43.70.Fg, 43.70.Hs [WS]

## INTRODUCTION

Some languages allow the clustering of the same consonant in vowel contexts and this phenomenon is known as "consonant gemination." Gemination plays a particular role in the phonology of such languages because several words change meaning as a function of singleton versus geminate consonants (minimal pairs). Phonetic theories agree in considering the gemination of a phoneme as a particular realization of the original one (Muljadic, 1972) which is modified in some of the acoustic parameters. Recent papers (Rochet and Rochet, 1995; Shrotriya *et al.*, 1995) report that there is an acoustic relationship between consonant closure duration and gemination as well as between the length of the vowel preceding the consonant and gemination. Moreover, these studies also report that there is a perceptual relationship between closure duration and gemination, whereas variation in the length of the vowel preceding the consonant does not seem to be perceptually relevant. Other acoustic parameters which appear to be related with gemination (Shrotriya *et al.*, 1995) are the burst energy and the  $F_0$  values at the offset of the vowel preceding the consonant. However, there is not extensive work on this phenomenon and the results reported are based on a small number of data.

Our interest was to examine the acoustic parameters which play a role in the production of geminate consonants and to validate their perceptual importance. To this end, we set up a series of experiments for collecting the acoustic data and for synthesizing the stimuli for the perceptual tests. Our aim was to try to give an answer to the following questions: (1) Which acoustic parameters are related with gemination? Specifically, does a feature of tension associated with gemi-

nation cause more extreme vowel formants [or some other spectral shape effects as for Hindi (Shrotriya *et al.*, 1995)] for Italian geminate stops? To this end, a data base of geminate and nongeminate utterances with no semantic meaning was built up and acoustic analysis of such data was carried out. (2) Are the significant acoustic attributes of a geminate phoneme production also perceptually significant?

## I. SPEECH MATERIALS AND MEASUREMENTS

### A. Data

A set of vowel-consonant-vowel (the nongeminate case) and vowel-consonant-consonant-vowel (the geminate case) utterances in which the consonant was [b,d,g,p,t,k] (the complete set of stop consonants in Italian) in the environment of the vowels [i,a,u] was recorded. These bisyllabic utterances were chosen because, in Italian, many minimal pairs, such as *papa* (pope) and *pappa* (baby food), *fato* (fate) and *fatto* (fact), *rito* (rite) and *ritto* (stand up), *pala* (shovel) and *palla* (ball), are bisyllabic words; therefore, the use of a bisyllabic structure is justified by the natural attitude of the native speakers in producing it. The use of nonsense words was necessary for having in all cases a symmetrical context and the same stress pattern. Since different acoustic parameters, including durational parameters, were measured, the utterances were not included in a carrier phrase; the stress and intonation pattern of the whole sentence would obviously influence such durational parameters in a way which would be difficult to control. The utterances were produced by six Italian speakers (three male and three female). Each

utterance was repeated three times for a total of 324 utterances in single form and 324 utterances in geminate form.

## B. Subjects

Six Italian adult speakers with no known articulatory impairment served as subjects. Four of them were students and two of them professors at Rome University “La Sapienza.” All subjects were native speakers of standard Italian that they learned during childhood. Their speech was characterized by the accent of the Italian region where they spent most of their life or by the accent of the closest relatives (parents and grandparents). Some of them showed no dialectal accent, whereas some others showed a Roman dialect accent.

## C. Recording procedures

The speech materials were produced by the speakers in a sound-treated room and recorded on a high-quality magnetic tape recording system. Care was taken to ensure that vocal effort and patterns of stress and intonation were reasonably natural and consistent from word to word. Before measurements were performed, the recorded utterances were evaluated and the utterances which were judged by the experimenter to be unacceptable samples of the phonemes in question were re-recorded. For example, a sample was judged to be unacceptable when mistakes happened in the recording procedure, or the speaker made pronunciation errors.

The recordings were made in the Speech Laboratory, INFOCOM Department, Rome University “La Sapienza” (Italy). The measurements were performed using the UNICE version 1.6 by VECSYS speech analysis program, which accepts user commands to read in waveform files and generates spectral displays of various types. The spectral representation used for the analysis of our data was the DFT (Discrete Fourier Transform) magnitude spectrum. The analysis window (Hamming window) duration was set to a default of 256 samples which corresponds roughly to 26 ms at a sampling rate of 10 kHz. The first step in the analysis procedure was to process the speech signals by a low-pass digital filter with a cutoff frequency of 5 kHz.<sup>1</sup>

The output of the filter was sampled at 10 kHz and stored both on a PC computer memory and on floppy disks. Sound spectrograms of all utterances and visual displays of the corresponding waveforms were also made. For these records, the criteria to perform the different measurements were established by visual inspection of the spectrogram, the waveform, and the spectrum to obtain a coherent set of measures.

## D. Measurements in the frequency domain

### 1. V1 formant frequencies

Using a DFT spectrum, measurements of the formant frequencies,  $F1$ ,  $F2$ , and  $F3$ , at the offset and in the middle of the vowel preceding the consonant were made. In order to avoid incorrect DFT estimations due to windowing and other side effects, these measurements were also checked for er-

rors by visual inspection of the spectrograms, which can give a coarse idea of each formant frequency range.

## 2. Parameters

Once the temporal sampling points corresponding to the consonant release were located, the burst offset, and the onset of the vowel following the consonant, the following parameters were examined, using software programs developed *ad hoc* for these tasks:

- (1) The burst energy: the energy in the temporal interval from the consonant release to the burst offset was computed by squaring and adding the samples contained in it. The result was divided by  $V1$  energy which was computed on three vowel frames located around the middle of  $V1$ .
- (2) The VOT energy: the energy in the temporal interval from the consonant release to the onset of the vowel following the consonant was computed. The result was divided by  $V1$  energy.
- (3) The burst power: the ratio between the burst energy and the burst duration;
- (4) The VOT power: the ratio between the VOT energy and the VOT duration.

These computations were only made for voiceless consonants because of the difficulty in defining the temporal sampling point corresponding to the burst offset for voiced consonants. For voiceless consonants, this temporal sampling point was identified as the time (after the consonant release) at which no energy below 1.5 kHz was visible in the spectrogram.

## 3. DFT spectra at consonant release

All the spectra were pre-emphasized. The 256-point analysis window was placed on the waveform in order to have the burst onset fall in the zone corresponding to the maximum value of the analysis window. A visual examination of the spectra was first performed. Then, the ratio of the 0–0.3 kHz to the 0–5 kHz frequency range signal energy was computed. Finally, the spectrum energy was quantized using a vector quantization framework (Vannucci, 1994; Rossetti, 1994) by dividing the 0.3–5 kHz frequency range into frequency bands. Since low-frequency samples are influenced by the voicing, the 0–0.3 kHz band was excluded from the centroid’s calculations. The total energy of the signal in the 0.3–5 kHz frequency range was computed and equally distributed over the frequency bands. Therefore, the frequency bands had different widths. Each band  $B_i$ ,  $i = 1, \dots, N = 10$  was then assigned to a centroid  $C_i$  defined by the pair of values  $C_{ix}$  and  $C_{iy}$ :

$$C_{ix} = \frac{\sum_{f_n \in B_i} f_n A(f_n)}{\sum_{f_n \in B_i} A(f_n)}, \quad (1)$$

$$C_{iy} = \sum_{f_n \in B_i} A(f_n) / |B_i|,$$

where  $A(f_n)$  is the magnitude of the  $n$ th DFT sample in the  $B_i$  frequency band,  $|B_i|$  is the number of DFT samples in the

$B_i$  frequency band, and  $f_n$  is the  $n$ th harmonic component in the  $B_i$  frequency band. The connected  $N$  code-points gave a smoothed representation of the spectrum energy distribution in the 0.3–5 kHz frequency range. Such representation was computed for each consonant in geminate and nongeminate forms and was considered as a template of the consonant burst spectrum.

## E. Measurements in the time domain

### 1. Duration of the vowel preceding the consonant (V1 duration) and the vowel following the consonant (V2 duration)

The temporal sampling point defining the vowel onset was identified as the temporal instant at which, in the waveform, a glottal pulse appears, followed by other regular glottal pulses. The vowel onset time was set by the experimenter by placing a cursor on the waveform display. In some cases, mainly for the vowel [a], a glottal excitation was visible before the regular vowel voicing. This glottal excitation was discarded in the vowel onset measurement. The temporal sampling point defining the vowel offset was identified as the temporal instant in which, in the spectrogram, the frequency energy was lower than 1 kHz. This criterion was also used to define the offset of the vowel following the consonant. The temporal sampling point defining the onset of the vowel following the consonant was identified as the temporal instant in which a frequency energy greater than 1 kHz appeared in the spectrogram.

### 2. Consonant closure duration

This measure was defined as the time interval from the offset of V1 to the consonantal release. To identify the temporal sampling point corresponding to the consonant release, the waveform, and the spectrogram were examined in parallel. The oral release is marked in the spectrogram by an abrupt onset of energy. An abrupt release does not always occur in the case of voiced consonants, and it is not always possible to identify the release by only looking at the spectrogram. In such cases, the examination of the waveform was useful because the amplitude of the consonantal voicing is lower than the amplitude of the vowel voicing. The consonant release was identified as the instant at which there was an abrupt onset of energy in the spectrogram and/or an amplitude change in the waveform. Other durational measures, such as VOT and complete utterance duration, were obtained as difference or sum of the measurements made in Secs. IE 1 and 2.

## II. RESULTS OF ACOUSTIC ANALYSES

The values of the twelve acoustic attributes, (1) V1 formant frequencies, (2) burst energy, (3) VOT energy, (4) burst power, (5) VOT power, (6) DFT spectra at (12) consonant release, (7) VOT duration, (8) burst duration, (9) utterance duration, (10) V1 duration, (11) V2 duration, and (12) consonant closure duration were computed for the 648 utterances (324 single and 324 geminate).

The aim of the production experiment was to try to understand whether the above parameters would show any sig-

nificant difference in the single versus the geminate consonants. To this purpose, each parameter was analyzed separately as described below. An ANOVA statistical analysis was performed on all the data. Gemination was treated as a between subjects factor, whereas other parameters, such as vowel category, formant values, and consonants, were treated as within subjects factors. Test of the main effect were performed when interactions were present, that is, the effect of one factor was explored at each level of the other factor. Furthermore, on some of the duration data was applied the Maximum Likelihood Criterion.

## A. Results in the frequency domain

Formant frequency values were averaged over all speakers and repetitions, keeping separate sentences differing in vowel identity, and single versus geminate consonants forms. An ANOVA analysis was performed on formant frequency values. Gemination was treated as a between subject factor, and vowel category [i,a,u], and formant values (measured in the middle and at the offset of the vowels) were treated as a within subjects factors. The ANOVA analysis showed that gemination had no effect on formant frequency values [ $F(1,10)=0.485$ ,  $p>0.1$  for  $F1$ ,  $F(1,10)=0.028$ ,  $p>0.1$  for  $F2$ ,  $F(1,10)=0.650$ ,  $p>0.1$  for  $F3$ ]. Furthermore, no interaction was found between gemination and vowel category [ $F(2,20)=0.113$ ,  $p>0.1$  for  $F1$ ,  $F(2,20)=0.006$ ,  $p>0.1$  for  $F2$ ,  $F(2,20)=0.179$ ,  $p>0.1$  for  $F3$ ], and between formant values measured in the middle and at the offset of the vowels [ $F(1,10)=0.485$ ,  $p>0.1$  for  $F1$ ,  $F(1,10)=0.018$ ,  $p>0.1$  for  $F2$ ,  $F(1,10)=0.872$ ,  $p>0.1$  for  $F3$ ]. Although, obviously, the formant frequency values change when the vowel or the consonant context were varied, results were very stable in terms of comparison between geminate and nongeminate consonants, suggesting that there was no relationship between formant frequencies and gemination.

The ANOVA analysis performed on the burst energy values showed that gemination had no effect on burst energy [ $F(1,16)=0.480$ ,  $p>0.1$  for female;  $F(1,16)=0.597$ ,  $p>0.1$  for male]. No interaction was found between gemination and vowels and between gemination and consonants both for female [ $F(2,32)=0.088$ ,  $p>0.1$ ;  $F(2,32)=0.259$ ,  $p>0.1$ ] and male [ $F(2,32)=2.956$ ,  $p>0.05$ ;  $F(2,32)=1.331$ ,  $p>0.1$ ] speakers. Burst energy was significantly affected by consonant category [ $F(2,32)=30.279$ ,  $p<0.0001$  for female;  $F(2,32)=28.988$ ,  $p<0.0001$  for male] and to a less extent by vowel category [ $F(2,32)=7.418$ ,  $p<0.01$  for female;  $F(2,32)=5.263$ ,  $p<0.05$  for male].

The ANOVA analysis performed on the VOT energy values showed that gemination had no effect on VOT energy [ $F(1,16)=0.117$ ,  $p>0.1$  for female;  $F(1,16)=0.428$ ,  $p>0.1$  for male]. No interaction was found between gemination and vowels and gemination and consonants both for female [ $F(2,32)=0.388$ ,  $p>0.1$ ;  $F(2,32)=0.437$ ,  $p>0.1$ ] and male [ $F(2,32)=4.115$ ,  $p>0.01$ ;  $F(2,32)=0.831$ ,  $p>0.1$ ] speakers. VOT energy was significantly affected by consonant category [ $F(2,32)=31.094$ ,  $p<0.0001$  for female;  $F(2,32)=29.597$ ,  $p<0.0001$  for male] and to a less extent by vowel category [ $F(2,32)=8.174$ ,  $p<0.01$  for female;  $F(2,32)=4.411$ ,  $p<0.05$  for male]. Figure 1(a) and



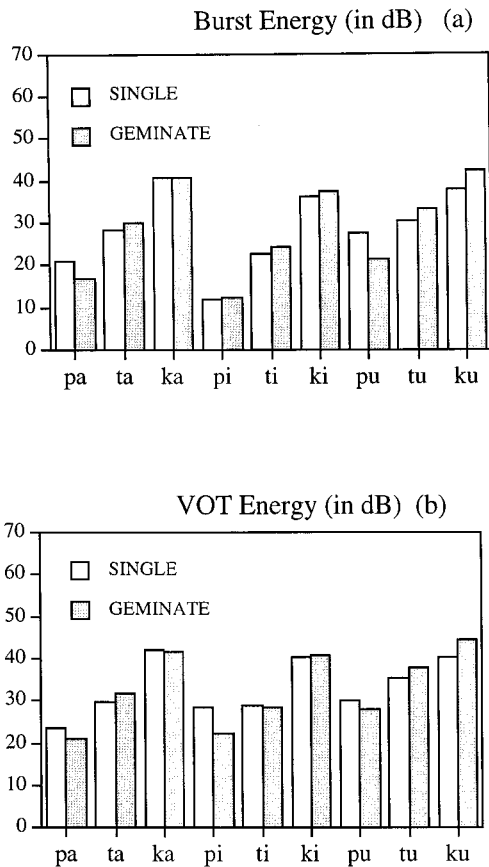


FIG. 1. Averaged values of burst energy (a) and averaged values of VOT energy (b) for single and geminate utterances. The data are averaged over speakers and repetitions.

(b) reports, respectively, the burst and the VOT energy values averaged over all the repetitions, keeping separate sentences differing in consonant place of articulation, vowel identity, and single versus geminate forms.

The ANOVA analysis performed on the burst power values showed that gemination had no effect on burst power [ $F(1,16)=0.834$ ,  $p>0.1$  for female;  $F(1,16)=0.384$ ,  $p>0.1$  for male]. No interaction was found between gemination and vowels and between gemination and consonants both for female [ $F(2,32)=0.074$ ,  $p>0.1$ ;  $F(2,32)=1.543$ ,  $p>0.1$ ] and male [ $F(2,32)=3.065$ ,  $p>0.1$ ;  $F(2,32)=.438$ ,  $p>0.1$ ] speakers. Burst power was significantly affected by consonant category [ $F(2,32)=14.811$ ,  $p<0.0001$  for female;  $F(2,32)=14.117$ ,  $p<0.0001$  for male] and to a lesser extent by vowel category [ $F(2,32)=14.065$ ,  $p<0.0001$  for female;  $F(2,32)=7.526$ ,  $p<0.01$  for male].

The ANOVA analysis performed on the VOT power values showed that gemination had no effect on VOT power [ $F(1,16)=1.093$ ,  $p>0.1$  for female;  $F(1,16)=0.007$ ,  $p>0.1$  for male]. No interaction was found between gemination and vowels [ $F(2,32)=1.372$ ,  $p>0.1$ , for female;  $F(2,32)=2.425$ ,  $p>0.1$ , for male] and gemination and consonants [ $F(2,32)=2.339$ ,  $p>0.1$ , for female;  $F(2,32)=0.378$ ,  $p>0.1$  for male]. VOT power was significantly affected by consonant category [ $F(2,32)=12.132$ ,  $p<0.001$  for female;  $F(2,32)=9.388$ ,  $p<0.001$  for male] and by

vowel category [ $F(2,32)=16.409$ ,  $p<0.0001$  for female;  $F(2,32)=9.000$ ,  $p<0.001$  for male].

Both burst energy and VOT energy values showed great variability among speakers even for the same place of articulation and the same vowel context. Burst power values showed great variability between the geminate and nongeminate case which were depending on the place of articulation, the vowel context, and the speakers. Geminate labial consonants showed lower burst power values than single ones, whereas the opposite was true for geminate dental consonants. Geminate velar consonants showed lower burst power values than single ones in the context of the vowels [i,a] and higher values in the context of the vowel [u]. However, this was not true for all speakers. For example, labial consonants in [a] context showed for some speakers higher burst power values in the nongeminate case, while for some others the opposite was true. No systematic difference between single and geminate forms was observed to be present across place of articulation and vowel context.

VOT power values showed a behavior similar to the burst power (such measurements were computed only for voiceless consonants) with no significant difference between single and geminate forms. This result was somewhat unexpected, because of the general feeling that geminates are produced with greater effort than nongeminates, resulting in a greater energy at the release.

The ANOVA analysis performed on the ratio of the signal energy in the 0–0.3 kHz to the 0–5 kHz frequency range showed that gemination had no effect on the signal-energy ratio both for female [ $F(1,16)=0.029$ ,  $p>0.1$ ] and male [ $F(1,16)=0.041$ ,  $p>0.1$ ] speakers. No interaction was found between gemination and vowels [ $F(2,32)=0.153$ ,  $p>0.1$  for female;  $F(2,32)=0.635$ ,  $p>0.1$ , for male]. No interaction was found between gemination and consonants for female speakers [ $F(2,32)=1.125$ ,  $p>0.1$ ]. For male speakers it was found an interaction between gemination and consonants [ $F(2,32)=3.965$ ,  $p<0.01$ ]. We explored, in this case, the effect of gemination for each consonant and we found that gemination was not significant for all the consonants [ $F(1,16)=0.587$ ,  $p>0.1$  for [t];  $F(1,16)=1.722$ ,  $p>0.1$  for [k];  $F(1,16)=1.006$ ,  $p>0.1$  for [b];  $F(1,16)=0.343$ ,  $p>0.1$  for [d];  $F(1,16)=2.837$ ,  $p>0.1$  for [g]] except for [p] [ $F(1,16)=11.609$ ,  $p<0.01$ ]. Since, the effect of gemination on consonant category was not systematic, we concluded that this statistical significance was not of practical importance. However, consonant category was significant both for single [ $F(5,80)=140.541$ ,  $p<0.001$ ] and geminate utterances [ $F(5,80)=135.587$ ,  $p<0.001$ ]. Figure 2 reports the ratio of the signal energy in the 0–0.3 kHz to the 0–5 kHz frequency range for each vowel and consonant in the single and geminate form.

The ratio of the signal energy was significantly affected by consonant category [ $F(5,80)=61.108$ ,  $p<0.0001$  for female;  $F(5,80)=272.162$ ,  $p<0.0001$  for male] and by vowel category [ $F(2,32)=16.021$ ,  $p<0.0001$  for female;  $F(2,32)=77.957$ ,  $p<0.0001$  for male]. An interaction was found between vowels and consonants both for female [ $F(10,160)=22.014$ ,  $p<0.0001$ ] and male [ $F(10,160)=35.660$ ,  $p<0.0001$ ] speakers. We explored, in this case,

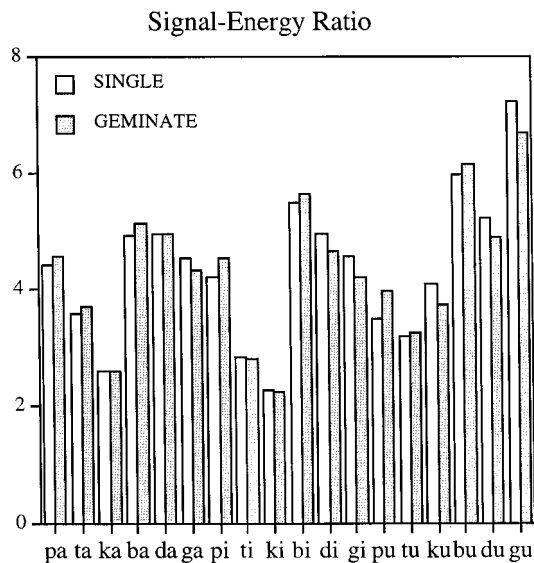


FIG. 2. Percentage of the ratio of the signal energy in the 0–0.3 kHz to 0–5 kHz frequency ranges [ $E_n(0-0.3\text{ kHz})/E_n(0-5\text{ kHz})$ ]. The reported values were averaged over three repetitions for each speaker.

the effect of vowels for each consonant and viceversa. We found that, for male speakers, the ratio of the signal-energy values were significantly affected by vowel category in the context of [p] [ $F(2,32)=7.386, p<0.01$ ], [k] [ $F(2,32)=77.677, p<0.001$ ], [b] [ $F(2,32)=9.340, p<0.01$ ], [g] [ $F(2,32)=119.917, p<0.001$ ], but not in the context of [t] [ $F(2,32)=3.850, p>0.01$ ], and [d] [ $F(2,32)=5.121, p>0.01$ ]. Moreover, the effect of the consonants was significant for each vowel [ $F(5,80)=105.865, p<0.01$  for [a];  $F(5,80)=169.329, p<0.01$  for [i];  $F(5,80)=116.040, p<0.01$  for [u]].

For female speakers, the ratio of the signal-energy values were significantly affected by vowel category in the context of [t] [ $F(2,32)=33.279, p<0.001$ ], [k] [ $F(2,32)=51.656, p<0.001$ ], [b] [ $F(2,32)=12.799, p<0.001$ ], [g] [ $F(2,32)=50.591, p<0.001$ ], but not in the context of [p] [ $F(2,32)=4.714, p>0.01$ ], and [d] [ $F(2,32)=0.606, p>0.01$ ]. Moreover, the effect of the consonants was significant for each vowel [ $F(5,80)=29.369, p<0.01$ , for [a];  $F(5,80)=43.321, p<0.01$ , for [i];  $F(5,80)=40.632, p<0.01$ , for [u]]. It is hard from these results to determine which of the two features (vowel or consonant category) play a major role in determining the signal-energy ratio values, even though the effect of consonant appear to be more systematic. Further research is necessary to evaluate this possibility, and at the moment it is beyond the aim of this paper.

DFT spectra of all consonants in single and geminate form were visually examined. The smoothed energy distribution (for the description of this measure see Sec. D3) did not show any difference between the geminates and nongeminates. Figure 3 shows the energy distribution for vowel [a] (squares), [i] (circles), and [u] (triangles) in labial [Fig. 3(a)], dental [Fig. 3(b)], and velar context [Fig. 3(c)] (respectively) in geminate and nongeminate forms. As shown, the plots of the geminate and nongeminate energy distribution overlapped quite closely. In conclusion, no relationship between

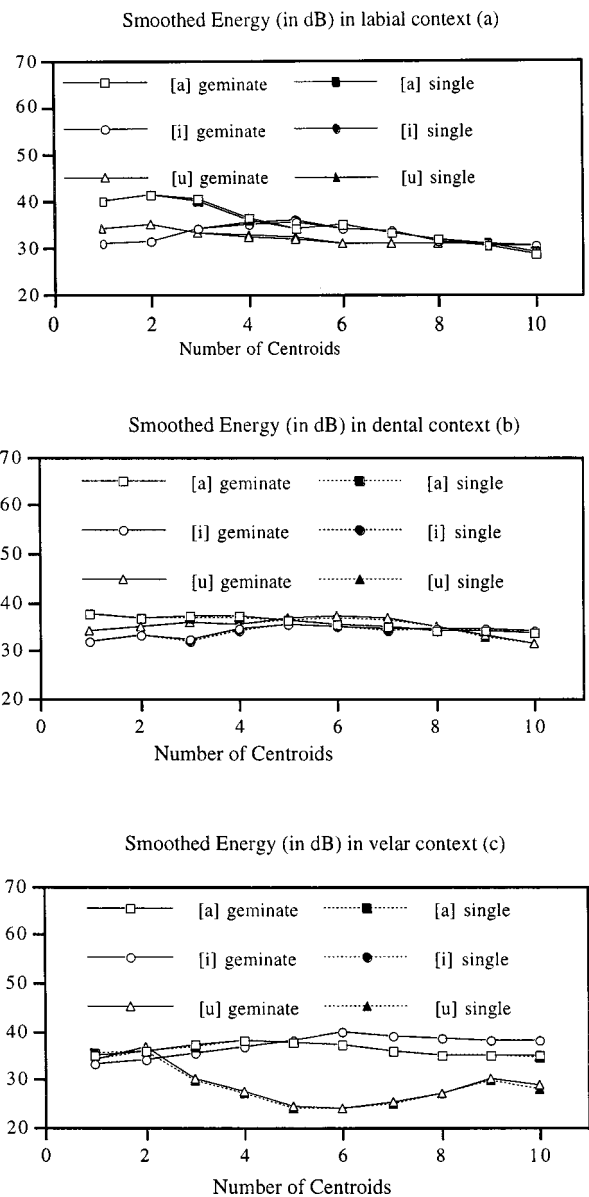


FIG. 3. Smoothed energy distribution of labial consonants in the [a] environment in geminate (empty squares) and nongeminate (filled squares) forms, in the [i] environment in geminate (empty circles) and nongeminate (filled circles) forms, and in the [u] environment in geminate (empty triangles) and nongeminate (filled triangles) forms. Labial (a), dental (b), and velar (c) consonants are plotted separately.

the energy representations and the acoustics of the gemination was observed.

## B. Results in the time domain

The ANOVA analyses in the time domain were performed first by taking into account separately female and male data. The reason for that was that we expect that gender could play a role on duration measures. Successively, as suggested by the reviewers, the ANOVA analyses were performed putting together the male and female data. This was done for all the duration measures reported below (VOT, burst duration, V2 duration, utterance duration, V1 duration, and closure duration). The results obtained (collapsing together male and female data) were consistent with those obtained considering the male and female data separately, ex-

cept for a little effect of gemination on utterance duration which will be discussed below. To be consistent with all the other data reported above, below are reported the ANOVA analyses performed separately on male and female data, where gemination has been considered as a between subject variable and vowels and consonants have been considered as within subject variables.

First, results which did not show any relationship with gemination are reported. The ANOVA analysis performed on VOT duration showed that gemination had no effect on VOT duration both for female [ $F(1,16) = .249, p > 0.1$ ] and male [ $F(1,16) = .005, p > 0.1$ ] speakers. No interaction was found between gemination and vowels [ $F(2,32) = .251, p > 0.1$  for female;  $F(2,32) = 1.048, p > 0.1$  for male], and between gemination and consonants [ $F(2,32) = 3.173, p > 0.01$  for female;  $F(2,32) = 5.326, p > 0.01$  for male]. As we expected, VOT duration was significantly affected by vowel category [ $F(2,32) = 52.820, p < 0.0001$  for female;  $F(2,32) = 73.254, p < 0.0001$  for male], and by consonant category [ $F(2,32) = 99.734, p < 0.0001$  for female;  $F(2,32) = 179.673, p < 0.0001$  for male].

The ANOVA analysis performed on burst duration showed that gemination had no effect on burst duration both for female [ $F(1,16) = 0.049, p > 0.1$ ] and male [ $F(1,16) = 0.735, p > 0.1$ ] speakers. No interaction was found between gemination and vowels [ $F(2,32) = 0.349, p > 0.1$  for female;  $F(2,32) = 0.423, p > 0.1$  for male]. No interaction was found between gemination and consonants for female speakers [ $F(2,32) = 0.051, p > 0.1$ ]. For male speakers an interaction was found between gemination and consonants [ $F(2,32) = 7.246, p < 0.01$ ]. We explored, in this case, the effect of gemination for each consonant and found that gemination was not significant for all the consonants [ $F(1,16) = 4.862, p > 0.01$  for [p];  $F(1,16) = 0.152, p > 0.01$  for [t];  $F(1,16) = 6.612, p > 0.01$  for [k]]. However, consonant category was significant both for single [ $F(2,32) = 16.117, p < 0.001$ ] and geminate utterances [ $F(2,32) = 60.375, p < 0.001$ ]. Hence, we concluded that also for male speakers, gemination had no effect on burst duration.

As we expected, burst duration was significantly affected by consonant category [ $F(2,32) = 33.661, p < 0.0001$  for female;  $F(2,32) = 69.245, p < 0.0001$  for male], whereas vowel category did not play a significant role because we found a small effect for male speakers [ $F(2,32) = 5.802, p < 0.01$ ] but no effect for female speakers [ $F(2,32) = 1.301, p > 0.1$ ].

The ANOVA analysis performed on V2 duration showed that gemination had no effect on V2 duration both for female [ $F(1,16) = 2.029, p > 0.1$ ] and male [ $F(1,16) = 1.545, p > 0.1$ ] speakers. No interaction was found between gemination and vowels [ $F(2,32) = 1.943, p > 0.1$  for female;  $F(2,32) = 0.958, p > 0.1$  for male], and between gemination and consonants [ $F(2,32) = 1.397, p > 0.1$  for female;  $F(2,32) = 0.731, p > 0.1$  for male]. V2 duration was significantly affected by vowel category [ $F(2,32) = 46.141, p < 0.0001$  for female;  $F(2,32) = 26.357, p < 0.0001$  for male], and by consonant category [ $F(2,32) = 5.692, p < 0.001$  for female;  $F(2,32) = 11.088, p < 0.0001$  for male]. Figure 4 reports the averaged V2 durations in the single and

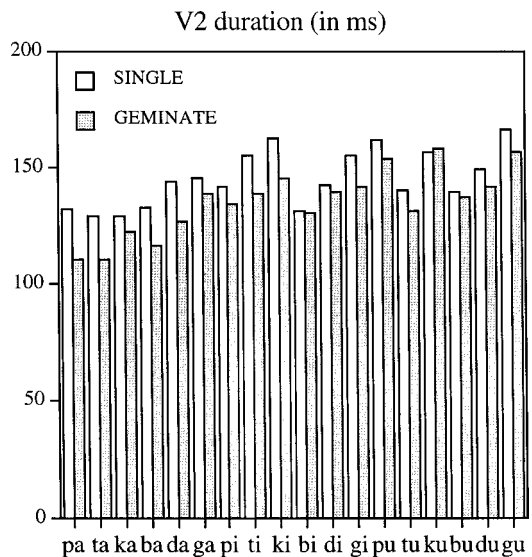


FIG. 4. V2 duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

geminate form, for the vowels [a,i,u] in the consonant context [p,t,k,b,d,g].

The ANOVA analysis performed on utterance duration showed that gemination had no effect on utterance duration both for female [ $F(1,16) = 7.952, p > 0.01$ ] and male [ $F(1,16) = 7.065, p > 0.01$ ] speakers. No interaction was found between gemination and vowels [ $F(2,32) = 0.772, p > 0.1$  for female;  $F(2,32) = 0.521, p > 0.1$  for male]. No interaction was found between gemination and consonants for female speakers [ $F(5,80) = 1.604, p > 0.1$ ]. For male speakers an interaction was found between gemination and consonants [ $F(5,80) = 3.749, p < 0.01$ ]. We explored, in this case, the effect of gemination for each consonant and we found that gemination was not significant for [p] [ $F(1,16) = 3.985, p > 0.01$ ], [b] [ $F(1,16) = 4.602, p > 0.01$ ], [d] [ $F(1,16) = 4.059, p > 0.01$ ], and [g] [ $F(1,16) = 0.690, p > 0.01$ ], whereas it was significant for [t] [ $F(1,16) = 10.473, p < 0.01$ ] and for [k] [ $F(1,16) = 16.220, p < 0.01$ ]. However, the consonant category was significant both for single [ $F(5,80) = 3.686, p < 0.01$ ] and geminate utterances [ $F(2,32) = 7.080, p < 0.001$ ]. Utterance duration was significantly affected by consonant category [ $F(5,80) = 7.882, p < 0.001$  for female;  $F(5,80) = 7.017, p < 0.0001$  for male], whereas vowel category do not play a significant role both for female [ $F(2,32) = 3.530, p > 0.01$ ] and male [ $F(2,32) = 0.112, p > 0.01$ ] speakers. When the data for male and female were collapsed together, an effect of gemination was found on utterance duration [ $F(1,32) = 14.936, p = 0.0005$ ] and an interaction between gemination and consonants [ $F(5,160) = 5.098, p = 0.0002$ ].

We explored, in this case, the effect of gemination for each consonant and found that gemination was not significant for [g] [ $F(1,32) = 3.720, p > 0.01$ ], whereas it was significant for [b] [ $F(1,32) = 8.485, p < 0.01$ ], [d] [ $F(1,32) = 8.485, p < 0.01$ ], [p] [ $F(1,32) = 11.206, p < 0.01$ ], [t] [ $F(1,32) = 23.921, p < 0.001$ ], and for [k] [ $F(1,32) = 23.596, p < 0.001$ ]. The consonant category also was significant both for single [ $F(5,160) = 5.276, p < 0.001$ ] and



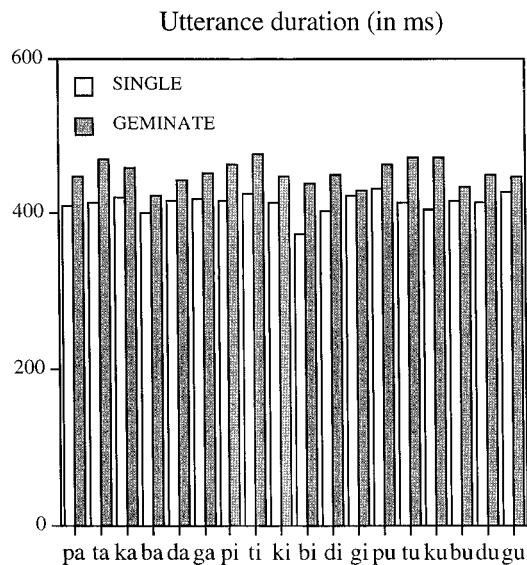


FIG. 5. Utterance duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

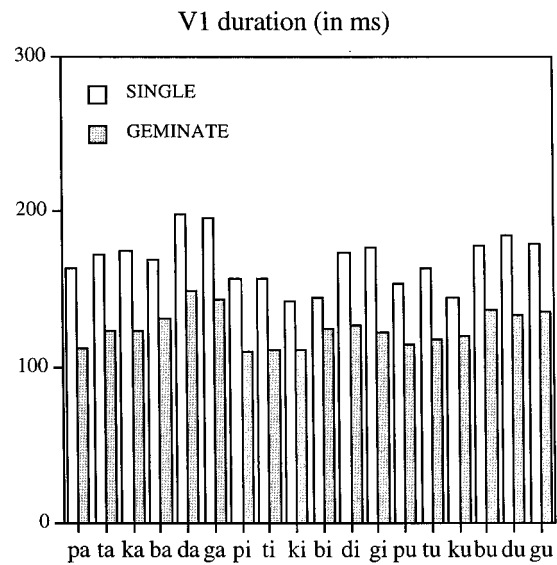


FIG. 6. V1 duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

geminate utterances [ $F(5,160) = 14.147, p < 0.001$ ].

From these results, we concluded that there was an effect of gemination on consonant category. However, since the consonant category also has a significant effect on gemination, it becomes difficult to define the role played by gemination on utterance duration. Further research is necessary to evaluate this effect, and to determine whether the present result generalizes to other consonant categories.

Figure 5 reports the averaged utterance durations in the single and geminate form, for the vowels [a,i,u] in the consonant context [p,t,k,b,d,g].

It is worth noting that the standard deviation values for VOT, burst duration, V2 duration, and utterance duration were very high in comparison with the differences among their values in the geminate and nongeminate case.

Moreover, for these parameters, there was a great deal of variability among speakers: Some speakers did not show any durational difference, whereas some others did. Some vowel contexts showed larger differences than others. Consonantal place and voicing also played a role for such variability. We can conclude from these data that V2 duration, burst duration, and VOT do not play a role in the geminate/nongeminate distinction, whereas further research is necessary to define the role played by gemination on utterance duration.

The only durational parameters which showed significant differences between geminate and nongeminate consonants were the V1 duration and the closure duration.

For V1 duration, the results of the ANOVA analysis showed that gemination plays a significant role both for female [ $F(1,16) = 12.531, p < 0.01$ ] and male [ $F(1,16) = 59.871, p < 0.0001$ ] speakers. No interaction was found between gemination and vowels both for female [ $F(2,32) = 3.915, p > 0.01$ ] and male [ $F(2,32) = 0.365, p > 0.1$ ] speakers. No interaction was found between gemination and consonants for male speakers [ $F(5,80) = 2.680, p > 0.01$ ]. For female speakers an interaction was found between gemination and consonants [ $F(5,80) = 6.605, p < 0.0001$ ]. We

explored, in this case, the effect of gemination for each consonant and found that gemination was significant for [p] [ $F(1,16) = 18.107, p < 0.01$ ], [t] [ $F(1,16) = 9.704, p < 0.01$ ], [d] [ $F(1,16) = 24.685, p < 0.001$ ], and [g] [ $F(1,16) = 11.158, p < 0.01$ ], whereas it was not significant for [k] [ $F(1,16) = 7.222, p > 0.01$ ] and for [b] [ $F(1,16) = 6.731, p > 0.01$ ]. Moreover, the consonant category was significant for both single [ $F(5,80) = 31.051, p < 0.001$ ] and geminate utterances [ $F(5,80) = 11.715, p < 0.001$ ]. From these results we concluded that gemination was of practical importance for the parameter we are considering (V1 duration). V1 duration was also significantly affected by the consonant category both for male [ $F(5,80) = 24.653, p < 0.0001$ ] and female [ $F(5,80) = 36.161, p < 0.0001$ ] speakers, and by vowel category both for female [ $F(2,32) = 17.488, p < 0.0001$ ] and male [ $F(2,32) = 28.989, p < 0.0001$ ] speakers. Averaged V1 durations (for geminate and nongeminate cases) as a function of vowels context and consonantal place are reported in Fig. 6.

The results of the ANOVA analysis on closure duration showed that gemination plays a significant role both for female [ $F(1,16) = 99.110, p < 0.0001$ ] and male [ $F(1,16) = 258.114, p < 0.0001$ ] speakers. No interaction was found between vowels and gemination for male [ $F(2,32) = 3.471, p > 0.01$ ] speakers. For female speakers an interaction was found between gemination and vowels [ $F(2,32) = 9.086, p < 0.001$ ]. We explored, in this case, the effect of gemination for each vowel and found that gemination was significant for all the vowels [ $F(1,16) = 100.819, p < 0.001$  for [a];  $F(1,16) = 81.534, p < 0.001$  for [i];  $F(1,16) = 97.257, p < 0.001$  for [u]], whereas the effect of vowel category was significant for geminate utterances [ $F(2,32) = 28.620, p < 0.001$ ] but not for single utterances [ $F(2,32) = 1.322, p > 0.1$ ]. An interaction was found between gemination and consonants both for male [ $F(5,80) = 12.403, p < 0.0001$ ] and female [ $F(5,80) = 10.935, p < 0.0001$ ] speakers. We explored, in this case, the effect of gemination for each consonant and found that gemination was significant for all con-



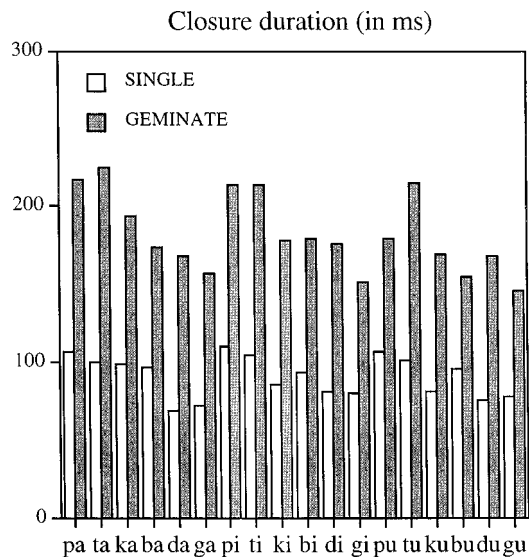


FIG. 7. Closure duration for single and geminate utterances. The reported values are averaged over speakers and repetitions.

sonants both for male [ $F(1,16)=205.908, p<0.001$  for [p];  $F(1,16)=254.019, p<0.001$  for [t];  $F(1,16)=208.164, p<0.001$  for [k];  $F(1,16)=105.225, p<0.001$  for [b];  $F(1,16)=175.347, p<0.001$  for [d];  $F(1,16)=157.184, p<0.001$  for [g]] and female [ $F(1,16)=69.345, p<0.001$  for [p];  $F(1,16)=72.764, p<0.001$  for [t];  $F(1,16)=82.609, p<0.001$  for [k];  $F(1,16)=69.284, p<0.001$  for [b];  $F(1,16)=97.300, p<0.001$  for [d];  $F(1,16)=92.591, p<0.001$  for [g]] speakers. Consonant category also was significant both for single [ $F(5,80)=14.893, p<0.001$  for female;  $F(5,80)=19.180, p<0.001$  for male] and geminate utterances [ $F(5,80)=37.786, p<0.001$  for female;  $F(5,80)=77.204, p<0.001$  for male]. From these results, we concluded that gemination plays a significant role on closure duration. Closure duration was significantly affected by consonant category both for male [ $F(5,80)=12.403, p<0.0001$ ] and female [ $F(5,80)=41.743, p<0.0001$ ] speakers, whereas vowel category showed a significant effect for female [ $F(2,32)=20.856, p<0.0001$ ] but not for male [ $F(2,32)=4.953, p>0.01$ ] speakers. Averaged closure durations (for geminate and nongeminate cases) as a function of vowels context and consonantal place are reported in Fig. 7.

V1 duration in the geminate case was observed to be reduced by about 25% with respect to its duration in the nongeminate case. Closure duration in the geminate case was significantly elongated, by about the 100%, with respect to the nongeminate case. This result was present for all speakers, vowel contexts, consonant place of articulation, and consonantal voicing.

Classification based on the Maximum Likelihood Criterion (Dillon and Goldstein, 1984) when applied to all the measured acoustic parameters confirmed that gemination was significant only for V1 and closure durations. The basic idea behind MLC is that the parameters of a set of data can be described through a Gaussian with mean  $m$  and a covariance matrix  $S$ . Generally this hypothesis is applied to all kinds of natural phenomena and is not a limitation. The mean

$m$  and the covariance matrix  $S$  are computed from the data. Once we compute  $m$  and  $S$  from two (or more) different set of data, the two Gaussians which describe the two sets of data are compared (to see how much they differ) using different criteria. One criterion is the MLC criterion. The method uses, as a measure of separability (separability scores) between the two Gaussians, the percentage of errors made, making, through a maximum likelihood criterion, an “*a posteriori*” classification of each data in the two groups.

The Maximum Likelihood Criterion applied on V1 alone gives a percentage of error equal to 20.4% (Error% in Table I) and as Errors/Utterances 132/648. When the MLC criterion was applied on different subsets of all the utterances, the misclassifying percentage for V1 remained stable (about 20%), suggesting that this parameter is not influenced by the context.

Table I shows the separability scores obtained by applying the Maximum Likelihood Criterion to closure duration, and to both closure and V1 duration (for different speakers, vowel contexts, and places of articulation) and to V1 duration alone.

The separability scores were good in both cases, although they did not improve when closure and vowel durations were considered together. These results suggested that closure duration can be considered a primary acoustic cue for the geminate/nongeminate distinction, whereas the role played by V1 duration must still be investigated.

### III. DISCUSSION ON THE ACOUSTIC DATA

The results obtained from the acoustic measurements showed that:

- (1) Formant frequency values of the vowel preceding the consonant showed no relationship with gemination, suggesting that no extra vocal effort is needed in a geminate production.
- (2) There is no relationship between any representation of the energy at consonant release and gemination, in contrast with the general feeling that geminate consonants must show, at the release, greater energy than singletons. We measured the burst energy, the burst power, the VOT energy, the VOT power, the spectrum shapes, the energy distribution, and other energy measures (see Sec. II A). None of these acoustic measurements showed any significant relationship with gemination. Shrotriya *et al.* reported that:
 

“...the burst of geminate consonant is stronger (by about 10 dB) as compared to the burst of nongeminate” (cf. Shrotriya *et al.*, 1995, “Acoustic and perceptual characteristics of geminated Hindi stop consonants,” ICPHS95 4, pp. 134).

 However, they did not give a precise definition of the measurements made and moreover, this result could be attributed to a language specific effect (Hindi vs Italian).
- (3) The acoustic parameters which appeared strongly related with gemination were the durations of the intervocalic consonant (which is close to twice the duration of the singletons) and of those of the vowel preceding it. The above parameters were found to be significant also in

TABLE I. Separability scores obtained by applying the Maximum Likelihood Criterion (MLC) to Closure Duration (CLd), to both Closure and V1 Duration (V1d), and to V1 duration alone. Percentage of errors and number of errors over number of utterances are reported. (The voiced/voiceless feature, the gender of speakers, the vowel context, the place of articulation, and each consonant are considered separately.)

Context	MLC on CLd		MLC on CLd and V1d	
	Errors%	Errors/Utterances	Errors%	Errors/Utterances
Overall	3.9%	25/648	3.9%	25/648
Voiceless	2.5%	8/324	1.9%	6/324
Voiced	2.8%	9/324	2.5%	8/324
Male	3.1%	10/324	2.8%	9/324
Female	5.6%	18/324	4.3%	14/324
[a]	2.8%	6/216	3.2%	7/216
[i]	2.8%	6/216	4.2%	9/216
[u]	4.6%	10/216	5.1%	11/216
Labials	2.8%	6/216	3.7%	8/216
Dentals	2.3%	5/216	2.8%	6/216
Velars	3.2%	7/216	2.8%	6/216
[b]	3.7%	4/108	2.8%	3/108
[d]	0.9%	1/108	0.9%	1/108
[g]	0.9%	1/108	0.9%	1/108
[p]	1.9%	2/108	1.9%	2/108
[t]	0.0%	0/108	0.0%	0/108
[k]	2.8%	3/108	2.8%	3/108
	MLC on V1 alone			
	Errors%	Errors/Utterances		
V1d	20.4%	132/648		

Hindi geminate consonants (Shrotriya *et al.*, 1995), suggesting that this effect is language independent.

Results obtained from the measurements in the time domain needed a more accurate discussion. Closure and V1 durations were found to be the only acoustic parameters which indicate the presence or absence of a geminate consonant, hence, it appeared natural to face the problem of what this implies at a higher level.

We observed that the length of V2 was shorter in the geminate case, but this shortening was not significant. VOT and burst lengths remained roughly unchanged in both the single and geminate case. In the geminate case, the length of the closure (on the average) was twice the length of the singleton closure and the length of V1 was reduced by about 25%. The length of utterance was, on the average, longer in the geminate case. However (even though the results from the ANOVA analysis showed that gemination has a little effect on utterance duration) for this parameter, there was a great deal of variability among speakers: Some speakers did not show any durational difference, whereas some others did; some vowel contexts and consonant contexts showed larger differences than others. Considering that there was a great lengthening of the consonant, a partial reduction of V1, and that the other durational parameters as VOT, burst duration, and V2 duration were not significantly affected by gemination, we would expect a great lengthening of the geminate utterance with respect to the single one. However, the lengthening of the geminate utterance was considerably lower than expected.

This effect cannot be explained on the basis of a hypoth-

esis of “anticipatory and backward compensation” suggested by Lindblom and Rapp (1973) for Swedish. The above theory attributes the segmental reduction of a phoneme to the number of syllables that precede (anticipatory compensation) and follow (backward compensation) it. Surely, our data showed an effect of anticipatory compensation which is evident in the reduction of V1. The effects of the backward compensation were negligible even though they can play a role in the similar timing of the single and geminate utterances. However, this compensation did not depend on the number of syllables in the word (which in our experiment are the same for the single and geminate utterances). Thus it could only be attributed to the acoustic attributes of the phonetic segments composing the utterance. Therefore, all other factors being equal, if the only significant acoustic attribute to discriminate a segment from another is its duration, and if it is necessary to maintain a constant timing because of the well-defined stress conditions, then speakers perform on the other segments a phonetic shortening. This could suggest that there are temporal constraints on the rhythmic structure of a word (for well-defined conditions of lexical stress and number of syllables) that speakers unconsciously tend to maintain, by balancing the durational change of some of the phonetic segments (which are distinctive only through a durational feature) with the durational change of others (which are distinctive through other features not related to segmental duration).

#### IV. PERCEPTUAL EXPERIMENT

In order to test the relevance of the acoustic parameters which distinguish geminate versus single consonants, a per-

ceptual experiment was carried out. The aim of this experiment was to define a closure length which works as a threshold in the perception of gemination (that is, below such average closure duration consonants are perceived as singleton; above it consonants are perceived as geminate). Moreover, this experiment was also devoted to investigate the perceptual role played by V1 duration in geminate and non-geminate contexts.

The experiment was carried out using /apa/ and /appa/ stimuli. The use of such stimuli (which are not representative of all consonants and vowel contexts) was aimed to evaluate the perceptual relevance of the durational difference observed in the acoustic data (see Sec. II). The perceptual experiment was intended as a validation of a general effect. In order to obtain quantitative estimations of this effect, the above experiment should be duplicated with others VCV and VCCV utterances.

A symmetrical context was used because it removes a spectral cue: the rate of change of formants through the closure. Huggins (1972a) showed that, in medial stops, a symmetrical vowel context had no effect on the perception of different closure durations, except for an increased sensitivity of the subjects to the durational changes. Since we are investigating the perceptual relevance of durational changes, a symmetrical context was selected for our experiment.

### A. Stimuli

The stimuli were synthesized as follows: a natural /apa/ token, spoken by one subject, whose durational behavior was close to the average durational behavior of all the subjects, was extracted from the database. The vowel and closure durations in the original stimulus were 176 ms and 99 ms, respectively.

The digitized signal was then modified by means of a waveform editor (UNICE editor) to produce two stimuli, by decreasing the length of V1 from 176 ms (V1 duration in the original token) to 116 ms, in a step corresponding to 5 F0 periods (60 ms). The reason for this choice was that, in voiced sounds, as shown by Huggins (1968), the minimum step size which could be used without introducing an abnormal pitch period was one period of the fundamental. The F0 periods were removed from the middle portion of the vowel to leave the VC and CV transitions intact. For each stimulus obtained, ten new stimuli were produced by increasing the length of the silent portion of the intervocalic consonant ([p]) from 100 to 235 ms in steps of 15 ms. This yielded to a total of 20 stimuli (2 vowel durations × 10 consonantal durations).

### B. Subjects

Subjects were 20 native speakers (12 males and 8 females) of standard Italian enrolled at Rome University, with no known hearing impairment. Their age ranged from 21 to 29 years.

### C. Experimental tasks

Subjects were asked to identify the stimulus words as geminate or singleton. The experiment was run separately for

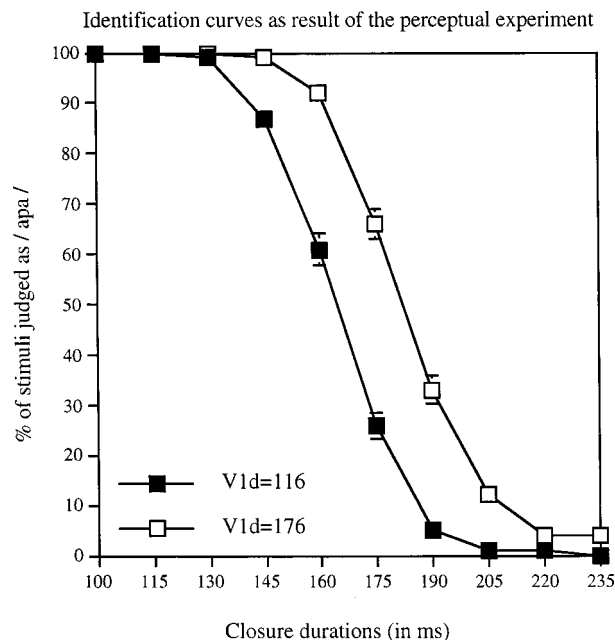


FIG. 8. Identification functions of /apa/ stimuli for V1=116 ms (filled squares) and V1=176 ms (empty squares) versus different consonant closure durations. The identification curves are average functions, where each point plots the mean of 10 observations for each of 20 subjects (also reported is the Standard Deviation around the data points).

the two vowel durations, 176 ms and 116 ms, with the order of presentation balanced across subjects. For each vowel duration, the ten closure durations were presented such that each stimulus was preceded and followed once, by every other. Subjects listened to a total of 101 stimuli in each stage. Between two subsequent stimuli, there was a 2-s silent interval to allow the subjects to answer, a 4-s music interval to minimize any psychoacoustic effect due to the previous stimulus, and a 2-s silent interval to let subjects concentrate on the next stimulus. The subject's task was to identify each stimulus as either /apa/ or /appa/ by typing an appropriate key on a computer keyboard. Subjects listened to a total of 202 stimuli played in random order via a computer program and delivered through good quality headphones. Some stimuli (about 25) were played, for practice, before the experiment start. Subjects were run individually, in sessions lasting about 32 min each.

### D. Results

Results are presented below (Fig. 8) by two identification functions where the varying closure durations are plotted for a fixed V1 duration [V1 duration=116 ms (filled squares), V1 duration=176 ms (empty squares)]. The two identification functions in Fig. 8 are average functions, where each point plots the mean of 10 observations on each of 20 subjects. The behavior of the identification functions of each speaker was similar to the above identification functions (Neter and Wassermann, 1974).

The identification functions reported in Fig. 8 show two different closure duration thresholds at which the perception of a consonant as geminate overcomes the chance. A shorter V1 duration (filled squares) requires a shorter closure duration to allow the perception of a geminate consonant; in this

TABLE II. Summary of the perceptual experiment results. Mean threshold values for the two different V1 durations (116 and 176 ms). The mean threshold differences (positive for each listener), the Standard Deviation (SD), and the result of a two-tailed student *t*-test applied over all the listeners (last row in the table) are also reported.

V1d=116 ms	V1d=176 ms	Differences in ms
Mean 165.8	182.7	16.9
SD 9.5	16.1	9.6
<i>t</i> (19)=7.882, <i>p</i> <0.001		

case, the average closure duration at which the perception of a geminate overcomes the chance was about 165 ms. However, a longer closure duration (about 183 ms) was required when V1 duration was longer (empty squares). The displacement (17 ms on the average) in the perceptual threshold for gemination, when V1 duration was longer, was present for all listeners. Table II summarizes the results of the perceptual test. The difference among the threshold values for the two different V1 duration is positive for all listeners, which reinforces the observation that when V1 is lengthened the closure duration must be longer for perceiving a geminate consonant. Moreover, a two-tailed student *t*-test was applied to the data which show that the differences are statistically significant [*t*(19)=7.882 at level *p*<0.001].

## V. DISCUSSION ON THE PERCEPTUAL DATA

The first result of our perceptual experiment was that stimuli with shorter V1 duration needed a shorter closure duration to allow geminate perception. This leads back to the discussion on the acoustic data, and particularly on the different lengths of the geminate and nongeminate utterances. In the previous discussion, we reported that the differences between the length of geminate and nongeminate utterances were not statistically significant, even though the geminate utterances were, on the average, longer than the single ones. That is not a contradiction with the fact that the closure length of a geminate is almost twice the length of a singleton. Our explanation was that speakers follow a sort of timing (rhythm) which is fixed in duration and depends on the number of syllables in the word: substituting a singleton to a geminate in words (leaving unchanged the number of syllables) does not significantly change the utterance length, an elongated segment (the geminate consonant) being partly compensated by the shortening of another (the vowels). The data from the perceptual experiment showed a similar effect in the listening process; i.e., changing the duration of the vowel in the utterance results, for a fixed consonant length, in a different discrimination threshold between singleton and geminate. Generally, if the vowel is shortened, a singleton could be perceived as a geminate. This confirms our hypothesis of a constant word timing for well-defined conditions of lexical stress and number of syllables. Speakers follow this timing even though the nature of the segments composing the word changes (as in the single versus the geminate case); listeners are sensitive to this timing by balancing changes in the utterance duration with changes of their perceptual durational thresholds (as in the shorter versus longer utterances).

The parameters (V1 and closure duration) which showed acoustic relationships with gemination were the object of a further analysis to define their perceptual role. We found that after a given threshold value which depends on the V1 duration, a consonant is always perceived as geminate. This means that the length of closure duration was used by listeners, as a perceptual cue, to distinguish between geminate and nongeminate consonants. However, it was also an acoustic attribute of the signal (see results on the acoustic data), supporting our hypothesis that closure duration is a distinctive feature for gemination. Since all segment durations which are to act as a cue must be perceived as reference to some baseline, this durational feature was relative rather than absolute. This is a very interesting finding because only durational parameters are found to play a role—whereas duration is often found to be relevant in connection to other acoustic attributes. Several authors pointed out that the pattern of duration of individual phonetic segments conveys information about the linguistic content of an utterance. Miller (1956), for example, has argued that duration should be classified as a distinctive feature for the recognition of vowels. Similarly, Lisker (1957) has shown that, in American English, if the stop-closure is lengthened in the [b] in **ruby** the word is heard as **rupee**, suggesting that it was a durational parameter (the VOT length) that made the distinction between voiced and voiceless stops. Bastian *et al.* (1961), and Liberman *et al.* (1961) reported similar results in studies on other consonants. Finally Klatt (1976) emphasized the role of durational features as a primary perceptual cue in the distinction of long versus short vowels, voiced versus voiceless fricatives, phrase final versus nonfinal syllables, stressed versus unstressed vowels, and presence or absence of emphasis. Moreover, results on the perceptual role of the closure length in the geminate distinction are in agreement with our finding. Pickett and Decker (1960) measured the phoneme boundary between a single and a double /p/ in the pair **topic** and **top pick**. At a speaking rate of six syllables per second, they found the phoneme boundary between **topic** and **top pick** to correspond to a closure duration of about 160 ms. Furthermore, Rochet and Rochet (1995), and Shrotriya *et al.* (1995) showed that for native speakers of different languages, closure duration is perceptually relevant in the geminate versus singleton distinction.

What about the reduction in length of V1? From an acoustic point of view, we expect that V1 before a geminate be shorter than before a nongeminate because in the first case it is part of a closed syllable and in the latter case it is part of an open syllable. However, our perceptual data show that the longer the V1 duration, the greater the crossover value (measured in ms) in perceiving a consonant as geminate. This result was present for all listeners, suggesting that the shortening of V1 could not be attributed only to syllable structure. Our interpretation is that there might be two superimposed effects: one due to syllable structure and one due to the presence of a geminate consonant. This conclusion is not in disagreement with the findings reported by Rochet and Rochet who showed that Italian listeners distinguished between *fato* and *fatto* on the basis of consonant duration but not on the basis of vowel duration. We concluded that V1 is shortened



to balance the abnormal lengthening of the closure in order to keep the rhythm constant, and make the utterance sound natural. This interpretation is also supported by Huggins (1972a) who found that subjects were much more sensitive to changes in vowel duration than to changes in closure duration, and that these changes were perceived as changes in the sentence rhythm, when the duration of the other segments in the utterance remained unchanged.

Informal listening of synthetic stimuli (unpublished data), in which the vowel duration was kept constant while the closure was lengthened up to a value typical of closure of a geminate showed that, in this case, the utterance was heard as broken into two syllables, a long [a] followed by a [pa]. This result, although only preliminary, support [together with the results reported by Kozhevnikov and Chistovich (1965) and by Huggins (1972b)] our hypothesis that the perception of timing in natural speech is based on rhythm rather than on sound segments, and explains why compensation occurs between vowels and consonants if the sentence is to remain temporally fluent.

## ACKNOWLEDGMENTS

Acknowledgment goes to Carmela Marangi, Maria Marinaro, Gian Antonio Mian, and Maurizio Ricciardi. The authors thank Riccardo Rossetti and Armando Vannucci who contributed to the acoustic measurement analyses and to the setup of the perceptual experiment. The authors would like to thank the editor Winifred Strange and two anonymous reviewers for their thorough reviews. This work was funded by Rome University "La Sapienza."

<sup>1</sup>The software UNICE which was used for digitizing the data allows the use of appropriate oversampling factors in order to obtain a correct A/D conversion. Thus if the filter passbands are 12 kHz, one cannot use a sampling rate lower than 40 kHz. To overcome this limitation, UNICE can perform a sampling rate conversion by a factor of two or four (the user sampling rate can be twice or four times lower than the real sampling rate). By using this capability, the available sampling frequency range starts from the half of the filter passband.

- Bastian, J., Eimas, P. D., and Liberman, A. M. (1961). "Identification and discrimination of a phonemic contrast induced by a silent interval," *J. Acoust. Soc. Am.* **33**, 842(A).
- Dillon, W. R., and Goldstein, M. (1984). *Multivariate Analysis* (Wiley, New York).
- Huggins, A. W. F. (1968). "How accurately must a speaker time his articulation," *IEEE Trans. Audio Electroacoust.* **AU-16**, 112–117.
- Huggins, A. W. F. (1972a). "Just noticeable differences for segment duration in natural speech," *J. Acoust. Soc. Am.* **51**, 1270–1278.
- Huggins, A. W. F. (1972b). "On the perception of temporal phenomena in speech," *J. Acoust. Soc. Am.* **51**, 1279–1290.
- Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," *J. Acoust. Soc. Am.* **59**, 1208–1221.
- Kozhevnikov, V. A., and Chistovich, L. A. (1965). "Speech: Articulation and perception." (Moscow-Leningrad), (English Translation: J.P.R.S., Washington D.C., No. JPRS 30543).
- Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., and Bastian, J. (1961). "An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance," *Lang. Speech* **4**, 175–195.
- Lindblom, B., and Rapp, K. (1973). "Some temporal regularities of spoken Swedish," *Papers from the Institute of Linguistics, Stockholm University*, **21**, pp. 1–62.
- Lisker, L. (1957). "Closure duration and the intervocalic voiced–voiceless distinction in English," *Language* **33**, 42–49.
- Miller, G. A. (1956). "The perception of speech," in *For Roman Jakobson* (Mouton, The Hague), pp. 353–360.
- Muljadic, Z. (1972). *Fonologia della lingua Italiana* (II Mulino, Bologna).
- Neter, J., and Wasserman, W. (1974). *Applied Linear Statistical Models* (Irwin, Momewood).
- Pickett, J. M., and Decker, L. R. (1960). "Time factors in perception of a double consonant," *Language and Speech* **3**, 11–17.
- Rochet, L. B., and Rochet, A. P. (1995). "The perception of the single-geminate consonant contrast by native speakers of Italian and Anglophones," in *Proceedings of ICPHS95*, edited by K. Elenius and P. Brand-erud, Vol. 3 (Arne Strömbergs Grafiska, Stockholm), pp. 616–619.
- Rossetti, R. (1994). "Gemination of Italian stops," *J. Acoust. Soc. Am.* **95**, 2874.
- Shrotriya, N., Siva Sarma, A. S., Verma, R., and Agrawal, S. S. (1995). "Acoustic and perceptual characteristics of geminate Hindi stop consonants," in *Proceedings of ICPHS95*, edited by K. Elenius and P. Brand-erud, Vol. 4 (Arne Strömbergs Grafiska, Stockholm), pp. 132–135.
- Vannucci, A. (1994). "Acoustic correlates of distinctive features of Italian stops," *J. Acoust. Soc. Am.* **95**, 2874.

# Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception

Andrew Faulkner and Stuart Rosen

*Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom*

(Received 7 December 1998; revised 17 May 1999; accepted 11 June 1999)

Auditory and audio-visual speech perception was investigated using auditory signals of invariant spectral envelope that temporally encoded the presence of voiced and voiceless excitation, variations in amplitude envelope and  $F_0$ . In experiment 1, the contribution of the timing of voicing was compared in consonant identification to the additional effects of variations in  $F_0$  and the amplitude of voiced speech. In audio-visual conditions only, amplitude variation slightly increased accuracy globally and for manner features.  $F_0$  variation slightly increased overall accuracy and manner perception in auditory and audio-visual conditions. Experiment 2 examined consonant information derived from the presence and amplitude variation of voiceless speech in addition to that from voicing,  $F_0$ , and voiced speech amplitude. Binary indication of voiceless excitation improved accuracy overall and for voicing and manner. The amplitude variation of voiceless speech produced only a small increment in place of articulation scores. A final experiment examined audio-visual sentence perception using encodings of voiceless excitation and amplitude variation added to a signal representing voicing and  $F_0$ . There was a contribution of amplitude variation to sentence perception, but not of voiceless excitation. The timing of voiced and voiceless excitation appears to be the major temporal cues to consonant identity. © 1999 Acoustical Society of America. [S0001-4966(99)01410-1]

PACS numbers: 43.71.Es, 43.71.Ma, 43.66.Ts [JMH]

## INTRODUCTION

The perceptual role of temporal structure in speech has until recently been rather neglected in comparison to spectral structure (Rosen, 1992; van Tasell *et al.*, 1987, 1992). Temporal speech information is likely to have special significance for hearing-impaired listeners, for users of cochlear implants, and more generally in listening conditions where noise masks spectral structure. Both electro-cochlear stimulation (Shannon, 1993) and acoustic stimulation to even profoundly hearing-impaired listeners (Faulkner *et al.*, 1992; Rosen *et al.*, 1990) are typically perceived with sufficient temporal resolution for temporal speech information to be processed. Cochlear hearing impairment of moderate or greater degree is typically associated with significant degradation of frequency selectivity (e.g., Moore, 1996). At frequencies where the hearing loss is profound, frequency selectivity can be completely absent (Faulkner *et al.*, 1990). The partial loss of spectral detail appears to have rather slight effects on the perception of speech at least in quiet (Baer and Moore, 1993; Shannon *et al.*, 1995). This result suggests that temporal cues and changes of gross spectral shape may be of greater significance in speech perception than has often been thought.

The notion that segmental and supra-segmental speech information needs to be clearly represented by speech perceptual prostheses is well-established. The present studies are largely concerned with the contributions from different aspects of temporal speech information to consonant perception in conditions where spectral structure is eliminated.

They have both a theoretical aim in relation to accounts of speech temporal structure (e.g., Rosen, 1992) and a practical aim in identifying those aspects of temporal structure that need to be effectively transmitted through a hearing aid or cochlear implant.

### A. Relation to previous studies

Many studies have examined the contribution of temporally coded cues to the audio-visual perception of connected speech (e.g., Breeuwer and Plomp, 1986; Grant *et al.*, 1985, 1991; Risberg and Lubker, 1978; Rosen *et al.*, 1981; Waldstein and Boothroyd, 1994). However, studies of sentence-level materials provide little insight into the phonetic information conveyed by temporal cues. Segmental perception from temporal cues, which can be analyzed to be informative of temporal contributions to phonetic features, has received much less attention.

Studies of consonant identification from temporally coded auditory information have rarely included both auditory and audio-visual conditions. Here we have studied consonant identification in both auditory and audio-visual conditions and also from purely visual input, so that the contributions of temporal structure can be observed both with and without partially complementary visual cues. Moreover, such studies of consonant perception have not sought to isolate components of temporal structure in relation to the acoustic-phonetic structures of speech (Baker and Rosen, 1994; Breeuwer and Plomp, 1986; Rosen, 1992; van Tasell *et al.*, 1987, 1992). In particular, the role of periodic and aperiodic excitation has largely been ignored. Furthermore,

in consonant identification, amplitude variation has never been dissociated from simple duration cues, although this has been done for connected speech (Breeuwer and Plomp, 1986; Grant *et al.*, 1985).

## B. The temporal structure of speech

Temporal information has several component elements which have been associated with both segmental and supra-segmental speech information. Rosen (1992) proposed a three-way classification of temporal components by rate. This is linked to Fourcin's (1977, 1990) analysis of speech in terms of acoustic speech patterns that relate to putatively significant productive and receptive features. Low rate temporal information (below about 50 Hz) conveys the presence or absence of acoustic excitation and the amplitude variation of periodically and aperiodically excited speech. This factor, termed *envelope* by Rosen (1992), is correlated with the effort underlying speech excitation, the degree of constriction of the vocal tract, and velar closure. Temporal information at rates between 50 and a few hundred Hz represents the periodicity or aperiodicity of excitation, and voice fundamental frequency ( $F_0$ ). This factor was termed *periodicity* in Rosen's classification. A still higher rate component of temporal speech information is that contained in the *fine structure* of the speech pressure waveform. This contains information related to vocal tract resonances. There is no strong evidence that speech spectral structure can be perceived from temporal fine structure, and it is not considered in the present study.

These classes of information, although broadly assigned to temporal rate ranges, cannot be properly separated by modulation rate. For example, both  $F_0$  and  $F_1$  may affect temporal modulations in the 200–500 Hz region, and at burst release, the onset of voiceless excitation can be extremely rapid. This conceptual framework nevertheless allows a broad classification of the production-related acoustic pattern information that exists over the speech modulation rate spectrum.

Temporal information at low and moderate rates has particular significance in speechreading, because it is largely uncorrelated with visible oral correlates of place of articulation, and is hence perceptually complementary to speechreading (Summerfield, 1987). Spectral speech information, conversely, is broadly correlated with visible information from speechreading.

## C. Phonetic information in temporal structure

### 1. Patterning of excitation

Much of the voicing and manner of articulation information in speech has acoustic correlates in the patterning in time of quasi-periodic laryngeal excitation, silence, and aperiodic excitation. Hence it would be expected that temporal cues can signal such information from the timing of the presence and absence of excitation and from its periodicity or aperiodicity. The extent to which the timing of voiced excitation contributes to consonant identification is examined in experiment 1. The additional consonantal information available from the timing of voiceless excitation is examined in

experiment 2. The significance of a cue to the timing of voiceless excitation in connected speech is addressed in experiment 3.

## 2. Amplitude variation

The degree of oral constriction and the opening of the nasal tract both affect the amplitude envelope of voiced speech. Hence, the variation of speech amplitude might be expected to provide cues to manner of articulation (e.g., Shinn and Blumstein, 1984). The envelope of voiceless speech is influenced both in its duration and its amplitude by the place of articulation of voiceless fricatives and voiceless plosives. Speech amplitude is also affected by the amplitude variations of both periodic and aperiodic excitation that follow from variation in sub-glottal pressure. Such variation is broadly associated with prosodic features and is not expected to be significant in consonant identity. Experiment 1 examines the role of the amplitude variation of voiced speech in consonant identification. Experiment 2 includes conditions which allow an assessment of the contribution of voiceless speech amplitude variation in consonant identification. Experiment 3 addresses the role of amplitude variation in the audio-visual perception of connected speech.

## 3. Fundamental frequency

The perceptual significance of voice  $F_0$  is primarily supra-segmental. However,  $F_0$  is affected by changes in the acoustic impedance of the vocal tract resulting from differing degrees of occlusion during consonant articulation (Fant, 1973, p. 80). Voiced consonants may thus contain  $F_0$  variation that carries some salient information related to manner of articulation. Experiment 1 examines this hypothesis.

## D. Interpretation of amplitude envelope

Investigators have used a number of different definitions and measurement methods for amplitude envelope. The perceptual contribution of amplitude envelope depends both on the speech frequencies over which it is measured and on the degree of smoothing applied to the extracted envelope (e.g., Grant *et al.*, 1991). Where the amplitude envelope smoothing filter extends into the voice  $F_0$  range, typical methods of envelope extraction lead to the periodicity of  $F_0$  being included in the envelope signal. In order to distinguish periodicity from other components of the amplitude envelope, we have used an envelope smoothing filter that is below the  $F_0$  range, and have derived periodicity information from larynx activity.

When amplitude envelope is measured directly from the speech signal, the choice of bandwidth from which amplitude envelope is measured determines whether voiceless speech energy is included. Rather than take this purely acoustic approach to separate voiced and voiceless excitation, we have used a reference signal from laryngeal activity to distinguish voiced and voiceless excitation. Henceforth, we use the term amplitude variation to refer to low rate amplitude fluctuations up to around 50 Hz. For voiced speech, amplitude envelope has been derived from frequencies up to

3 kHz. For voiceless speech components, amplitude envelope was measured here over the frequency range 3–10 kHz.

## I. TEMPORAL INFORMATION IN CONSONANT IDENTIFICATION

### A. Experiment 1: Segmental information from voicing pattern, amplitude variation, and $F_0$ variation

This first study examined the contributions to consonant identification of the gross timing, amplitude variation, and  $F_0$  variation of voiced components of speech. Because of the segmental correlates of these three temporal factors, each may be expected to have a possible role in cueing consonantal features and identity. This experiment is designed to evaluate their relative significance both in audio-visual and purely auditory perception using auditory signals with an invariant spectral envelope.

#### 1. Method

*a. Conditions.* The contributions of these temporal components in auditory and audio-visual consonant identification are assessed here by differences in performance across nine different presentation conditions. Four auditory signals were used, both with and without visual lipreading information. A purely visual lipreading condition, **L**, was also included. The gross timing of voicing was represented by signal **V**, a pulse-train of fixed frequency and amplitude, which was present only during vocal fold vibration. Signal **V(A)** differed from signal **V** in carrying amplitude envelope variation, derived from the original speech.  $F_0$  variation without amplitude variation was present in signal **F**, where a fixed-amplitude pulse-train during voicing followed the speaker's  $F_0$ . Finally, signal **F(A)** varied in both frequency and amplitude according to  $F_0$  and the amplitude envelope of voiced speech, as well as indicating the timing of voicing.

*b. Stimulus processing.* Fundamental frequency and the timing of laryngeal excitation were derived from an electro-laryngograph signal that had been recorded with the speech signal. This processing made use of a Voiscope™ (Laryngograph Ltd.<sup>1</sup>) which generated a brief logic pulse at each vocal fold closure, and a flip-flop logic signal that indicated voicing. The voicing detector was used to gate a pulse-train signal. For voiced speech the rate of this pulse train was fixed at 200 Hz for signals **V** and **V(A)**. For signals **F** and **F(A)**, the pulse rate was controlled cycle-by-cycle from the signal generated at each larynx closure so as to follow the voice fundamental frequency.

Amplitude envelope was derived by full-wave rectifying the speech signal after 3-kHz low-pass filtering (Kemo VBF/3: 48 dB/octave) and then by smoothing the rectifier output with a low-pass filter having its  $-3$ -dB point at 22.5 Hz, and subsequent attenuation of 18 dB at 50 Hz and 53 dB at 80 Hz (Kemo VBF 8<sup>2</sup>).

Finally, the pulse signal was multiplied by the amplitude envelope and then low-pass filtered at 400 Hz (Kemo VBF/14: 18 dB/octave). The processing of voicing and  $F_0$  involved an audio delay of one period. The amplitude envelope

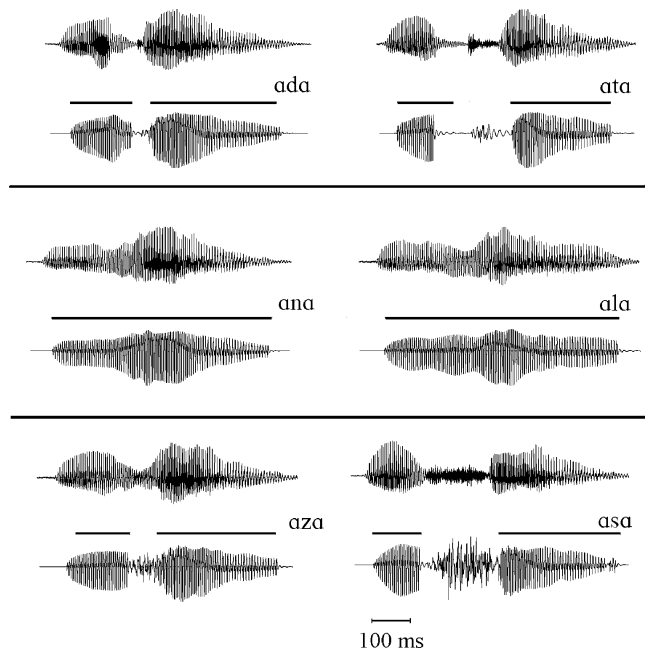


FIG. 1. Speech (upper) and stimulus (lower) waveforms for six alveolar VCVs. The stimulus waveform is for condition **F(A)+N(A)** of experiment 2. Signal **F(A)** used in experiment 1 is similar except that the signal level is zero where voicing was not detected. The lines above the stimulus waveforms mark the timing of voiced excitation as detected from the Laryngograph signal.

measure involved a delay of 19.5 ms arising from the impulse response of the smoothing filter. Speech input and output signals illustrating the result of this processing are shown in Fig. 1.

Speech materials comprised each of the 24 English consonants between the vowel /a/. Five video-recorded lists from a female speaker of standard southern British English were used, comprising in total ten distinct tokens of each consonant, with two in each of the lists. Each list used a different pseudo-random ordering. The stimuli were originally recorded onto Hi Band U-matic video, with the speech signal and an electro-laryngograph signal on the two FM audio channels. These recordings were then edited and copied to S-VHS using the FM audio tracks. The stimuli presented in the experiment itself were copied to VHS video tapes, using the FM audio tracks to record all five lists with each of the auditory signals.

#### 2. Subjects

Five normally hearing subjects aged between 18 and 30 with normal or corrected-to-normal vision took part. All were native speakers of British English. They were paid for their participation.

#### 3. Procedure

All the experiments reported here took place in an acoustically isolated room. The video image was presented on a 36-cm Panasonic color monitor. The audio signal was presented free-field at approximately 65 dBA through either a Rogers A75 amplifier and Rogers LS3/5a loudspeaker or a



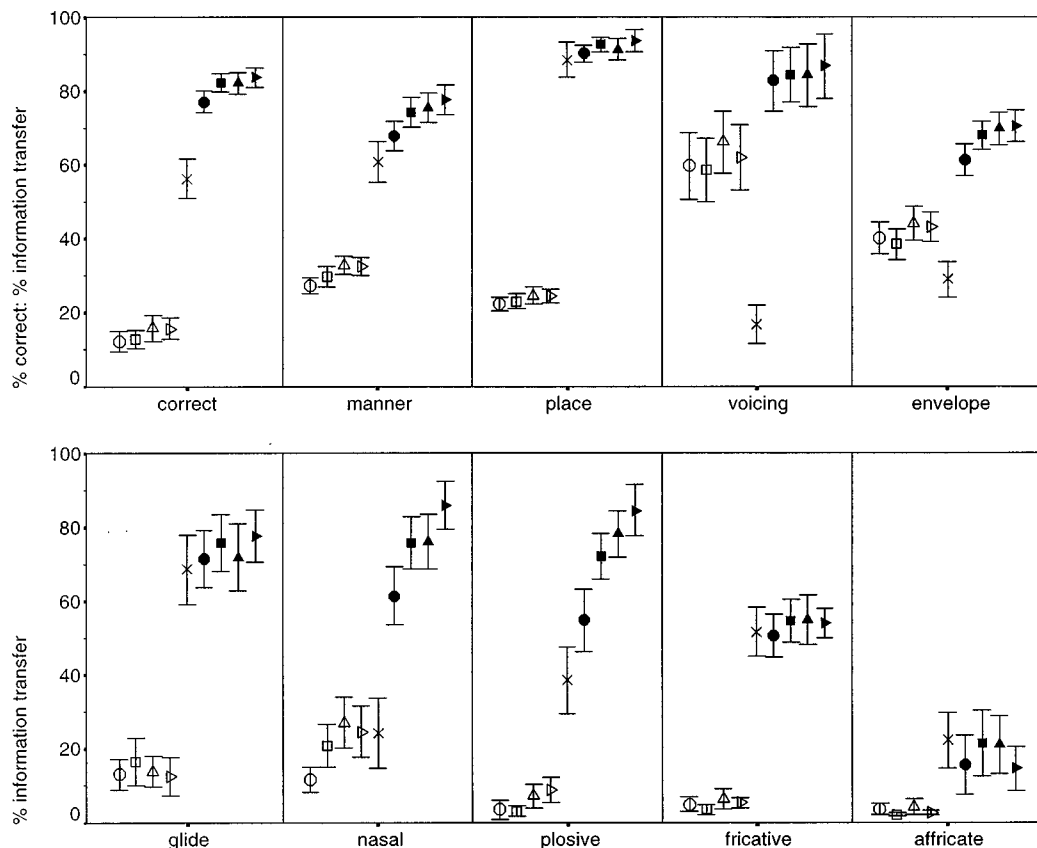


FIG. 2. Upper panel: percent correct and percent information transfer for manner, place, voicing, and envelope features. Lower panel: percent information transfer for binary manner features. Data from experiment 1, five subjects. Conditions are shown as V:  $\circ$ ; V(A):  $\bullet$ ; F:  $\triangle$ ; F(A):  $\blacktriangle$ ; L:  $\times$ . Audio-visual conditions have the same symbol as the corresponding auditory conditions, but filled. Error bars show  $\pm 2$  standard errors over 30 data points (subjects  $\times$  sessions).

Yamaha P2050 amplifier and a QUAD PRO-63 loudspeaker. Subjects sat approximately 1.5 m from the monitor and loudspeaker.

The first of the five lists of consonants was reserved for training, which was provided before the first test list in each condition. Training was performed by the experimenter providing verbal feedback to subjects after they had attempted to identify each one of the 48 training stimuli. The eight audio and audio-visual conditions were presented in blocks of four test lists for which the audio signal remained the same. Five such blocks were presented in total, interleaved with five blocks of two lists in condition L, with the blocks in a randomized order. The order of auditory and audio-visual conditions within blocks was counterbalanced, and the ordering of test lists in each block was randomized to minimize learning of the stimulus order.

Subjects responded in writing on a printed response sheet during a blank time interval following each stimulus. In total, 10 lists of 48 consonants were administered to each subject in each condition.

#### 4. Results

Consonant identification data from each test list were analyzed to give overall proportion correct scores, together with information transfer measures (Miller and Nicely, 1955) for the binary voicing feature, for place of articulation (distinguishing bilabial, labiodental, dental, alveolar, palatal, ve-

lar, and pharyngeal place), for manner (plosive, affricate, fricative, nasal, or glide), and for the modified envelope feature defined by van Tasell *et al.* (1992). The envelope feature distinguishes four classes of consonant: voiceless plosives; voiced plosives and fricatives; voiceless fricatives; nasals and glides. Subjects showed improving performance over the early sessions. The analyses presented here are based on the last six lists in each condition, which showed more stable levels of performance over sessions.

Preliminary repeated measures analyses of variance for the overall accuracy and feature measures used factors of condition, test session, and subject. These revealed significant effects of condition, session, and subject for all measures. There were no significant interactions between session and condition and the session factor was henceforth ignored. A second series of repeated measures ANOVAs of overall correct and feature scores that excluded condition L treated the experiment as a  $2 \times 2 \times 2$  factorial design, with factors of amplitude variation,  $F_0$  variation, and visual information, in addition to subject. Between session variability was included in the residual error term.<sup>3</sup>

Mean scores in each condition are shown in Fig. 2. Auditory only scores were always rather low except for voicing and envelope features. Visual information had a highly significant effect on all of the measures.  $F_0$  variation was a significant main effect for overall accuracy [ $F(1,8) = 15.9, p = 0.004$ ], for manner [ $F(1,12) = 20.3, p = 0.001$ ],

voicing [ $F(1,228)=4.0, p=0.047$ ], and envelope information [ $F(1,228)=20.1, p=0.001$ ]. The effects of  $F_0$  variation did not interact with any other factor.

Amplitude variation showed a significant main effect for both overall accuracy [ $F(1,216)=5.17, p=0.024$ ] and for manner information [ $F(1,12)=6.41, p=0.026$ ]. However, there was a significant interaction of amplitude variation with the presence of visual information for overall accuracy [ $F(1,216)=4.18, p=0.042$ ] and the envelope feature [ $F(1,228)=5.16, p=0.024$ ]. This interaction also came close to significance for manner information [ $F(1,212)=3.34, p=0.069$ ]. Hence, further analyses were carried to test the effect of amplitude variation for auditory and audio-visual conditions separately. In auditory conditions, the effect of amplitude variation was never significant for these measures ( $p$  was always greater than 0.3). However, in audio-visual conditions, these analyses were able to reveal small but significant effects of amplitude variation on overall accuracy [ $F(1,112)=9.4, p=0.003$ ], manner [ $F(1,112)=9.4, p=0.003$ ], and envelope information [ $F(1,112)=4.84, p=0.030$ ].

The effects of  $F_0$  and amplitude variation as they act on consonant manner contrasts have been examined in more detail by computing information transfer scores for binary manner features of nasal, plosive, fricative, affricate, and glide. These are included in Fig. 2. Affricate, fricative, and glide features proved to be affected only by the presence or absence of visual information. Nasal and plosive features were affected by visual information and also by amplitude and  $F_0$  variations. The presence of  $F_0$  variation significantly increased nasal information across auditory and audio-visual conditions [ $F(1,4)=11.1, p=0.029$ ], while  $F_0$  variation showed an interaction with visual information for the plosive feature. Further analysis showed a significant effect of  $F_0$  on plosive information in audio-visual conditions only [ $F(1,112)=40.3, p<0.001$ ]. Amplitude variation showed significant interactions with visual information for both nasal and plosive features. In auditory conditions, there was no significant effect of amplitude variation, while in audio-visual conditions, there were significant amplitude effects for both nasal [ $F(1,112)=11.7, p<0.001$ ] and plosive information [ $F(1,112)=17.9, p<0.001$ ].

## 5. Discussion

The results indicate that the bulk of the auditory information in these stimuli comes from the timing of voiced excitation.  $F_0$  variation had a significant, but modest effect, increasing overall accuracy by just over 3%. This followed from the improved perception of nasal and plosive manners. The amplitude variation of voiced speech at rates of up to 50 Hz led to significant increments in accuracy (again of about 3%) and manner information in audio-visual conditions, but had no effect in purely auditory conditions.

## B. Experiment 2: Segmental information from voiceless excitation

A second study addressed the role of voiceless frication and the amplitude envelope of voiceless speech. Here there

were three different acoustic signals each presented with and without lip-reading.

## 1. Conditions

The **F(A)** signal used in experiment 1, encoding timing of voicing,  $F_0$  and voiced speech amplitude, was retained as a reference condition and was compared to two new signals that represented voiceless excitation. In condition **F(A)+N**, the **F(A)** signal was combined with a fixed-level aperiodic noise that was present during purely voiceless excitation. The contribution of amplitude variation in voiceless speech was examined by comparing **F(A)+N** with **F(A)+N(A)**, in which the noise component was modulated by the amplitude envelope of voiceless speech.

## 2. Stimuli

Voiced speech was processed as before. Voiceless excitation was detected by a threshold applied to the output of a spectral balance circuit that compared the amount of energy above and below 3 kHz in the speech signal. Voiceless excitation was represented by a 400-Hz low-pass filtered noise, only in the absence of voicing as detected from the recorded electrolaryngograph signal. Amplitude envelope for voiceless speech was derived by full-wave rectifying the speech signal after a 3-kHz high-pass filter (Kemo VBF/3: 48 dB/octave), and smoothing the result as for voiced speech using a low-pass filter with its  $-3$ -dB point at 22.5 Hz (Kemo VBF/8, as experiment 1). The pulse and noise signals were added together after multiplication by the separately extracted amplitude envelopes of voiced and voiceless speech.

## 3. Subjects and procedure

Five new subjects with normal hearing and normal or corrected to normal vision took part and were paid for their services. Procedures were the same as those for experiment 1 and again ten test lists of 48 stimuli were presented in each condition.

## 4. Results

Data from the last six test lists in each condition only were analyzed. Results are shown in Fig. 3. As for experiment 1, ANOVAs by condition, session, and subject were carried out for each measure. Test session had a significant effect on most scores, but there were no significant interactions between condition and session. The main analyses used a partial factorial design, with factors of visual information, a cue to the presence of voiceless excitation [comparing signals **F(A)** and **F(A)+N**], and the encoding of amplitude variation for voiceless speech [comparing signals **F(A)+N** and **F(A)+N(A)**]. These revealed highly significant ( $p<0.001$ ) main effects of visual information for overall accuracy and for each of the features of manner, place, voicing, and envelope. Cueing the presence of voiceless excitation led to significant increases in all of these measures except for place information; overall accuracy,  $F(1,116)=30.7, p<0.001$ ; manner,  $F(1,162)=84.4, p<0.001$ ; voicing,  $F(1,17.2)=11.35, p=0.004$ ; envelope,  $F(1,4)=55.67, p=0.002$ .

The encoding of amplitude variation for voiceless speech showed a main effect only for place information, where there was a small but significant increase in scores in

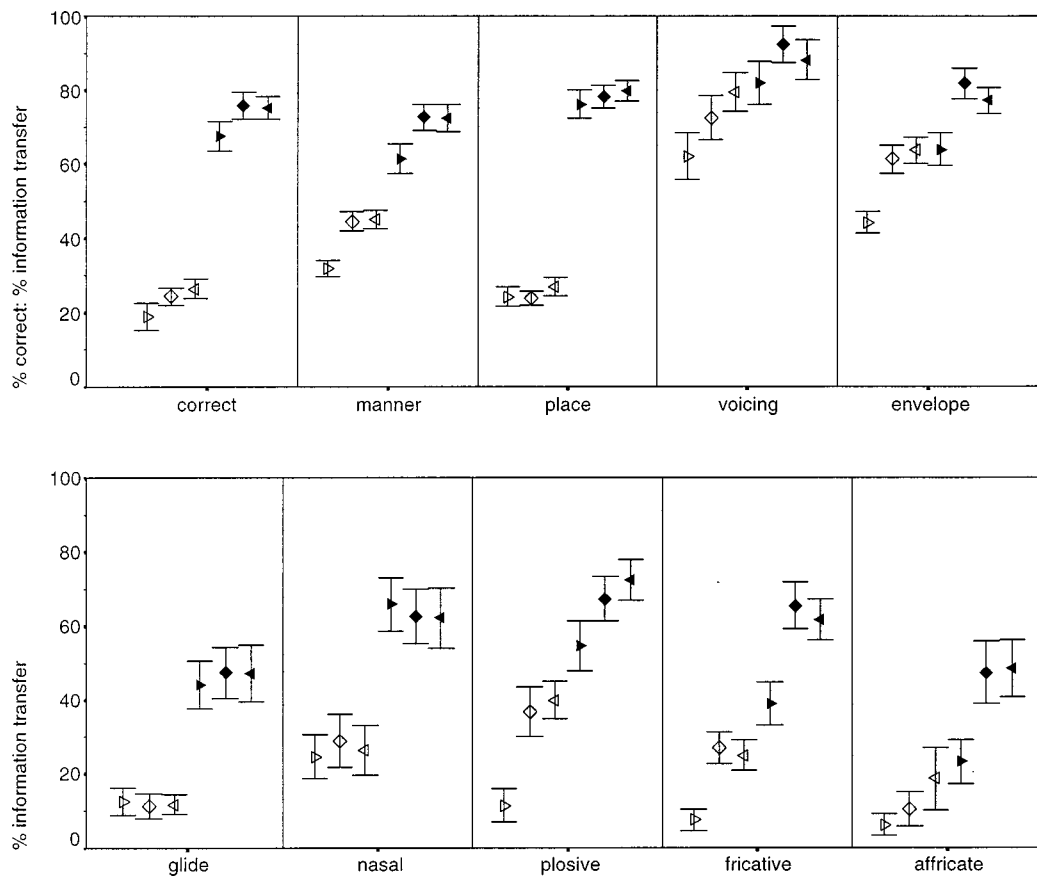


FIG. 3. Upper panel: percent correct and percent information transfer for manner, place, voicing, and envelope features. Lower panel: percent information transfer for binary manner features. Data from experiment 2, five subjects. Conditions are shown as **F(A)**:  $\triangleright$ ; **F(A)+N**:  $\diamond$ ; **F(A)+N(A)**:  $\triangleleft$ . Audio-visual conditions have the same symbol as the corresponding auditory conditions, but filled. Error bars show  $\pm 2$  standard errors of 30 data points.

both auditory and audio-visual conditions in the presence of this information [ $F(1,158)=4.75, p=0.031$ ]. There were significant interactions between voiceless speech amplitude variation and visual information for voicing [ $F(1,158)=5.93, p=0.016$ ] and envelope features [ $F(1,158)=5.66, p=0.019$ ]. Both were crossover interactions, reflecting an increase in scores with amplitude variation in auditory conditions but a decrease in audio-visual conditions.

Binary manner features (glide, nasal, plosive, fricative, and affricate: see Fig. 3) all showed highly significant increases with the presence of visual information ( $p < 0.006$ ). As in experiment 1, glide and nasal features were unaffected by differences in the auditory signals. Encoding the presence of voiceless excitation led to significant increases in information transfer for plosive [ $F(1,166)=54.6, p < 0.001$ ], fricative [ $F(1,10.32)=56.8, p=0.001$ ], and affricate features [ $F(1,158)=19.3, p < 0.001$ ]. For plosive and affricate features, this factor also showed a significant interaction with the presence of visual cues. For the plosive feature, this interaction reflects a somewhat larger increase in scores in auditory conditions than audio-visually. In the case of the affricate feature, however, there was little effect in auditory conditions of the voiceless cue, but a large effect audio-visually. The only binary manner feature affected by the amplitude variation of voiceless speech was the plosive feature [ $F(1,17.6)=5.61, p=0.029$ ].

## 5. Discussion

The main outcome of this experiment is that a temporal indication of the presence of aperiodicity representing voiceless excitation leads to significant increments in overall performance in consonant identification. This derives from improved voicing and manner perception. The manner information gained from the signaling of aperiodic excitation can be associated with binary manner features marking affricate, fricative, and plosive consonants, for which voiceless excitation would be expected to be perceptually salient. The amplitude variation of voiceless speech had few effects, but did lead to increased place and plosive information. The effect on place information presumably derives from the use of burst and frication intensity cues.

## C. Experiment 2A

This subsidiary experiment was performed for completeness and to allow us to examine a condition representing the timing of both voiced and voiceless excitation in the absence of amplitude variation. This experiment measures the effect of a cue to the presence of voiceless excitation in the absence of amplitude cues, and the effect of amplitude variation when the timing of both voiced and voiceless speech excitation is cued.

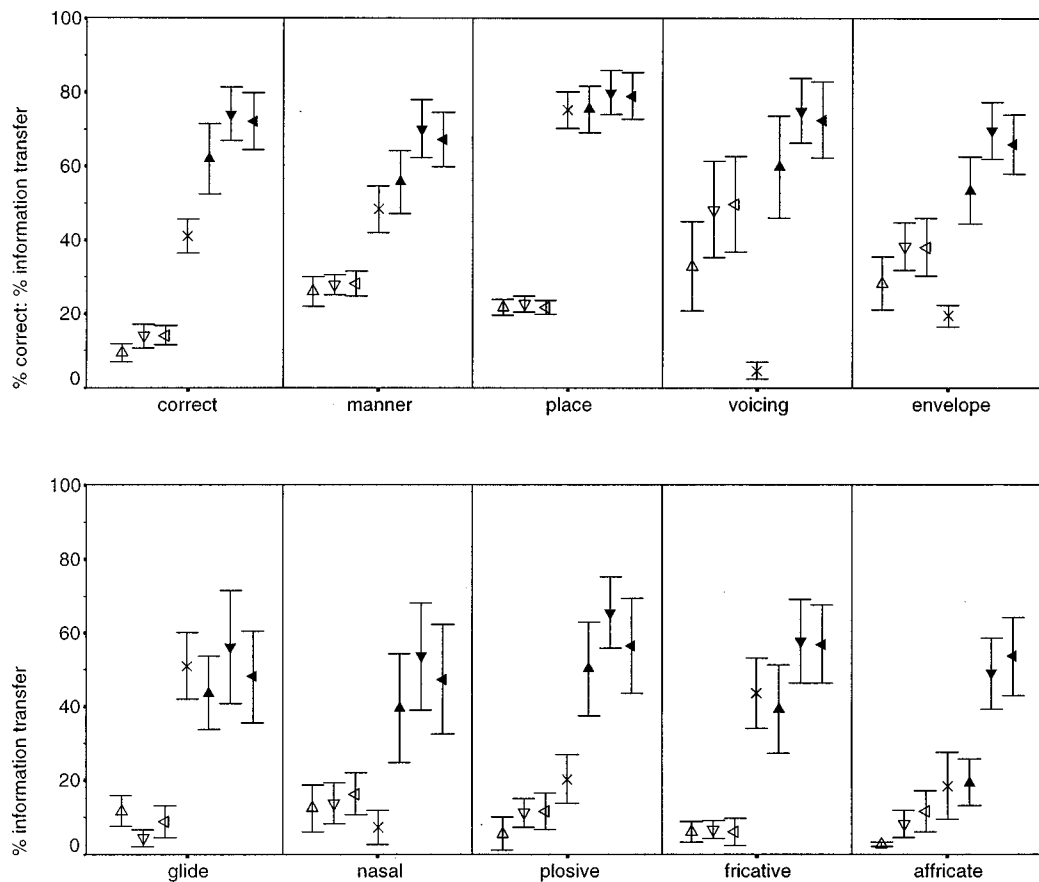


FIG. 4. Upper panel: percent correct and percent information transfer for manner, place, voicing, and envelope features. Lower panel: percent information transfer for binary manner features. Data from experiment 2A, three subjects. Conditions are shown as **F**:  $\Delta$ ; **F+N**:  $\nabla$ ; **F(A)+N(A)**:  $\triangleleft$ ; **L**:  $\times$ . Audio-visual conditions have the same symbol as the corresponding auditory conditions, but filled. Error bars show  $\pm 2$  standard errors over 18 data points.

### 1. Method

Experiment 2A used the same methods as previously, and three naive observers. Seven conditions were used, involving lipreading alone as condition **L**, and three acoustic signals both with and without lipreading: **F**, **F+N**, and **F(A)+N(A)**. Condition **F+N** encoded the presence of voiced or purely voiceless excitation and  $F_0$  with no amplitude variation, and had not been used before.

### 2. Results and discussion

As before, the last 6 of 10 test lists of 48 consonants were analyzed for each subject. Results for each measure and condition are shown in Fig. 4. One subject showed lower performance overall than the subjects of experiments 1 and 2, and this resulted in a wider range and lower mean scores than previously. The analyses proceeded similarly to that for experiment 2. A series of ANOVAs using factors of condition, session, and subject showed that here neither test session nor its interaction with condition was a significant factor for any measure. The main analyses<sup>4</sup> excluded condition **L** and employed a factorial approach to examine the effects of the presence of visual cues, the presence of a cue to voiceless excitation (comparing signals **F** and **F+N**), and the representation of amplitude envelope [comparing signals **F+N** and **F(A)+N(A)**]. As previously, visual cues led to highly significant increases in overall accuracy and manner, place, voicing, and envelope features. As in experiment 2, the pres-

ence of the voiceless excitation cue produced significant effects for all measures except place information; overall accuracy,  $F(1,2)=69.13$ ,  $p=0.014$ ; manner,  $F(1,2)=38.19$ ,  $p=0.025$ ; voicing,  $F(1,2)=24.03$ ,  $p=0.039$ ; envelope,  $F(1,2)=137.27$ ,  $p=0.007$ . For manner, the effect of the voiceless cue interacted with the presence of visual information [ $F(1,2)=141.7$ ,  $p=0.007$ ] and was greater in audio-visual conditions. This interaction term was also significant for place information [ $F(1,2)=31.8$ ,  $p=0.030$ ], where the voiceless cue had no detectable effect in auditory conditions, but led to a significant increase in audio-visual scores [ $F(1,2)=138.4$ ,  $p=0.007$ ].

There were no main effects of amplitude variation. However, as in experiment 1, there were significant interactions involving amplitude variation and visual information. Here the interactions occurred for overall accuracy [ $F(1,2)=334.1$ ,  $p=0.003$ ], manner information [ $F(1,2)=205.7$ ,  $p=0.005$ ], voicing information [ $F(1,2)=23.91$ ,  $p=0.039$ ], and the envelope feature [ $F(1,2)=20.74$ ,  $p=0.045$ ]. Owing to the small number of subjects here, ANOVA has insufficient power to detect the small effects of amplitude variation that these interactions suggest may be present here in audio-visual conditions.

As in experiment 2, information transfer for binary affricate, fricative, and plosive manner features (Fig. 4) were significantly affected by the acoustic signal. Both plosive



[ $F(1,2)=334, p=0.003$ ] and affricate [ $F(1,2)=355, p=0.003$ ] information were significantly increased by the cue to the presence of voiceless excitation. For the affricate feature this effect was, as in experiment 2, much larger in audio-visual conditions. Fricative information also tended to be increased by the voiceless cue, but here only in audio-visual conditions.

#### D. Discussion

Experiment 2A replicates the finding of experiment 2 that the timing of voiceless excitation can contribute significantly to consonant identity. It shows the same pattern of interaction as previously between visual cues and amplitude variation, whereby audio-visual scores for overall accuracy and manner information are typically higher when amplitude variation was encoded.

### II. EXPERIMENT 3: AUDIO-VISUAL SENTENCE PERCEPTION

The third experiment was performed to examine the contribution of these same temporally coded information elements to the audio-visual perception of connected speech. Here, the perception of supra-segmental prosodic information is likely to contribute to performance through lexically and syntactically based stress. While the principal source of prosodic information is generally thought to be intonation, the acoustic correlates of stress also include increased amplitude. Hence it may be expected that amplitude variations are of more importance in connected speech than for consonant identification. Breeuwer and Plomp (1986) have already shown that audio-visual perception of connected speech was significantly improved when amplitude envelope variation was added to a signal that conveyed  $F_0$  information, with syllable correct scores being around 13% higher when amplitude variation was included.

One purpose of this final experiment was to ensure that the signal processing methods used here did not have some artifact that reduced the contribution of speech amplitude variation to consonant identification. In addition, since consonant identification in experiments 2 and 2A was significantly enhanced by the availability of a cue to the presence and duration of voiceless excitation, this final experiment also provides a test of the significance of this factor in sentence-level perception. In the present experiment, the role of  $F_0$  variation was not examined in comparison to a fixed-frequency signal, as the limited number of sentence test lists constrained the number of conditions that could be employed. However, other studies, notably that of Waldstein and Boothroyd (1994), have clearly established that  $F_0$  variation plays a major role in audio-visual sentence perception.

#### A. Experimental method

##### 1. Conditions

The three auditory supplements employed were **F**, **F(A)**, and **F(A)+N(A)** as above. The fourth condition was lipreading (**L**) with no acoustic signal.

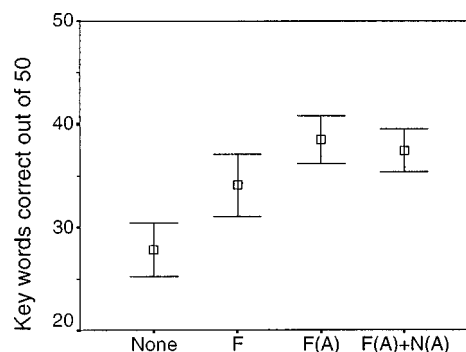


FIG. 5. Mean number of key words identified by condition. The error bars show  $\pm 2$  standard errors over 30 data points (subject  $\times$  test session).

#### 2. Speech processing

Speech processing was essentially the same as previously, except that the pulse and carrier signals had a 5 kHz rather than a 400-Hz bandwidth.

#### 3. Speech materials

Speech materials were taken from the UCL EPI audio-visual recording of the BKB sentences (Foster *et al.*, 1993). The speaker was an adult female. Each of the 21 sentence lists comprises 16 sentences, each with 4 or 5 “key” content words that are scored for correctness.

#### 4. Subjects

Six normally hearing subjects took part. All were Speech Science students at UCL aged between 21 and 35. All had normal hearing and normal or corrected-to-normal vision. None had taken part in the previous experiments.

#### 5. Procedure

Each subject first received an unscored practice sentence list in each of the four conditions. Because of the limited number of available lists, the same list was used for practice in each condition. This was List 1 which, according to Foster *et al.* (1993), differs most in difficulty from the other 20 lists. Subjects subsequently received five test sessions comprising one test list in each of the four conditions. The order of the four conditions was counterbalanced over five testing sessions. The subjects were split into two groups, and two counterbalanced orders were used to distribute lists between the test conditions.

#### B. Results

Each sentence list was scored according to the “key-word tight” (KW-T) procedure (Bamford and Wilson, 1979). This was preferred to the “key-word loose” method since it requires the key words to be identified exactly, and hence was expected to reflect more accurately the perception of detailed phonetic information such as the presence of voiceless /s/ in indicating plurality. The group results are shown in Fig. 5.

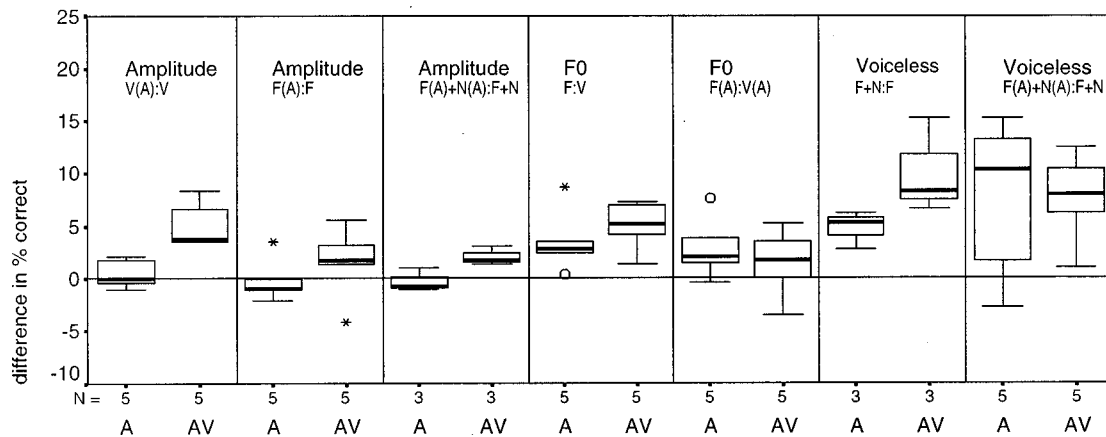


FIG. 6. Differences in % correct scores produced by the addition of amplitude variation (leftmost three panels),  $F_0$  variation (central two panels), and a cue to the presence of voiceless excitation (rightmost two panels). The legends A and AV on the horizontal axis indicate auditory and audio-visual conditions, respectively. The compared auditory signals are indicated in each panel. The plots show the interquartile range (box), the median (bar), and the extreme values (whiskers) of the range of differences in each case. Each data set is from a single experiment. The number of subjects in each data set is also shown ( $N$ ).

### 1. Analysis

Because scores from some subjects approached the upper bounds of the test in some conditions, an arcsine transformation was applied to the data prior to a repeated-measures analysis of variance. Comparisons between conditions were made using Tukey HSD tests. As expected, all three of the audio-visual conditions showed significantly higher scores than the visual only condition. Scores in conditions  $L+F(A)$  and  $L+F(A)+N(A)$  did not differ significantly, while both showed a significantly higher score than condition  $L+F$ , the difference being around 7.5 words out of 50. There was also a significant practice effect over sessions [ $F(4,20) = 10.55, p < 0.001$ ], but no significant interaction of condition and session.

### 2. Discussion

In showing a significant contribution of amplitude envelope variation, these results are consistent with others in the literature. Since we have found only a small contribution of amplitude variation at the level of segmental (consonant) perception, we presume the more substantial effect here is partially due to supra-segmental correlates of amplitude variation.

A contribution of voiceless excitation information is not apparent here, despite it being consistently significant in consonant identification. Presumably the contribution it can make in consonantal manner perception and the enhancement of voicing contrasts is made less significant here by the availability of syntactic and lexical context.

## III. SUMMARY AND CONCLUSIONS

### A. Perceptual contributions of elements of temporal structure

As previous studies also show, auditory alone scores in consonant identification from spectrally invariant signals are always rather low (e.g., Rabinowitz *et al.*, 1992). In audio-visual conditions, however, these auditory signals lead to quite substantial improvements in performance over purely visual presentation. Several clear findings that hold in both

audio-visual and auditory consonant identification have emerged. First, the bulk of the useful temporal information comes from the timing of voiced excitation. Second, the timing of voiceless excitation provides an additional significant contribution that provides cues to both voicing and manner of articulation. Smaller contributions can be associated with  $F_0$  variation, which provides a subtle but significant cue to consonant manner, and in audio-visual conditions only, with amplitude variation.

The importance of  $F_0$  information in the perception of connected speech is well-established. Amplitude variation was confirmed to have significance here, as other studies of audio-visual speech perception have already shown (e.g., Breeuwer and Plomp, 1987; Grant *et al.*, 1985, 1991). A similar conclusion was reached for the auditory perception of sine-wave speech by Remez and Rubin (1990). They suggested that amplitude envelope variations were significant as a cue to syllable boundaries, especially around plosive consonants. Owing to the very low performance levels expected from sentence perception in purely auditory conditions with the signals used here, it is not readily possible to make comparable measures of any contribution of amplitude variation in these conditions

The relative effects of temporal cues to aperiodic excitation,  $F_0$ , and amplitude variation to consonant identification in experiments 1, 2, and 2A are shown together in Fig. 6. This displays the within-subject differences in performance attributable to the presence or absence of these cues from auditory and audio-visual conditions that differ only in respect of the added information. The median difference in auditory performance produced by the addition of amplitude variation is very close to zero in each comparison. The audio-visual performance increment with the addition of amplitude variation is, however, always greater than zero, but rarely exceeds 5% in any individual subject.  $F_0$  variation led to small performance increments more generally in both auditory and audio-visual presentation, typically of around 3%. Somewhat larger effects ranging up to 15% increases in accuracy were found for the voiceless cue.

That amplitude variation does not, in the absence of

spectral cues, lead to substantial increments in consonant identification may explain the lack of an effect of amplitude compression in consonant identification. For example, Souza and Turner (1996) have shown that the auditory identification of signal-correlated noise stimuli based on consonants is unaffected by compression of the amplitude envelope.

A temporal coding of the timing of voiceless excitation was shown to contribute to consonant identification in both auditory and audio-visual conditions, but not to the audio-visual identification of sentences. This is consistent with Articulation Index importance functions that place greater weight on higher frequency bands for nonsense word materials when phonemes occur with equal frequency than for meaningful connected speech (Studebaker *et al.*, 1987).

## B. Temporal cues to consonant features and sentence perception

In terms of perceptual features for consonant identity, we find evidence that voicing information is conveyed by the timing of both periodic and aperiodic excitation. Manner information is also provided by the timing of periodic and aperiodic excitation and here there is in addition a modest but significant contribution from  $F_0$  variation, and from amplitude variation in audio-visual conditions. For binary manner features, the timing of voiceless excitation is a cue for plosive, fricative, and affricate manners, while  $F_0$  and amplitude variations contribute to plosive and nasal information.

The transfer of place of articulation information solely by the auditory signals used here is slight in each case. However, experiment 2 did show a significant increase in place identification due to the amplitude variation of voiceless speech. There is thus no evidence that place-related variations in voice-onset time (Lisker and Abramson, 1967) that would be encoded in the timing of voiced and voiceless excitation lead to salient perceptual cues to place of articulation in the present experiments. However, the amplitude (but not the timing) of voiceless speech does appear to have small effects on place perception.

In the classification of temporal speech features, it seems important to distinguish two components of low-rate *envelope* information, the simple presence of energy, and the perhaps less significant low-rate amplitude variation. Temporal *periodicity* information clearly contributes to consonant perception through contrasting periodic and aperiodic excitation, and also through subtle effects of  $F_0$  on manner perception, in addition to the well known supra-segmental role of  $F_0$ .

## C. Implications for prosthesis design

Since amplitude variation has only minor effects on the detailed perception of consonantal features, it may be expected that, at least for speech in quiet, considerable compression of amplitude variation could be introduced without substantial cost provided that the typically gross supra-segmental variations in amplitude remain audible. It would therefore be expected that speech receptive prostheses intended for those who largely rely on audio-visual perception

need not convey finer details of speech amplitude variation. It is perhaps not surprising in the light of these findings that several studies show amplitude compression to have rather slight effects on speech perception at both segmental and sentence levels (Drullman and Smoorenburg, 1997; Souza and Turner, 1996; van Tasell and Trine, 1996). In noisy environments, however, it is plausible that variations of amplitude may contribute to the ability to segregate speech from noise, and here compression may be more detrimental.

At least where auditory signals lack spectral structure, fundamental frequency information is well-established as a source of useful information in sentence-level audio-visual speech perception. The results of experiment 1 suggest that prostheses which ensure a salient percept of  $F_0$  may also aid in the identification of consonants. That the explicit temporal representation of voiced and voiceless excitation can contribute to consonant identification may also be of practical significance for the design of hearing prostheses, since it would be likely to contribute to the audio-visual perception of low-redundancy messages. Voiceless excited speech is of course distinct from voice excited speech not only by aperiodicity, but also typically by the presence of predominantly higher frequency energy. In listeners for whom higher frequencies are not audible, a temporal coding of periodicity and aperiodicity is likely to be the only possible means of preserving this excitation contrast. Where sufficient frequency range and useful spectral resolution is retained, this contrast may also be accessible from spectral structure, and the significance of temporal cues to aperiodicity when spectral cues are also available merits further investigation.

## ACKNOWLEDGMENTS

Supported by TIDE Projects Nos. TP133/206 (STRIDE) and TP1217 (OSCAR), Medical Research Council Grant No. G9020214, and a Wellcome Trust Vacation Scholarship. We are grateful to Kirsti Reeve, Kerensa Smith, and Athena Euthymiades for carrying out the data collection, and to Winifred Strange, James Hillenbrand, and two anonymous reviewers for constructive comments on the manuscript.

<sup>1</sup>The use of the Voicscope for the detection of voiced excitation involves a thresholding procedure on a signal that follows the r.f. impedance across the larynx. This signal gives an unambiguous indication of vocal fold closure, but a less exact indication of smaller quasi-periodic movements of the vocal folds that do not result in vocal fold contact. It is possible for weak laryngeal excitation to be missed. The threshold was set as low as possible without allowing noise from the recorded larynx electrode output to trigger the voicing detector.

<sup>2</sup>The "pulse" setting of the VBF8 filter was used. This setting selects a filter type with a maximally flat delay response over frequency. The  $-3$  dB and other cutoff frequencies quoted above are measured values using a nominal 50-Hz cutoff frequency.

<sup>3</sup>Given the small subject numbers, the conventional mixed-effect ANOVA model has been modified to maximize the power of these analyses (experiments 1 and 2). The conventional error term, which is the interaction of the tested factor with subject, was excluded from the ANOVA model when it was clearly nonsignificant ( $p > 0.2$ ). Hence the error term for a given test becomes a higher order interaction, which now includes the variability from the excluded interaction. Where each relevant higher order interaction was also clearly nonsignificant, the overall MS error was used. This pooled error has the effect of increasing the error term  $df$  and the power of each  $F$  test. Having so modified the ANOVA model, the typical power for a dif-

ference in percent correct scores of around 3%, as for example, the effects in experiment 1 of  $F_0$  variation in auditory and audio-visual conditions, and that of amplitude variation in audio-visual conditions, was between 0.85 and 0.95.

<sup>4</sup>For this experiment, it proved impossible to increase the power of these ANOVAs by discarding nonsignificant interactions with the subject factor, since these interaction terms were generally substantially smaller than the residual error.

Baer, T., and Moore, B. C. J. (1993). "Effects of spectral smearing on the intelligibility of sentences in noise," *J. Acoust. Soc. Am.* **94**, 1229–1241.

Baker, R., and Rosen, S. (1994). "Temporal information in consonant identification," *Speech Hearing and Language; Work in Progress*, Dept. Phonetics and Linguistics, University College London, **7**, 3–29.

Bamford, J., and Wilson, I. (1979). "Methodological considerations and practical aspects of the BKB sentence lists," in *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children*, edited by J. Bench and J. Bamford (Academic, London), pp. 147–187.

Breeuwer, M., and Plomp, R. (1986). "Speech reading supplemented with auditorily presented speech parameters," *J. Acoust. Soc. Am.* **79**, 481–499.

Drullman, R., and Smoorenburg, G. F. (1997). "Auditory-visual perception of compressed speech by profoundly hearing impaired subjects," *Audiology* **36**, 165–177.

Fant, G. (1973). *Speech Sounds and Features* (MIT Press, Cambridge, MA).

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.

Faulkner, A., Ball, V., Rosen, S., Moore, B. C. J., and Fourcin, A. J. (1992). "Speech pattern hearing aids for the profoundly hearing-impaired: Speech perception and auditory abilities," *J. Acoust. Soc. Am.* **91**, 2136–2155.

Foster, J. R., Summerfield, A. Q., Marshall, D. H., Palmer, L., Ball, V., and Rosen, S. (1993). "Lip-reading the BKB sentence lists; corrections for list and practice effects," *Br. J. Audiol.* **27**, 233–246.

Fourcin, A. J. (1977). "English speech patterns with special reference to artificial auditory stimulation," in *A Review of Artificial Auditory Stimulation: Medical Research Council Working Group Report*, edited by A. R. Thornton (Institute of Sound and Vibration Research, University of Southampton), pp. 42–44.

Fourcin, A. (1990). "Prospects for Speech Pattern Element Aids," *Acta Oto-Laryngol. Suppl.* **469**, 257–267.

Grant, K. W., Ardell, L. H., Kuhl, P. K., and Sparks, D. W. (1985). "The contribution of fundamental frequency, amplitude envelope and voicing duration cues to speechreading in normal-hearing subjects," *J. Acoust. Soc. Am.* **77**, 671–677.

Grant, K. W., Braida, L. D., and Renn, R. J. (1991). "Single-band amplitude envelope cues as an aid to speechreading," *Q. J. Exp. Psychol.* **43A**, 647–678.

Lisker, L., and Abramson, A. S. (1970). "The voicing dimension: some experiments in comparative phonetics," in *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague, 1967 (Academia, Prague), pp. 563–567.

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. (1996). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," *Ear Hear.* **17**, 133–161.

Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (1992). "Relations among different measures of speech reception in subjects using a cochlear implant," *J. Acoust. Soc. Am.* **92**, 1869–1881.

Remez, R. E., and Rubin, P. E. (1990). "On the perception of speech from time-varying acoustic information: Contributions of amplitude variation," *Percept. Psychophys.* **48**, 313–325.

Risberg, A., and Lubker, J. L. (1978). "Prosody and speechreading," Report of STL-QPSR, Dept. of Linguistics, University of Stockholm, Stockholm, Sweden, **4**, 1–16.

Rosen, S., Fourcin, A. J., and Moore, B. C. J. (1981). "Voice pitch as an aid to lipreading," *Nature (London)* **291**, 150–152.

Rosen, S., Faulkner, A., and Smith, D. A. J. (1990). "The psychoacoustics of profound hearing impairment," *Acta Oto-Laryngol. Suppl.* **469**, 16–22.

Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory and linguistic aspects," *Philos. Trans. R. Soc. London, Ser. B* 367–373.

Shannon, R. V. (1993). "Psychophysics," in *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular Publishing, San Diego), pp. 357–388.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.

Shinn, P., and Blumstein, S. E. (1984). "On the role of amplitude envelope for the perception of [b] and [w]," *J. Acoust. Soc. Am.* **75**, 1243–1252.

Souza, P. E., and Turner, C. W. (1996). "Effect of single-channel compression on temporal speech information," *J. Speech Hear. Res.* **39**, 901–911.

Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (1987). "A frequency importance function for continuous discourse," *J. Acoust. Soc. Am.* **81**, 1130–1138.

Summerfield, A. Q. (1987). "Some preliminaries to a comprehensive account of audio-visual speech perception," in *Hearing by Eye: The Psychology of Lipreading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum Associates, Hove, U.K.), pp. 3–51.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. (1987). "Speech waveform envelope cues for consonant recognition," *J. Acoust. Soc. Am.* **82**, 1152–1181.

Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (1992). "Temporal cues for consonant recognition: Training, talker generalization, and use in the evaluation of cochlear implants," *J. Acoust. Soc. Am.* **92**, 1247–1257.

Van Tasell, D. J., and Trine, T. D. (1996). "Effects of single-band syllabic amplitude compression on temporal speech information in nonsense syllables and in sentences," *J. Speech Hear. Res.* **39**, 912–922.

Waldstein, R. S., and Boothroyd, A. (1994). "Speechreading enhancement using a sinusoidal substitute for voice fundamental frequency," *Speech Commun.* **14**, 303–312.



# Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors

Ann R. Bradlow

*Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208*

David B. Pisoni

*Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405*

(Received 3 August 1998; revised 26 March 1999; accepted 17 June 1999)

In order to gain insight into the interplay between the talker-, listener-, and item-related factors that influence speech perception, a large multi-talker database of digitally recorded spoken words was developed, and was then submitted to intelligibility tests with multiple listeners. Ten talkers produced two lists of words at three speaking rates. One list contained lexically “easy” words (words with few phonetically similar sounding “neighbors” with which they could be confused), and the other list contained lexically “hard” words (words with many phonetically similar sounding “neighbors”). An analysis of the intelligibility data obtained with native speakers of English (experiment 1) showed a strong effect of lexical similarity. Easy words had higher intelligibility scores than hard words. A strong effect of speaking rate was also found whereby slow and medium rate words had higher intelligibility scores than fast rate words. Finally, a relationship was also observed between the various stimulus factors whereby the perceptual difficulties imposed by one factor, such as a hard word spoken at a fast rate, could be overcome by the advantage gained through the listener’s experience and familiarity with the speech of a particular talker. In experiment 2, the investigation was extended to another listener population, namely, non-native listeners. Results showed that the ability to take advantage of surface phonetic information, such as a consistent talker across items, is a perceptual skill that transfers easily from first to second language perception. However, non-native listeners had particular difficulty with lexically hard words even when familiarity with the items was controlled, suggesting that non-native word recognition may be compromised when fine phonetic discrimination at the segmental level is required. Taken together, the results of this study provide insight into the signal-dependent and signal-independent factors that influence spoken language processing in native and non-native listeners. © 1999 Acoustical Society of America. [S0001-4966(99)02410-8]

PACS numbers: 43.71.Es, 43.71.Hw [JMH]

## INTRODUCTION

Speech perception and spoken word recognition accuracy depend on a wide range of talker-, listener-, and utterance-related characteristics, all of which can vary across communicative situations. A large and continuously growing body of work has provided us with important new information regarding the way in which talkers modify their speech production and articulation depending on a variety of linguistic and paralinguistic factors. For example, Lindblom (1990) showed how speakers vary their output along a continuum of hyper- and hypo-speech, using hyper-speech to assist a listener under “difficult” listening conditions, and hypo-speech when the talker believes less articulatory precision can be tolerated by the listener. A similar idea has been investigated over the past decade or so in a series of studies that examined the acoustic-phonetic factors that differentiate a “conversational” style of speech from a “clear” style of speech, such as one might use when addressing a person with a hearing loss (Picheny *et al.*, 1985, 1986, 1989; Uchanski *et al.*, 1996). Similarly, under the “Lombard effect,” talkers increase their vocal effort when talking in a noisy environment (Hanley and Steer, 1949; Draegert, 1951; Lane and Tranel, 1971), and adults adopt a hyper-articulated style of

speech when addressing infants (Fernald and Simon, 1984; Fernald *et al.*, 1989; Grieser and Kuhl, 1988; Kuhl *et al.*, 1997). These studies, and many others, have provided a great deal of new information about the way in which individual talkers modify and adjust their articulatory patterns to accommodate situational demands. However, aside from establishing that the “clear” speech style does indeed provide an intelligibility advantage over “conversational” speech (Picheny *et al.*, 1985), considerably less attention has been paid to the direct perceptual consequences, from the listener’s point of view, of different styles of speech (see Summers *et al.*, 1988; Lively *et al.*, 1993). Important questions that remain to be answered are: (1) Which of the clear speech transformations are most effective in aiding speech communication? And (2), how do listeners tune their performance according to communicative and situational demands? In order to develop a more complete understanding of the interplay between the talker-, listener-, and item-related factors that influence speech production and perception, we need to look at how the speech signal varies across a range of conditions, as well as how these variables affect listener performance.

With this overall goal in mind, recent work in our laboratory has focused on some of the factors that contribute to

variability in speech perception at the word and sentence levels. Our general approach stems from a basic view of speech communication as a highly adaptive process on the parts of both the talker and the listener. In carrying out our research, we believe that the use of large multi-talker multi-listener speech databases is essential for gaining a deeper understanding of the stimulus variability that is inherent in real-world speech production and perception.

To date, several factors have been shown to directly influence overall speech intelligibility by native listeners of American English. First, the degree of variability in the stimulus materials has been shown to have a major impact on the listener's speech recognition accuracy. For example, word recognition accuracies decrease and response times increase when listeners are presented with spoken word lists that incorporate a high-degree of stimulus variability due to the presence of multiple talkers and multiple speaking rates, relative to spoken word lists in which such stimulus variability is minimized (Mullennix *et al.*, 1989; Sommers *et al.*, 1994). Second, familiarity on the part of the listener with the talker's voice and articulatory characteristics enhances word recognition accuracy under difficult listening conditions. For example, Nygaard *et al.* (1994) showed recently that listeners were more accurate at identifying novel words in noise when the words were spoken by a talker who they had been trained to identify than when the same words were spoken by a novel talker (see also Nygaard and Pisoni, 1998). Third, the lexical characteristics of the particular words in a stimulus set exert a strong influence on overall intelligibility. Several recent studies have shown that lexically "easy" words (i.e., words with few phonetically similar "neighbors" with which they could be confused) are recognized better than lexically "hard" words (i.e., highly confusable words with many phonetically similar neighbors) (Pisoni *et al.*, 1985; Luce, 1986; Luce *et al.*, 1990; Luce and Pisoni, 1998). Finally, in a first attempt at identifying the talker-specific acoustic-phonetic characteristics that correlate with intertalker intelligibility differences, Bradlow *et al.* (1996) showed that talkers who exhibited a high-degree of "articulatory precision" in their speech generally had higher overall speech intelligibility scores than talkers who tended to produce more "reduced" speech (see also Wright, 1997). Taken together, these recent studies demonstrate that a range of talker-, listener-, and item-related factors affect the observed variability in overall speech intelligibility.

The present study extends this line of research by investigating the *combined* effects of various talker-, listener-, and item-related characteristics on isolated word recognition. The rationale of this study was that, in order to develop a comprehensive understanding of variability in speech production and perception, we need to directly investigate the ways in which multiple sources of variability operate in combination. Specifically, we hypothesized that perceptual difficulties introduced by one factor might be attenuated or amplified by the presence of another factor. For example, we expected that a relatively high degree of phonetic reduction introduced by a fast speaking rate might be tolerated when a listener becomes familiar with the speech of a particular talker. Conversely, we expected that hard word recognition would be

especially difficult for non-native listeners when there is a mismatch between the native and target language phoneme inventories. In order to test these predictions, we conducted two experiments, each of which examined spoken word recognition under conditions that manipulated talker-, listener-, and item-related factors both separately and in combination.

In experiment 1, we used a large database of digital speech recordings to assess the effects of speaking rate, lexical discrimination, and listener-talker adaptation on isolated word intelligibility. By directly examining the separate and combined effects of these characteristics on native-language speech intelligibility, we hoped to gain insight into perceptual processes that underlie native language word recognition. Specifically, we wanted to investigate the separate and combined effects of "signal-dependent" factors, such as speaking rate, and "signal-independent" factors, such as knowledge of the sound-based structure of the lexicon (Lindblom, 1990). Furthermore, the availability of this carefully constructed, multi-talker, multi-listener database provided us with a set of digital speech recordings along with normed intelligibility scores that could then be used in experiments that directly investigate spoken word recognition in a variety of special populations like non-native listeners or listeners with hearing impairments. Accordingly, in experiment 2 we used the same materials as in experiment 1 to investigate stimulus variability and spoken word recognition by non-native listeners. We wanted to see how non-native listeners cope with stimulus variability, and which demographic and linguistic variables correlate with non-native speech intelligibility.

The overall goal of these experiments was to describe in detail, and ultimately to provide a principled account of the relations between the various talker-, listener-, and item-related factors that influence spoken word recognition by both native and non-native listeners. While this was primarily an exploratory study, we believe that this type of fundamental knowledge about the way in which listeners compensate for multiple sources of variability in speech provides insight into the perceptual mechanisms that underlie spoken language processing.

## I. EXPERIMENT 1

### A. Method

#### 1. The "easy" and "hard" word lists

An "easy" list and a "hard" list of words (75 items each) were constructed such that the two lists differed in terms of three lexical characteristics (Pisoni *et al.*, 1985; Luce, 1986; Luce *et al.*, 1990; Luce and Pisoni, 1998). First, using the word frequency counts provided by the Brown Corpus of printed text (Kucera and Frances, 1967), the words were selected such that the mean word frequency of the easy list was significantly higher than the mean frequency of the words in the hard list (309.7 versus 12.2. per million). Second, using an on-line version of Webster's Pocket Dictionary (20 000 entries) in conjunction with a custom-designed lexical search program, words were selected such that the mean neighborhood density (the number of phonetic "neighbors") of the easy list was lower than the mean neighborhood den-

TABLE I. Descriptive statistics for the “easy” and “hard” word lists. Familiarity and frequency are characteristics of the target word itself. Density is the number of lexical neighbors, and mean neighborhood frequency is the mean frequency of all of these neighbors.

Easy words	Familiarity	Frequency	Density	Mean neighborhood frequency
Mean	6.97	309.69	13.53	38.32
Median	7	106	14	33.3
Standard deviation	0.08	1127.65	4.42	21.87
Minimum	6.5	36	1	2.33
Maximum	7	9816	20	79.67
Range	0.5	9780	19	77.33
Hard words				
Mean	6.81	12.21	26.61	282.23
Median	6.92	3	26	216.48
Standard deviation	0.23	45.85	4.91	215.96
Minimum	6.25	1	11	74.85
Maximum	7	365	39	1066.59
Range	0.75	364	28	991.75

sity of the hard list (13.5 versus 26.6). In these neighborhood density counts, a neighbor of a given target word was defined as any word that differed from the target word by a one phoneme addition, substitution, or deletion in any position (Greenberg and Jenkins, 1964). For example, some of the neighbors of the word “cat” are “pat, cot, cap, scat, at.” Third, the two word lists were constructed such that the mean neighborhood frequency (the mean frequency of the neighbors) of the easy list was much lower than the mean neighborhood frequency of the hard list (38.3 versus 282.2 per million). The net result of these three lexical manipulations was that the easy list consisted of a set of words that occur frequently in the language, and have few phonetically similar, low-frequency neighbors with which they could be confused. In contrast, the hard list consisted of words with many neighbors that are high in frequency relative to the target word. Thus, easy words “stand out” from sparse neighborhoods; hard words are “swamped” by dense neighborhoods. Finally, in order to ensure that subjects would be familiar with all of the words in both lists, all words had been judged as highly familiar by normal-hearing adults, i.e., received a familiarity rating of 6.25 or higher on a 7-point scale where 1 indicated the lowest and 7 indicated the highest degree of familiarity (Nusbaum *et al.*, 1984). Table I provides descriptive statistics for the various lexical characteristics of the words in the two word lists. The items in the two lists of words are provided in the appendix.<sup>1</sup>

## 2. Digital speech recordings

Ten talkers (five males and five females) were recorded producing both the easy and the hard word lists at three different speaking rates (fast, medium, and slow), giving a total of 4500 tokens (150 words  $\times$  3 speaking rates  $\times$  10 talkers). None of the talkers had any known speech or hearing impairments at the time of recording, and all were native speakers of General American English. The talkers were recruited from the Indiana University community and were paid for their participation. All talkers were told in advance that they would be asked to produce three word lists of 150 words each at three different speaking rates. Each individual

talker was allowed to regulate his/her own speaking rate, so long as the three rates were distinct. An analysis of the word durations for each talker at each of the three rates confirmed that each talker successfully produced the three lists with three distinct speaking rates. The mean durations were 809 ms (range 576–1030 ms), 525 ms (range 466–579 ms), and 328 ms (range 264–413 ms) for the slow, medium, and fast words, respectively, confirming that the talkers were successful at producing three distinct rates of speech.

All 150 words (75 easy plus 75 hard) were presented to the talkers in random order on a CRT monitor in a sound-attenuated booth (IAC model 401A). The stimuli were transduced with a Shure (SM98) microphone, and digitized online (16-bit analog-to-digital converter (DSC Model 240) at a 20-kHz sampling rate). The recordings were all live monitored by an experimenter for gross misarticulations and hesitations. Each individual digital file was then edited by hand to remove the silent portions at the beginning and end of each stimulus. The root-mean-square amplitude of each of the digital speech files was then equated. Finally, the files were converted to PC WAV format for presentation to listeners using a PC-based perceptual testing system (Hernandez, 1995).

## 3. Speech intelligibility tests

Speech intelligibility scores were collected from *independent* groups of ten normal-hearing listeners, each of whom transcribed the full set of 150 words from one talker at one speaking rate, for a total of 30 groups of 10 listeners (10 talkers  $\times$  3 speaking rates). The listeners were all recruited from the Indiana University community and were paid for their participation. None of the listeners reported any prior history of a hearing or speech impairment at the time of testing. The words were presented to the listeners in random order over matched and calibrated Beyer DT-100 headphones via a PC-based perceptual testing system (Hernandez, 1995). The words were presented in the clear (no background noise was added) at a comfortable listening level (70 dB SPL). On each trial, the listeners heard the word and then typed in their response on a computer keyboard. Each

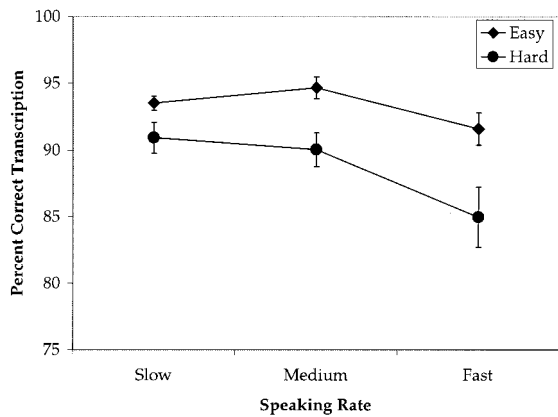


FIG. 1. Mean transcription accuracy scores across all talkers and listeners for the easy and hard words at the slow, medium, and fast speaking rates. The error bars represent the standard error of the mean.

listener received a different randomization of the 150 test words. In the data scoring, a word was counted as correct if all of the letters were present and in the correct order, if all the letters were present but not in the correct order (to allow for obvious typographical errors), or if the transcribed word was a homophone of the intended word.

These transcription scores provided a means of investigating the effects of speaking rate (fast versus medium versus slow) and lexical discrimination (easy versus hard) on isolated word intelligibility. Additionally, since each group of listeners transcribed the full set of 150 words by a single talker at a single rate in a single transcription session, we could also use these intelligibility data to investigate whether listeners adapted to talker-specific characteristics by comparing intelligibility scores from the beginning to scores from the end of the transcription session. We hypothesized that this kind of listener–talker “attunement” on the part of the listener, which occurs over the course of exposure to the speech of a particular talker, mediates the effects of lexical difficulty (easy versus hard) and speaking-rate (fast versus medium versus slow) such that some of the perceptual difficulty introduced by these stimulus factors could be overcome by listener–talker adaptation.

## B. Results

Figure 1 shows the overall percent correct transcription scores across all talkers and listeners for the easy and hard word lists at each of the three speaking rates. As expected based on earlier investigations of the effects of these lexical characteristics on speech perception (Pisoni *et al.*, 1985; Luce, 1986; Luce *et al.*, 1990; Luce and Pisoni, 1998), the easy word lists were consistently transcribed more accurately than the hard word lists. As shown in Table II, the higher transcription accuracy for the easy list relative to the hard list held true for most of the talkers at all three speaking rates. The exceptions were for talkers 2 and 7 at the slow rate and for talker 7 at the medium rate, where there was a very small advantage for the hard word list. Thus, the word identification advantage for easy words over hard words is a highly robust effect that generalizes across multiple talkers and speaking rates. The critical difference between easy and hard

TABLE II. Mean intelligibility scores across all ten listeners for the easy and hard word lists by each talker at each speaking rate.

Talker	Easy			Hard		
	Slow	Medium	Fast	Slow	Medium	Fast
1	91.07	92.40	86.13	82.67	81.20	72.27
2	94.40	95.47	94.27	94.80	94.40	89.33
3	94.67	94.00	94.93	88.93	89.60	92.53
4	92.40	96.00	88.27	88.67	87.20	78.00
5	94.00	94.40	86.27	89.47	91.33	75.47
6	92.93	93.87	91.87	92.80	90.40	89.73
7	90.67	89.20	89.47	91.07	90.26	87.87
8	94.93	96.27	92.93	93.60	88.40	89.47
9	95.07	96.67	95.73	92.40	92.13	84.40
10	95.07	98.40	96.27	94.93	95.46	90.67
Mean	93.52	94.67	91.61	90.93	90.04	84.97

words is that hard words require the listener to discriminate between a large set of competitors. In other words, in order to recognize a hard word correctly, the listener must make fine phonetic discriminations between words at the segmental level. The fact that this lexical competition effect is observed even under highly favorable listening conditions suggests that the ability to make fine phonetic discriminations is a skill that is prone to disruption, and as such is likely to be affected even more when conditions are less than favorable such as in the case of non-native listeners, noisy listening environments, or a hearing impairment.

Figure 1 also shows a substantial decline in transcription accuracy for the fast rate relative to the medium and slow rates for both the easy and the hard word lists; however, there was no intelligibility advantage for the slow rate over the medium rate. This pattern of results was somewhat surprising in view of the fact that, on average, the slow words were about 54% longer in duration than the medium words (see also Torretta, 1995). Thus, it appears that isolated word intelligibility is not enhanced by slowing the speaking rate. However, the absence of any difference may have been due to a ceiling effect for word intelligibility in quiet listening conditions.

These initial observations were all confirmed by a repeated-measures ANOVA (nested design) on the arcsine transformed data (Studebaker, 1985) with both speaking rate (fast, medium, slow) and lexical discrimination (easy, hard) as within subject variables, and the intelligibility scores for each talker in each condition averaged across all ten listeners as the dependent variable (see Table II). There was a main effect of speaking rate [ $F(2,18) = 11.127$ ,  $p < 0.001$ ], and a main effect of lexical discrimination [ $F(1,18) = 28.494$ ,  $p < 0.001$ ]. There was also a significant speaking rate by lexical discrimination interaction [ $F(2,18) = 5.862$ ,  $p = 0.011$ ], due to the increasing intelligibility difference between easy and hard words as the speaking rate increases. An examination of the paired contrasts showed a significant difference (at the  $p < 0.005$  level) between the fast and medium rates for both the easy and the hard words. There was no difference between the medium and slow rates for the hard words, whereas for the easy words there was a small but significant ( $p = 0.038$ ) advantage for the medium rate over the slow rate. Furthermore, at all three rates, the easy versus hard



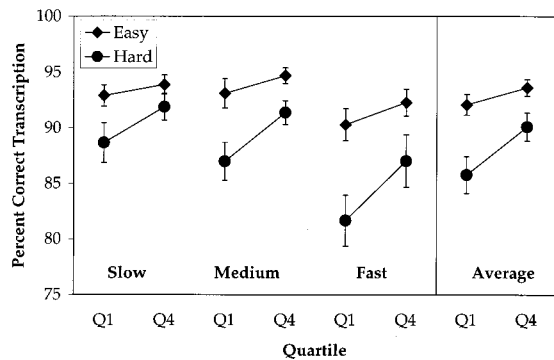


FIG. 2. Mean transcription accuracy scores across all talkers and listeners for the easy and hard words in the first and fourth quartiles at the slow, medium, and fast speaking rates, and averaged across all three speaking rates. The error bars represent the standard error of the mean.

difference was significant at the  $p < 0.005$  level.

The words in the easy and hard lists in this database were selected so that the effect of lexical difficulty could be assessed *across* the lists. In other words, the easy–hard difference across lists is largely categorical, rather than gradient. However, as shown in Table I, there is some degree of intralist variability in lexical difficulty.<sup>2</sup> Thus, we were able to perform correlational analyses on the various lexical characteristics and word intelligibility across the entire set of 150 words. Results showed a significant negative correlation between neighborhood density and intelligibility at all three speaking rates (slow:  $r = -0.213$ ,  $p < 0.01$ ; medium:  $r = -0.356$ ,  $p < 0.0001$ ; fast:  $r = -0.360$ ,  $p < 0.0001$ ). Furthermore, using a measure of target word “prominence,” which we defined as mean neighborhood frequency minus target word frequency, we found a trend towards a negative correlation between prominence and intelligibility at the medium and fast speaking rates (medium:  $r = -0.143$ ,  $p = 0.08$ ; fast:  $r = -0.155$ ,  $p = 0.06$ ). These results provide additional support for the fundamental assumptions of the neighborhood activation model of spoken word recognition, specifically, the assumption that spoken words are recognized relationally in the context of other phonetically similar words in the mental lexicon (Luce and Pisoni, 1998).

The final step in our analysis of these intelligibility data was to investigate whether isolated word intelligibility improves as the listener becomes accustomed to the talker’s voice. In particular, we wondered whether hard words that were presented later in a transcription session would be more accurately transcribed than hard words presented earlier in the session. We were interested in whether listener–talker adaptation might compensate for the processing difficulties introduced by the lexical discrimination factor.

Figure 2 shows the percent correct transcription scores for the easy and hard words in the first quartile (Q1) and fourth quartile (Q4) of the transcription sessions at each of the three speaking rates as well as across all three rates. In each case, the first and fourth quartiles were taken as the first and last 38 words presented to the listeners, respectively. Because each listener received a different randomization of the 150 words, differences due to particular items were controlled for over the entire group of listeners. As shown in Fig.

TABLE III. Mean intelligibility scores for each speaking rate in the first and fourth quartile.

	First quartile	Fourth quartile
slow	90.80	92.90
medium	90.05	93.04
fast	85.98	89.67

2, hard words presented in the last quartile were generally more accurately transcribed than hard words presented in the first quartile across all three speaking rates. In contrast, there was no noticeable difference between easy words presented in the first and fourth quartiles at all three speaking rates, a finding that may be due to a “ceiling” effect for easy words.

An ANOVA on the arcsine transformed data (Studebaker, 1985) for the intelligibility scores averaged across all three rates showed the expected main effect of lexical category [ $F(1,9) = 27.826$ ,  $p < 0.005$ ]. There was also a main effect of quartile [ $F(1,9) = 22.648$ ,  $p = 0.001$ ], indicating that the Q4 intelligibility scores were significantly higher than the Q1 intelligibility scores. Furthermore, there was a significant quartile by lexical category interaction [ $F(1,9) = 8.344$ ,  $p = 0.018$ ], due to the greater Q4–Q1 difference for the hard words than for the easy words. Interestingly, a pairwise comparison showed a nonsignificant difference between the easy words in the first quartile and the hard words in the fourth quartile. Separate ANOVA’s on the arcsine transformed data for each speaking rate showed that for all three rates there was a main effect of quartile, such that the Q4 intelligibility scores were consistently higher than the Q1 intelligibility scores [slow:  $F(1,9) = 9.298$ ,  $p = 0.014$ ; medium:  $F(1,9) = 12.166$ ,  $p < 0.007$ ; fast:  $F(1,9) = 19.322$ ,  $p < 0.002$ ]. There was also a main effect of lexical discrimination, such that easy words had higher intelligibility scores than hard words [slow:  $F(1,9) = 7.301$ ,  $p = 0.024$ ; medium:  $F(1,9) = 19.937$ ,  $p < 0.002$ ; fast:  $F(1,9) = 22.538$ ,  $p < 0.001$ ]. Furthermore, there was a tendency towards a quartile by lexical category interaction for the medium and fast rates [slow:  $F(1,9) = 1.270$ ,  $p = 0.289$ ; medium:  $F(1,9) = 5.074$ ,  $p = 0.051$ ; fast:  $F(1,9) = 3.857$ ,  $p = 0.081$ ].

These data on the time-course of word recognition indicate that as the listener becomes accustomed to the talker’s voice and specific articulatory patterns, the intelligibility difficulty introduced by the lexical characteristics of hard words can be overcome to a large extent. Furthermore, a comparison of the first and fourth quartile intelligibility scores across the three speaking rates (see Table III) showed that the intelligibility of fast rate words in the fourth quartile (mean = 89.67%) approached the intelligibility scores for the slow and medium rate words in the first quartile (means = 90.80% and 90.05%, respectively). In other words, the listener’s experience with the talker’s speech tended to compensate for the intelligibility difficulty introduced by the fast speaking rate. In general, this pattern of results suggests that listener–talker adaptation and attunement are important factors in speech perception that combine with other talker- and item-related factors, such as speaking rate and lexical

discrimination, in determining the overall intelligibility of normal speech by normal listeners.

### C. Summary and discussion

The primary goal of this initial experiment was to examine the combined effects of various talker-, item-, and listener-related factors on spoken word recognition by native listeners by using a carefully constructed multi-talker, multi-listener speech database. Results showed that overall word intelligibility was adversely affected by lexical discriminability: easy words had higher overall intelligibility than hard words. This effect of lexical discrimination was a listener-related factor that results from knowledge on the part of the listener regarding the sound-based structure of the lexicon of the language. We also observed a decline in overall intelligibility for the fast speaking rate: slow and medium rate words both had higher overall intelligibility scores than fast rate words. This speaking rate effect was a signal-related factor that presumably results from acoustic-phonetic adjustments on the part of the talker when he or she is required to consciously adjust speaking rate. We also observed a relationship between the various factors whereby the difficulties imposed by one factor, such as a fast speaking rate or an inherently difficult lexical item, could be overcome by the advantage gained through the listener's experience with the speech of a particular talker. Taken together, these data demonstrate that speech intelligibility is subject to a multitude of highly dynamic variables that have their basis in specific talker-, item-, and listener-related factors. These findings underscore the view of speech communication as an adaptive process from both the talker's and the listener's points of view. In the next experiment, we extended our investigation of factors affecting recognition of spoken words to another listener population, non-native listeners of English.

## II. EXPERIMENT 2

Spoken word recognition by non-native speakers depends on a wide range of skills including novel contrast categorization, the adoption of non-native processing strategies, and vocabulary development in the target language. Current research on non-native speech perception has been dominated by the study of the first of these skills, namely, non-native phoneme perception [e.g., see Strange (1995) and references therein]. The bulk of this research has focused on understanding the effects of the first language phoneme inventory on the ability to discriminate and identify second language phonemic contrasts. The findings have led to the development of several models that account for the different degrees of difficulty associated with the perception and production of different non-native contrasts (Best, 1995; Flege, 1995), and has provided researchers with important information about the effects of linguistic background on speech sound perception and categorization. However, we still do not know to what extent the perception of larger linguistic units by non-native listeners depends on fine-grained phoneme discrimination and identification. Is accurate phoneme categorization a necessary prerequisite for accurate word recognition by non-native listeners? Or, does novel phonemic

contrast perception arise from the ability to recognize word-sized units that contrast minimally with each other in the target language?

A similar issue is central to the study of first language acquisition in children. Current research in infant speech perception and early word learning has suggested that the system of meaningful contrasts develops only after infants have developed the skills to perceive and extract words-sized units from the speech stream. As Jusczyk (1997) notes,

“...it is unlikely that filling in a phonetic inventory is the primary force that drives infants' acquisition of the sound structure of their native language. Rather, the acquisition of phonemic categories and phonemic distinctions falls out of learning to segment and recognize words in the fluent speech of one's native language” (p. 109).

While adult second-language acquisition differs in many respects from infant first-language acquisition, it is likely that the need to recognize words is the primary force behind both processes. According to this point of view, sensitivity to non-native phonemic contrasts develops in response to the addition of new lexical items that reflect the specific contrast in question. While the adult second-language learner has the advantage of mature analytic skills that can aid the perception of phonological features at the segmental level, it is likely that novel phoneme perception can function in a linguistically meaningful manner only once the contrast in question signals a known lexical contrast. In other words, acquiring knowledge of the sound-based structure of the target language lexicon is just as important in non-native speech perception as gaining experience with the structure of the target language phoneme inventory. In order to fully understand non-native speech perception, we need to investigate recognition of word-sized units by non-native listeners using stimulus materials that are well controlled in terms of the sound-based structure of the target language. Accordingly, in experiment 2 we used the stimuli from the multi-talker database developed in experiment 1 to investigate spoken word recognition by non-native listeners.

In particular, we wanted to determine whether non-native listeners of English show the same effect of lexical discriminability as native listeners. Specifically, do non-native listeners have greater difficulty with “hard” words than with “easy” words? This outcome would suggest that non-native listeners develop lexicons of their second language using the same sound-based organizational principles as native listeners. We also wanted to know how non-native listeners perform under conditions of high stimulus variability due to a change in talker across items in a spoken word list. Previous research has shown that native listener word recognition is more accurate when surface characteristics, such as talker-related characteristics, remain consistent across items in a list (Mullennix *et al.*, 1989; Sommers *et al.*, 1994). Furthermore, as we found in experiment 1 above, native listeners show evidence of adaptation and tuning to these talker-related characteristics especially under conditions where word recognition is more difficult (i.e., lexically “hard” words). Thus, as a step towards gaining further in-

sight into the factors affecting recognition of spoken words, we wanted to see how non-native listeners cope with talker variability across items in a list.

Furthermore, in this experiment we assessed both spoken word recognition and written word familiarity. This comparison across these two modalities in adult second-language learners allowed us to look at non-native aural proficiency and non-native lexical development independently of each other. This independent measurement of non-native spoken word recognition and lexical development was particularly important because these two abilities might be confounded in non-native listeners. We know that spoken words are recognized by native listeners in the context of other words and that words requiring fine phonetic discrimination are more difficult to recognize (Luce and Pisoni, 1998). Thus, we might expect that non-native listeners will have particular difficulty with hard words since we know that fine phonetic discrimination of foreign language phonemes is particularly difficult for non-native listeners. However, lexically hard words are defined as words of lower frequency in the language, thus we might expect non-native listeners to be less familiar with hard words than easy words and therefore less likely to recognize them correctly. Thus, in order to understand the interaction of phonetic and lexical effects on non-native word recognition independently of word familiarity, we need to obtain independent measures of spoken word recognition and knowledge of the lexicon of the target language. Accordingly, we obtained both measures in experiment 2.

## A. Method

### 1. Subjects

Two groups of subjects participated in this experiment. The first group, the experimental group, included 20 non-native listeners of English who were recruited from the Indiana University community. They ranged in age from 21 to 33 years, and had studied English for 2 to 18 years. The group included 8 males and 12 females. They came from diverse language backgrounds, with the breakdown as follows: 6 Koreans, 4 Chinese, 3 Russians, 2 Japanese, 2 Spanish, 1 Bengali, 1 Nepali, and 1 Dani. The second group, the control group, included 20 native English listeners. They were also recruited from the Indiana University community, and ranged in age from 20 to 42 years. This group included 6 males and 14 females. All subjects were paid for their participation. None reported any known speech or hearing impairment at the time of testing.

### 2. Stimuli and procedures

All subjects performed two separate tasks. The first task was a spoken word recognition task in which subjects heard a word over headphones and typed what they heard into a computer keyboard. The stimuli for this task came from the multi-talker database of words that was described in experiment 1 above. Only words from the medium rate set were used in this experiment. Two separate lists of words were compiled. The first list consisted of 78 items produced by a single female talker whose mean intelligibility score for the medium rate words was closest to the average intelligibility

score across all ten talkers. Within this “single-talker” list, half of the words ( $n=39$ ) came from the easy list and half of the words ( $n=39$ ) came from the hard list. The second list consisted of 72 items, half of which were easy ( $n=36$ ) and half of which were hard ( $n=36$ ). The items in this “multiple talker” list were produced by the nine remaining talkers, four females and five males, with each talker producing four of the easy words and four of the hard words. There was no overlap between the items in the two lists. The single- and multiple-talker lists were presented to the listeners binaurally over matched and calibrated (DT-100 Beyer) headphones at a comfortable level (70 dB SPL). The order of list presentation (single-talker versus multiple-talker) was counterbalanced across listeners. Within each list, the words were presented in random order and the listeners were instructed to type the word they heard into the keyboard. Each word was presented only once with no possibility of repetition. However, the experiment was self-paced, allowing the listeners to correct spelling errors or make best guesses when entering their responses on the computer keyboard.

The second task was a word familiarity rating task in which subjects rated their familiarity with a list of English words. In this task, subjects responded to 300 words that were presented in standard American English orthography on a computer monitor. Subjects entered their response by pushing a button on a custom-made 7-button box after the word appeared on the screen. Subjects were instructed to use a 7-point scale where 1 indicated “I have never seen this word,” 4 indicated “I have seen this word but don’t know its meaning,” and 7 indicated “I know this word.” Of the 300 words used in this task, 150 came from the “easy” and “hard” lists used in experiment 1 and in the spoken word recognition task of experiment 2. The remaining 150 words were a subset of words that were taken from a longer list of words that had been used in a previous familiarity rating task with native listeners (Lewellen *et al.*, 1993). Of these, 50 received low ratings, 50 received medium ratings, and 50 received high ratings from the native listeners in this earlier study.

Taken together, the list of 300 words used in the present experiment included all of the words used in the spoken word recognition task plus a set of words known to cover a wide range of familiarity ratings from native listeners. Thus, this list provided us with a measure of the receptive vocabulary size of our non-native subjects relative to native subjects. Furthermore, these familiarity rating data allowed us to assess the extent to which non-native spoken word recognition depends on familiarity with the target word. All subjects performed the familiarity rating task after having completed the spoken word recognition task.

## B. Results

### 1. Spoken word recognition

Figure 3 shows the overall percent correct transcription scores for the easy and hard words for the control subjects (left panel) and for the non-native subjects (middle panel) in the single-talker and multiple-talker conditions, respectively. As expected, the control subjects displayed higher overall



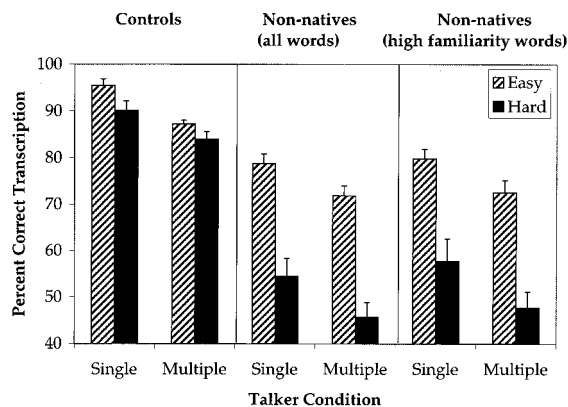


FIG. 3. Mean transcription accuracy scores for the easy and hard words in the single and multiple talker conditions for the control subjects (left panel), the non-native subjects (middle panel), and only the items of high familiarity to the non-native subjects (right panel). The error bars represent the standard error of the mean.

word recognition scores than the non-native listeners. The overall mean and standard deviation for the control subjects were 89.22% and 6.83%, respectively. For the non-native subjects, the mean and standard deviation were 62.73% and 12.24%, respectively. However, both subject groups showed similar patterns of results across the single- and multiple-talker conditions, as well as across the easy and hard words. For both groups, the overall percent correct recognition rate in the multiple-talker condition was lower than in the single-talker condition, indicating that both groups were able to take advantage of the consistent talker information in the single-talker condition. The difference between word recognition accuracy scores in the single- and multiple-talker conditions was 7.2% for the control subjects and 7.9% for the non-native subjects. Additionally, both groups showed higher recognition accuracy scores for the easy than for the hard words. However, there was a strong interaction between subject group and lexical category. Whereas the control subjects showed a difference of 4.3% between easy and hard words, the non-native subjects showed a much larger difference of 25.2%, and this difference was present for both the single- and multiple-talker conditions. The pattern of results displayed in Fig. 3 was confirmed by a three-factor ANOVA on the arcsine transformed data (Studebaker, 1985) with group (non-native, control), talker (single, multiple) and lexical category (easy, hard) as factors. This analysis showed main effects of group [ $F(1,38) = 113.234, p < 0.001$ ], talker [ $F(1,38) = 48.085, p < 0.001$ ], and lexical category [ $F(1,38) = 127.146, p < 0.001$ ]. There was also a significant lexical category  $\times$  group interaction [ $F(1,38) = 38.861, p < 0.001$ ]. None of the other interactions was significant.

The significant difference in word recognition performance between the single- and multiple-talker conditions for both groups of subjects suggests that the ability to take advantage of consistent surface information about a particular talker's voice is a skill that transfers easily from first to second language. Conversely, this result suggests that the processing difficulty introduced by a high degree of variability in the stimulus set due to a change in talker from item to item

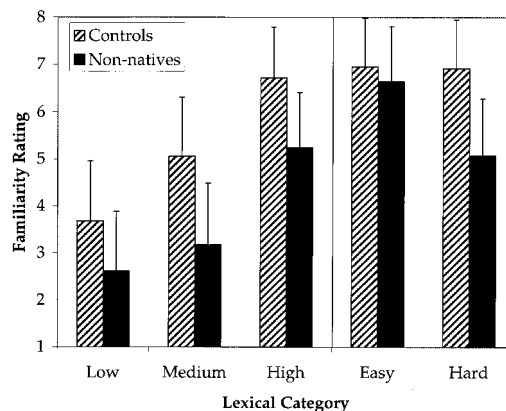


FIG. 4. Mean familiarity ratings for the control and non-native subjects on words of previously determined low, medium, and high familiarity (left panel), and the easy and hard words used in the present word recognition tests (right panel). The error bars represent the standard error of the mean.

is not particularly acute for non-native listeners. Rather, all listeners, regardless of language background, respond similarly to indexical, surface-level variability. The highly significant easy-hard word difference for the non-native listeners suggests that these listeners are developing an English language lexicon with the same sound-based structure as the native English listener lexicon. However, the fact that the non-native listeners showed much lower scores for hard words relative to the control subjects suggests that they have much greater difficulty when fine phonetic discrimination at the segmental level is required by the task. Nevertheless, these non-native subjects appear to be recognizing spoken words relationally in the context of other words they know, although at somewhat lower levels of accuracy relative to native speakers.

## 2. Familiarity ratings

Figure 4 shows the mean familiarity ratings given by the control and non-native subjects in response to the five word lists used in this task. The words in the low, medium, and high lists shown in the left panel were classified into these three categories based on earlier ratings from a large number of native listeners (Lewellen *et al.*, 1993). The easy and hard lists shown in the right panel contained the same easy and hard words that were presented to the subjects in the spoken word recognition task. For the low, medium, and high lists, the non-native listeners gave substantially lower familiarity ratings than the control subjects. However, both groups showed the expected pattern of increasing familiarity ratings from the low to the medium to the high word lists, suggesting that this task is indeed a valid measure of word familiarity in non-native listeners (see Lewellen *et al.*, 1993).

Of greater interest are the results of the familiarity rating task with the easy and hard words lists. These words were originally selected so that native listeners would be highly familiar with all the test words. This native listener familiarity is indicated in Fig. 4 by the high mean ratings for the control subjects (striped bars) for both the easy (mean rating = 6.9) and hard (mean rating = 6.9) words lists. In contrast, the non-native listeners (black bars) had a high mean



TABLE IV. Correlations between spoken word recognition accuracy, word familiarity ratings, and demographic variables. Numbers in parentheses refer to the range for each variable.

	Word recognition		Familiarity ratings	
	Easy (60%–89%)	Hard (25%–74%)	Easy (3.95–7.00)	Hard (3.69–7.00)
Age of English study onset (4–23 yrs)	+0.09	–0.22	+0.04	–0.61 <sup>b</sup>
No. of years of English study (<1–18)	+0.11	+0.28	+0.17	+0.37
No. of years in English environment (<1–8)	+0.37	+0.45 <sup>a</sup>	+0.04	+0.12

<sup>a</sup> $p < 0.05$ .

<sup>b</sup> $p < 0.005$ .

familiarity rating for the easy words (mean rating=6.6), but their ratings for the hard words were much lower (mean rating=5.1). Thus, the pattern of familiarity ratings parallels the pattern of word recognition scores for the non-native subjects, suggesting that part of their difficulty in recognizing hard words may stem from a lack of familiarity with the words themselves rather than from a difficulty with fine phonetic discrimination.

In order to assess the relationship between word familiarity and spoken word recognition performance in the non-native listeners, we reanalyzed the non-native spoken word recognition data by limiting our analysis to only those words that received a familiarity rating of 6 or higher. In this manner, both the non-native and the control subjects' scores reflect word recognition accuracy for words that are judged to be highly familiar to the listeners. The right panel of Fig. 3 shows the non-native subjects' mean word recognition accuracy scores in the single and multiple talker conditions only for the easy and hard words that received a familiarity rating of 6 or higher. On average, across all 20 non-native subjects, 105 of the original 150 words (70%) were included in this analysis. This includes an average of 54/75 (72%) of the easy words and 51/75 (68%) of the hard words.

As shown in Fig. 3, the general pattern of results for the non-natives that we observed for all words (middle panel) is present even when we remove the confounding factor of word familiarity (Fig. 3, right panel). A three-factor ANOVA [on the arcsine transformed data (Studebaker, 1985) for only the high familiarity non-native word recognition scores] with group (non-native, control), talker (single, multiple), and lexical category (easy, hard) as factors showed main effects of group [ $F(1,38) = 97.340$ ,  $p < 0.001$ ], talker [ $F(1,38) = 38.760$ ,  $p < 0.001$ ], and lexical category [ $F(1,38) = 72.944$ ,  $p < 0.001$ ]. There was also a significant lexical category  $\times$  group interaction [ $F(1,38) = 20.139$ ,  $p < 0.001$ ]. None of the other interactions were significant.

Thus, while non-native word recognition accuracy may be affected by familiarity with the lexical items, even when we controlled for familiarity, we observed a strong easy–hard lexical effect for these listeners. This pattern suggests that non-native listeners develop second-language mental lexicons that follow the same sound-based structure as the first-language mental lexicon, and that the fine phonetic discrimi-

nation required for accurate recognition of hard words is especially difficult for these listeners.

### 3. Correlational analyses

In order to further investigate the factors that underlie non-native listener responses to spoken words, we performed a series of correlational analyses between the mean spoken word recognition accuracy scores for each of the 20 non-native subjects and various demographic factors that we obtained from subjects at the start of the data collection sessions. We also performed a similar set of correlational analyses between these demographic variables and the mean familiarity rating score for each of the non-native subjects. In each case, we performed separate correlations for the easy word scores and the hard word scores. Table IV shows the results of these correlational analyses for the variables of greatest interest. For each variable, the numbers in parentheses represent the range of scores across all 20 subjects.

For all dependent variables, none of the correlations with the easy words were significant. This may be because the ranges of word recognition and familiarity rating scores for the easy words were more restricted than for the hard words. There was little or no variance in these measures for the easy words because of ceiling effects in performance. However, for the hard words several interesting correlations emerged. The data showed no correlation between age of onset of English study and hard word recognition; however, number of years in an English environment was significantly positively correlated with hard word recognition scores ( $r = +0.45$ ). In contrast, there was no correlation between hard word familiarity and number of years in an English environment; however, age of onset of English study was significantly negatively correlated with hard word familiarity ( $r = -0.61$ ). Number of years of formal English study was not significantly correlated with either hard word recognition or hard word familiarity. These correlations with the demographic variables suggest that spoken word recognition is an essentially aural skill that requires exposure to spoken language, whereas written vocabulary development is most aided by an early onset of formal second-language study.

### C. Summary and discussion

In experiment 2 we investigated some of the characteristics of non-native spoken word recognition as they relate to known characteristics of native spoken word recognition. We found that spoken word recognition by non-native listeners displayed the same overall patterns as for native listeners. Specifically, both groups of listeners recognized words more accurately when all the test words were spoken by the same talker relative to a condition where the talker changed from item to item. This finding suggests that the ability to take advantage of consistent surface phonetic information, such as consistencies in the talker's voice and articulatory patterns, is a language-independent skill that transfers easily from first-language to second-language word recognition.

We also found that both groups of listeners were more accurate at recognizing words that were distinctive or easily discriminated in their lexical neighborhood than those that had many similar sounding neighbors with which they can easily be confused. However, this effect was much more prominent for the non-native listeners, suggesting that these listeners have particular difficulty in recognizing words that require perception of fine phonetic detail for lexical discrimination. This pattern of results was observed even when we controlled for word familiarity across the easy and hard word lists.

Additionally, we found a dissociation between word recognition accuracy and word familiarity ratings with each representing a different skill. Hard word recognition correlated positively with number of years immersed in an English language environment but not with total number of years of English study or age of English study onset, suggesting that hard word recognition may be a good index of non-native aural proficiency independently of vocabulary development. In contrast, hard word familiarity was correlated negatively with age of onset of English study but not with number of years in an English language environment or with total number of years of English study, suggesting that hard word familiarity may be a good index of non-native lexicon development independently of non-native language aural proficiency.

### III. GENERAL DISCUSSION

Taken together, these two perceptual experiments demonstrate various characteristics of word recognition by native and non-native listeners. From a methodological point of view, our results show the utility of a large multi-talker multi-listener digital speech database for investigations into spoken language processing. An important aspect of the database that was developed in the present study was that it included a large number of stimulus items produced by a large number of talkers that were then submitted to intelligibility tests by a large number of listeners. This approach to speech database development—one that always includes both production and perception data—has proved particularly effective as a means of investigating the effects of variability in the speech signal from the points of view of both the talker and the listener. We believe that an important goal of research in spoken language processing is to understand

both the sources of variability in the speech signal, and the effects of this variability on the listener (Stevens, 1996). In order to achieve these goals, researchers will need to devise new ways of investigating the separate and combined effects of various sources of stimulus variability in speech. Our multi-talker multi-listener database approach has proved particularly useful in this regard.

From a theoretical standpoint, the findings of the present study point to several key features of spoken language processing. The data demonstrate that spoken word recognition accuracy depends on a combination of at least three types of factors: (1) signal-related characteristics, such as speaking rate, (2) lexical factors, such as knowledge of the sound-based structure of the mental lexicon, and (3) instance-specific factors, such as the listener's prior experience with the talker's voice and articulatory habits. All three factors combine to determine overall speech intelligibility.

Of particular theoretical interest in this study is the finding that listeners adapt to the demands of the communicative situation in much the same way as talkers do. Just as talkers adapt their speech patterns to match the demands of the communicative situation, so do listeners tune and adjust their speech perception mechanisms to take advantage of surface level or paralinguistic consistencies in the signal (see also Nygaard *et al.*, 1994; Kakehi, 1992; Nygaard and Pisoni, 1998 for similar findings). This finding raises the basic question of what listeners are learning over the course of exposure to the speech of a particular talker. Recently, Nygaard and Pisoni (1998) suggested two possible mechanisms that underlie this form of perceptual learning. One possibility is that the listener becomes more efficient at performing the operations that map the talker-specific phonetic implementations to their abstract phonemic representations. In other words, the listener becomes well practiced at the specific procedures required to normalize across the particular talker's idiosyncratic phonetic implementation characteristics, in order to arrive at the intended symbolic representation of the speech signal. This view assumes that the linguistic and indexical (i.e., talker-specific) information conveyed by the speech signal are orthogonal, and that the recovery of the linguistic content is aided by more efficient separation of the linguistic and indexical aspects of the speech signal. The other possibility considered by Nygaard and Pisoni (1998) is that the linguistic and indexical aspects of the signal are integral. According to this view, the talker-specific indexical information and the linguistic content of a signal are carried by the same kinds of time-varying acoustic characteristics. Thus, a high degree of sensitivity to the talker-specific indexical aspects implies an equally high degree of sensitivity to the linguistic aspects of the signal. Consequently, talker familiarity and enhanced word recognition performance necessarily go together. While the data from the present study do not support either one of these alternatives over the other, our results do demonstrate that this type of sensitivity to consistent surface characteristics across items in a list is a feature of spoken language processing that functions independently of whether the listener is perceiving his or her native language or a foreign language.

The present findings also demonstrate a strong effect of

fine-grained phonetic discrimination on word recognition. Word recognition accuracy was always compromised when fine phonetic discrimination was needed to recognize a word, as in the case of hard words spoken at a fast rate for native listeners (experiment 1) or hard words spoken at a medium rate for non-native listeners (experiment 2). A fast speaking rate results in reduced acoustic-phonetic cues in the signal. Similarly, non-native listeners have reduced sensitivity to crucial acoustic-phonetic cues due to their lack of experience with speech in the target language. Thus, when fine-grained acoustic-phonetic discrimination is reduced, whether due to signal-related factors or listener-related factors, word recognition accuracy is reduced accordingly.

This finding suggests that while listeners may be primarily motivated to recognize word-sized units (Jusczyk, 1997), their ability to access lexical items is limited by the degree of low-level acoustic-phonetic detail that is available from the signal. In other words, spoken language processing relies on both accurate phoneme categorization and knowledge of the sound structure of the target language (Luce and Pisoni, 1998). Any attempt to enhance speech intelligibility for non-native listeners or for native listeners under difficult listening conditions due to hearing loss or environmental noise should consider both the degree of acoustic-phonetic detail available in the signal and the phonological and lexical nature of the stimulus materials to be recognized. Depending on various factors, such as those explored in this study, more or less acoustic-phonetic reduction may be tolerated without significant loss of intelligibility.

Consistent with this view, we might predict that, since talkers are presumably also attuned to the sound-based structure of the mental lexicon, they will tend to hyperarticulate hard words. Wright (1997) tested this prediction by performing acoustic analyses of the materials in the same database that we used in the present study. He found that the vowels in the easy words were significantly more centralized (i.e., reduced) than the vowels in the hard words. Nevertheless, as demonstrated by the highly robust easy-hard effect observed in the present study, this hyperarticulation was not sufficient to overcome the effect of lexical difficulty on the part of the listener.

Thus, while both talkers and listeners apparently adapt and modify their performance to the demands of the communicative situation, the effectiveness of each adaptive strategy in enhancing intelligibility can only be judged in relation to other factors that are known to affect speech intelligibility. For this reason, it is critical that speech researchers investigate the separate and combined effects of a wide range of talker-, listener-, and item-related factors on spoken language processing.

## ACKNOWLEDGMENTS

We are grateful to Gina Torretta for data collection and processing, and to Luis Hernandez for technical support. We are also grateful to Chris Darwin, James Hillenbrand, John Kingston, and Terrance Nearey for many insightful and helpful comments. This research was supported by NIDCD Training Grant No. DC-00012 and by NIDCD Research Grant No. DC-00111 to Indiana University. Earlier versions

of this work were presented in the Fall of 1997 at the 134th meeting of the Acoustical Society of America in San Diego, CA (2–6 December 1997), and at the International Symposium on Speech Perception by Non-Native Listeners in Boston, MA (19–21 November 1997).

## APPENDIX: WORD LISTS

Easy words			Hard words		
was	live	dog	ban	rum	pawn
down	move	vote	bead	sane	bun
work	food	league	bean	soak	gut
long	size	thick	bug	suck	lice
both	cause	page	bum	tan	mid
thought	wrong	hung	chat	wed	wick
does	chief	join	cheer	white	hurl
put	faith	shop	comb	whore	moat
give	pool	roof	cot	wrong	teat
young	deep	leg	den	con	hash
thing	firm	lose	dune	doom	hid
peace	serve	theme	fade	hick	hoot
god	reach	soil	fin	rut	mace
five	mouth	pull	goat	toot	main
gave	teeth	chain	knob	wad	moan
death	gas	curve	lad	bud	mum
shall	jack	path	mall	dame	rim
real	check	dirt	mat	lace	rout
south	king	vice	mitt	lame	wail
job	shape	rough	mole	pad	hum
love	learn	fool	pat	chore	sill
full	ship	noise	pet	cod	beak
wife	neck	wash	pup	hack	hag
voice	watch	balm	rat	kin	wade
girl	judge	fig	rhyme	kit	weed

<sup>1</sup>The entire on-line version of Webster's Pocket Dictionary, which includes the lexical characteristics for all of the 20 000 entries in this dictionary, is available in spreadsheet format (Microsoft Excel) from the Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405.

<sup>2</sup>There was also a small amount of overlap between the two lists on all three of the lexical characteristics (namely, frequency, density, and mean neighborhood frequency). In fact, one word ("wrong") appeared in both word lists. While this was a somewhat regrettable situation, removing it from the analysis did not alter the overall intelligibility scores of either list at any speaking rate in a significant way (less than 0.11% difference in all cases).

Best, C. T. (1995). "A direct-realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 171–206.

Bradlow, A. R., Torretta, 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.

Dræger, G. L. (1951). "Relationships between voice variables and speech intelligibility in high level noise," *Speech Monogr.* **18**, 272–278.

Fernald, A., and Simon, T. (1984). "Expanded intonation contours in mothers' speech to newborns," *Dev. Psychobiol.* **20**, 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," *J. Child Lang.* **16**, 477–501.

Flege, J. E. (1995). "Second language speech learning: Theory, findings and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–272.



- Greenberg, J. H., and Jenkins, J. J. (1964). "Studies in the psychological correlates of the sound system of American English," *Word* **20**, 157–177.
- Grieser, D., and Kuhl, P. K. (1988). "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese," *Dev. Psychobiol.* **24**, 14–20.
- Hanley, T. D., and Steer, M. D. (1949). "Effect of level of distracting noise upon speaking rate, duration and intensity," *J. Speech Hear. Disord.* **14**, 363–368.
- Hernandez, L. R. (1995). "Current computer facilities in the Speech Research Laboratory," *Research on Spoken Language Processing, Progress Report 20*, Indiana University, Bloomington, IN, pp. 389–394.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language* (MIT, Cambridge, MA).
- Kakehi, K. (1992). "Adaptability to differences between talkers in Japanese monosyllabic perception," in *Speech Perception, Speech Production, and Linguistic Structure*, edited by Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (OHM, Tokyo), pp. 135–142.
- Kucera, F., and Francis, W. (1967). *Computational Analysis of Present Day American English* (Brown U. P., Providence, RI).
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Koshevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, E. I., and Lacerda, F. (1997). "Cross-language analysis of phonetic units in language addressed to infants," *Science* **277**, 684–686.
- Lane, H. L., and Tranel, B. (1971). "The Lombard sign and the role of hearing in speech," *J. Speech Hear. Res.* **14**, 677–709.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 403–439.
- Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., and Greene, B. G. (1993). "Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization," *J. Exp. Psychol.* **122**, 316–330.
- Lively, S. E., Pisoni, D. B., Summers, W. V., and Bernacki, R. H. (1993). "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences," *J. Acoust. Soc. Am.* **93**, 2962–2973.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**, 1–36.
- Luce, P. A. (1986). "Neighborhoods of words in the mental lexicon," *Research on Speech Perception, Technical Report No. 6*, Indiana University, Bloomington, IN.
- Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). "Similarity neighborhoods of spoken words," in *Cognitive Models of Speech Processing: Psycholinguistics and Computational Perspectives*, edited by G. Altmann (MIT, Cambridge, MA), pp. 122–147.
- Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). "Some effects of talker variability on spoken word recognition," *J. Acoust. Soc. Am.* **85**, 365–378.
- Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," *Research in Speech Perception, Progress Report 10*, Indiana University, Bloomington, IN, pp. 357–376.
- Nygaard, L. C., and Pisoni, D. B. (1998). "Talker-specific learning in speech perception," *Percept. Psychophys.* **60**, 335–376.
- Nygaard, L. C., Sommers, M. C., and Pisoni, D. B. (1994). "Speech perception as a talker-contingent process," *Psychol. Sci.* **5**, 42–46.
- Picheny, M. A., Durlach, N. I., and Braidia, 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.
- Picheny, M. A., Durlach, N. I., and Braidia, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Hear. Res.* **29**, 434–446.
- Picheny, M. A., Durlach, N. I., and Braidia, L. D. (1989). "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech," *J. Speech Hear. Res.* **32**, 600–603.
- Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and Slowiaczek, L. M. (1985). "Speech perception, word recognition and the structure of the lexicon," *Speech Commun.* **4**, 75–95.
- Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (1994). "Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude," *J. Acoust. Soc. Am.* **96**, 1314–1324.
- Stevens, K. N. (1996). "Understanding variability in speech: A requisite for advances in speech synthesis and recognition," *J. Acoust. Soc. Am.* **100**, 2634.
- Strange, W. (editor) (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York, Timonium, MD).
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). "Effects of noise on speech production: Acoustic and perceptual analyses," *J. Acoust. Soc. Am.* **84**, 917–928.
- Torretta, G. M. (1995). "The 'easy-hard' word multi-talker speech database: An initial report," *Research on Spoken Language Processing, Progress Report 20*, Indiana University, Bloomington, IN, pp. 321–334.
- Uchanski, R. M., Choi, S., Braidia, 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.
- Wright, R. (1997). "Lexical competition and reduction in speech: A preliminary report," *Research on Spoken Language Processing, Progress Report 21*, Indiana University, Bloomington, IN, pp. 471–486.



# Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children

Ann R. Bradlow

*Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208*

Nina Kraus, Trent G. Nicol, Therese J. McGee, Jenna Cunningham, and Steven G. Zecker

*Auditory Neuroscience Laboratory, Northwestern University, 2299 North Campus Drive, Evanston, Illinois 60208*

Thomas D. Carrell

*Department of Communication Sciences, University of Nebraska, Lincoln, Nebraska 68538*

(Received 2 November 1998; revised 17 May 1999; accepted 8 June 1999)

In order to investigate the precise acoustic features of stop consonants that pose perceptual difficulties for some children with learning problems, discrimination thresholds along two separate synthetic /da-ga/ continua were compared in a group of children with learning problems (LP) and a group of normal children. The continua differed only in the duration of the formant transitions. Results showed that simply lengthening the formant transition duration from 40 to 80 ms did not result in improved discrimination thresholds for the LP group relative to the normal group. Consistent with previous findings, an electrophysiologic response that is known to reflect the brain's representation of a change from one auditory stimulus to another—the mismatch negativity (MMN)—indicated diminished responses in the LP group relative to the normal group to /da/ versus /ga/ when the transition duration was 40 ms. In the lengthened transition duration condition the MMN responses from the LP group were more similar to those from the normal group, and were enhanced relative to the short transition duration condition. These data suggest that extending the duration of the critical portion of the acoustic stimulus can result in enhanced encoding at a preattentive neural level; however, this stimulus manipulation on its own is not a sufficient acoustic enhancement to facilitate increased perceptual discrimination of this place-of-articulation contrast. © 1999 Acoustical Society of America. [S0001-4966(99)00210-6]

PACS numbers: 43.71.Pc, 43.71.Ft [JMH]

## INTRODUCTION

Numerous studies have established that a subset of children with diagnosed language and/or learning problems exhibit deficits perceiving certain acoustically similar speech sounds (e.g., Tallal and Piercy, 1974, 1975; Brandt and Rosen, 1980; Godfrey *et al.*, 1981; Tallal and Stark, 1981; Werker and Tees, 1987; Reed, 1989; Elliott *et al.*, 1989; Leonard and McGregor, 1992; Sussman, 1993; Stark and Heinz, 1996a, b; Kraus *et al.*, 1996; Mody *et al.*, 1997, and others). However, the precise nature of the underlying perceptual deficit and its effect on language development and academic achievement has yet to be fully explained. As part of a comprehensive study of speech sound discrimination abilities, academic achievement, and neurophysiologic responses to speech stimuli in normal and learning-disabled school-aged children, we were interested in investigating the specific acoustic-phonetic features that provoke perceptual difficulty in the disabled population. While the sound structure of naturally spoken language is sufficiently rich in acoustic and contextual cues to tolerate some degree of imprecision in perceptual acuity, measures of fine-grained speech sound perception are potentially very useful for the identification and monitoring of children whose learning problems may have an auditory-phonetic basis. Accordingly,

the present paper presents a systematic examination of fine-grained discrimination of one aspect of stop-vowel syllables that has been proposed as a perceptually vulnerable feature, namely the characteristically brief formant transition duration in the initial portion of the vowel following the stop release.

An important aspect of this study is the use of a combined behavioral and neurophysiologic approach (see also Kraus *et al.*, 1993, 1996, in press). With this multidisciplinary technique, we hoped to gain insight into both stimulus representation at a preattentive neural level and perception of these speech signals in a task requiring focused attention to the stimuli. Furthermore, in both the neurophysiologic and behavioral testing paradigms, we used a task that tested the subject's ability to detect small differences between synthetic speech stimuli that varied along a single acoustic dimension. The use of this type of fine-grained discrimination task with "stripped down" synthetic speech stimuli allowed us to test speech sound perception under conditions that stress the system beyond what is required in natural spoken language processing. This testing procedure was deemed appropriate for investigating the auditory-perceptual deficits of school-aged children with learning problems based on the assumption that, rather than having difficulty perceiving naturally produced speech under favor-

able listening conditions, these subjects have difficulty under less-than-optimal listening conditions. For example, perceptual difficulties may become apparent in noisy listening environments, or when fine categorical distinctions at the segment level are important, such as when learning the sound-letter associations required for reading.

The literature on speech sound perception in impaired children is vast and somewhat diffuse in terms of subject, task, and stimulus factors. Nevertheless, several general findings have emerged, which we attempt to summarize here. With respect to subject population, various studies have examined children diagnosed as specific language impaired (SLI), reading impaired (dyslexics), or learning impaired. For example, several studies found speech perception deficits in children with “developmental aphasia,” or SLI (e.g., Tallal and Piercy, 1974, 1975; Frumkin and Rapin, 1980; Tallal and Stark, 1981; Stark and Heinz, 1996a, b; Elliott and Hammer, 1988; Elliott *et al.*, 1989; Sussman, 1993, and others). Similar deficits have also been documented in reading-disabled children (e.g., Tallal, 1980; Godfrey *et al.*, 1981; Brandt and Rosen, 1980; Werker and Tees, 1987; De Weirtd, 1988; Reed, 1989; Mody *et al.*, 1997, and many others). Finally, Kraus *et al.* (1996) recently reported deficits in the biological representation of speech sounds in children diagnosed with a learning disability, attention deficit disorder, or both. While these subject population groups have apparently disparate diagnostic labels, it is clear that individuals within these groups exhibit similar speech perception deficits, suggesting that these deficits are somehow correlated, or coexist with impaired language development, reading, and general academic achievement.

A variety of task-related variables have been employed in studies of speech perception abilities in impaired children; however, most tasks fall into one of two basic types. The first type of task requires subjects to respond to stimulus pairs, or longer strings of stimuli, in which each member represents a good exemplar of a particular speech sound category (e.g., Tallal and Piercy, 1974, 1975; Tallal, 1980; Frumkin and Rapin, 1980; Tallal and Stark, 1981; Stark and Heinz, 1996b; Mody *et al.*, 1997). In these tasks, subjects are typically required to identify a given stimulus as a member of one of two categories, to discriminate between the two stimuli, or to judge the order of presentation of the two stimuli. Since subjects hear only good category exemplars, these tasks tap into the subjects’ abilities to make judgments that rely on perception of cross-category differences. This type of task using stimulus pairs that are minimally different has revealed identification, discrimination, and temporal order judgment impairments in clinical populations relative to normal populations.

In the second type of task, subjects are presented with series of stimuli that represent points along an acoustic continuum. In the traditional categorical perception paradigm, both identification and discrimination functions along synthetic speech continua are measured in order to determine the “sharpness” of category boundaries. Using this paradigm, several studies found that, when compared to normal children, disabled children were less consistent in identifying, and less accurate at discriminating, stimuli close to category

boundaries, indicating less sharply defined phonetic categories than normal children (Brandt and Rosen, 1980; Godfrey *et al.*, 1981; Werker and Tees, 1987; De Weirtd, 1988; Reed, 1989; Sussman, 1993; Mody *et al.*, 1997).

Synthetic speech continua have also been used to determine discrimination thresholds. For this measurement of fine-grained discrimination, subjects are presented with stimuli that become increasingly similar along a particular acoustic dimension according to an adaptive procedure. The discrimination threshold, or just noticeable difference score (jnd), is determined as the point at which the subject’s ability to discriminate two stimuli reaches a preset criterion, such as 70% correct. Using this kind of task, several studies have shown that disabled children require greater acoustic distances between stimuli along certain synthetic speech continua in order to tell them apart (Elliott and Hammer, 1988; Elliott *et al.*, 1989; Kraus *et al.*, 1996; Stark and Heinz, 1996a).

The third variable to be manipulated across the numerous studies on this topic is the stimulus variable. While task-by-stimulus (e.g., Sussman, 1993) and population-by-stimulus (e.g., Frumkin and Rapin, 1980) interactions have been observed, several general stimulus-related effects have been reported in the literature, which we have summarized in Table I. This list is not exhaustive; nevertheless, it provides an overview of the kinds of stimulus-related factors that various researchers have investigated. With respect to consonant contrasts, voicing and place-of-articulation appear to be highly vulnerable to perceptual disruption. For voicing contrasts, the critical acoustic dimension is temporal (voice onset time). For place-of-articulation contrasts the critical acoustic dimension is either dynamic-spectral, as in the case of stop-vowel syllables, or static-spectral, as in the case of fricatives. In contrast to voicing and place-of-articulation features, the stop-glide manner feature has not been shown to provoke perceptual difficulty, nor has a combination of place, manner, and voicing features. In the case of the stop-glide contrast, the critical acoustic dimension is temporal in nature (formant transition duration.) In the case of a combination of place, manner, and voicing features, the stimulus contrasts do not differ minimally along an acoustic dimension, but rather the combination of various features differentiates the two stimuli, thus making them acoustically and phonetically more distinct.

For vowels, perceptual difficulties have been shown to arise, particularly for contrasts between vowels that are close together in the acoustic-phonetic vowel space, that is, for vowels that are acoustically similar. A contrast between peripheral vowels that are at the edges of the acoustic-phonetic vowel space (/i-æ-a-u/) has only been shown to provoke difficulty when embedded in a tri-syllabic string (/dab\_ba/). A contrast between vowels that are more similar in terms of their formant frequencies (/ε-æ/) was shown to be less resistant to perceptual difficulty for impaired subjects. However, the data for this vowel contrast are somewhat contradictory. While a recent study found impaired perception for the /ε-æ/ contrast, but not for the /a-i/ contrast, regardless of stimulus length (Stark and Heinz, 1996b), earlier studies found normal perception for the /ε-æ/ contrast when the stimuli were

TABLE I. Summary (nonexhaustive) of findings regarding stimulus factors and impaired speech perception in children with learning problems.

Contrast	Impaired perception?	Study
<b>Consonants</b>		
<b>Voicing</b>		
/ba-pa/	Yes	9
/da-ta/	Yes	5
<b>Place</b>		
/ba-da/ short FTD	Yes	1,3,4,5,6,8 9,10,11,15
/ba-da/ varying FTD	Yes	12
/da-ga/ short FTD	Yes	3,4,9,14
/sa-ʃa/	Yes	2
/das-daʃ/	Yes	10
/pʌ-tʌ/ short FTD	Yes	7
/ba-da/ long FTD	No	2
<b>Manner</b>		
/ba-wa/	No	14
<b>Voicing+Place+Manner</b>		
/ba-sa/	No	15
/da-ʃa/	No	15
<b>Vowels</b>		
<b>Dissimilar</b>		
/a-i/ long (40–240 ms)	No	13
/a-i/ short (10–40 ms)	No	13
/a-æ/ /d_b/ (250 ms)	No	5,10
/i-u/ 100 ms	No	10
/i-u/ /dab_ba/ (100 ms)	Yes	10
<b>Similar</b>		
/ε-æ/ long (250 ms)	No	1,8
/ε-æ/ short (40 ms)	No	5
/ε-æ/ /_I/ (40 ms)	Yes	2
/ε-æ/ long (40–240 ms)	Yes	13
/ε-æ/ v. short (10–40 ms)	Yes	13

FTD=Formant transition duration; 1=Tallal and Piercy, 1974; 2=Tallal and Piercy, 1975; 3=Brandt and Rosen, 1980; 4=Godfrey *et al.*, 1981; 5=Tallal and Stark, 1981; 6=Werker and Tees, 1987; 7=De Weirdt, 1988; 8=Reed, 1989; 9=Elliott *et al.*, 1989; 10=Leonard and McGregor, 1992; 11=Sussman, 1993; 12=Stark and Heinz, 1996a; 13=Stark and Heinz, 1996b; 14=Kraus *et al.*, 1996; 15=Mody *et al.*, 1997.

either long (250 ms) or short (40 ms), provided that they were presented in isolation with no immediately following phonetic material (Tallal and Piercy, 1974, 1975). It is likely that this apparent cross-study discrepancy regarding the /ε-æ/ contrast is due to differences in both stimulus and task details. For our purposes, however, the important point is that this contrast has been known to provoke perceptual difficulty, whereas contrasts between highly dissimilar vowels have generally been shown to be resistant to perceptual breakdown.

In the present study, we focused on a population of school-aged children with learning problems, a task that measured discrimination thresholds along synthetic speech continua, and stimuli that allowed us to investigate some of the precise acoustic-phonetic characteristics of stop consonant place-of-articulation contrasts that provoke perceptual difficulty. Concurrent with the behavioral speech perception measurement, we investigated the neurophysiologic encoding of these stimulus contrasts. The specific questions that we addressed in this study were: (1) Does lengthening the consonant–vowel formant transition duration in a synthetic

/da-ga/ continuum result in improved discrimination thresholds for children known to have elevated thresholds along this place-of-articulation continuum? (2) How are behaviorally determined discrimination abilities in normal and impaired subject groups reflected in an electrophysiologic response that is known to reflect the brain's preattentive representation of a change from one auditory stimulus to another? The answer to the first question should contribute directly to our understanding of the precise acoustic-phonetic features that provoke perceptual difficulties. The answer to the second question should contribute to our understanding of the nature of the underlying deficit as stemming from faulty stimulus encoding at a neural level, deficient representation at an acoustic-phonetic level, or a combination of both these factors. Taken together, the responses to these two questions should provide information that is relevant to the design of auditory training procedures for children with speech perception deficits.

Formant transition duration in stop-vowel syllables has been the subject of several previous studies. Tallal and Piercy (1974, 1975) claimed that children with developmental aphasia had impaired perceptual abilities relative to normal children for /ba-da/ syllables with short (43 ms) formant transition durations, but not for /ba-da/ syllables with long (95 ms) formant transition durations. They interpreted this finding as indicating that lengthened transition durations resulted in improved perception due to the increased processing time allowed by the longer stimulus duration. This view of the underlying perceptual impairment as resulting from a basic deficit in the processing of rapidly changing signals led to the development of a training program which exposes impaired children to speech that has been modified so that the fast transitional elements are lengthened and amplified (Tallal *et al.*, 1996; Merzenich *et al.*, 1996).

In a critical assessment of this approach to the problem, Mody *et al.* (1997) claimed that the apparent improved perception of /ba-da/ stimuli with lengthened formant transition durations may instead be due to increased phonetic distance across the place-of-articulation contrast, which results from the glidlike quality of the stimuli with lengthened formant transitions (see also Reidel and Studdert-Kennedy, 1985). Thus, in their view, perception of this place-of-articulation contrast is affected far more by the phonetic nature of the contrast than by the processing time allowed for by the stimulus duration. Central to their view is the finding that nonspeech sine wave analogs of the second and third formants in the /ba-da/ contrast did not provoke perceptual difficulties. In other words, these acoustic elements were vulnerable to perceptual disruption only when they formed part of a speech contrast.

Our broad research agenda is aimed specifically at investigating the neural encoding and perception of speech signals. Thus, we do not directly address the question of whether the perceptual deficits exhibited by impaired children are general auditory or speech specific in nature. Rather, by focusing on the effect of lengthening the formant transition duration in a /da-ga/ continuum, we hoped to add to the

empirical data addressing the nature of the underlying perceptual deficit as it pertains to speech perception in particular.

## I. METHOD

### A. Subjects

A group of 104 school-aged children served as subjects: 72 were classified as normal, and 32 were classified as having learning problems. The normal children had no history of learning or attention problems (based on a detailed parent questionnaire) and scored within normal limits (including no discrepancy between ability and achievement) on all tests in a psychoeducational test battery that was administered as part of our complete testing protocol. We therefore refer to this group as the WNL (within normal limits) group. (The psychoeducational test battery included portions of the Woodcock-Johnson Psycho-educational Battery, Woodcock-Johnson Psycho-educational Battery-Revised, and the Wide Range Achievement Test, 3rd ed.). The children with learning problems had been formally clinically diagnosed (prior to entry into the study) as having a learning disability ( $n = 11$ ), attention deficit disorder ( $n = 14$ ), or both ( $n = 7$ ). Furthermore, these impaired subjects had significantly lower scores than the normal subjects on all tests in the psychoeducational test battery. We therefore refer to this group as the LP (learning problems) group. All children had normal intelligence (Brief Cognitive Score  $> 85$ ) and normal hearing. (The Brief Cognitive Score is a measure of overall mental aptitude. Normal hearing was defined as thresholds better than 20 dB HL for 500–8000 Hz.)

It is important to note that the psychoeducational test battery administered as part of our testing protocol served as a study-internal verification of subject group assignment. In order for a subject to be assigned to the LP group, she or he had to have a formal (professional) diagnosis, as well as below normal scores on our test battery. Similarly, in order for a subject to be assigned to the WNL group, she or he had to have no clinical diagnosis and normal scores on our test battery.

The experimental design specifically included children from various clinical diagnostic categories. The rationale behind these broadly defined inclusion criteria stems from our larger scale studies which have revealed that biologic and perceptual deficits cut across diagnostic categories such as learning disability, attention deficit disorder, and dyslexia. Our previous report on 91 children with various diagnostic categories (learning disability, attention deficit disorder, or both) indicated that 35%–40% of children in each of these categories have fine-grained auditory perceptual deficits (Kraus *et al.*, 1996). Ongoing accumulated data on 161 LP children uphold the observation that these deficits cut across diagnostic categories and are not correlated with any specific diagnostic group. Thus, a child with an auditory perceptual deficit and a diagnosis of learning disabled may have more in common (perceptually) with a child with similar perceptual deficits with another diagnostic category, such as attention deficit disorder, than with another learning-disabled child who does not exhibit an auditory perceptual deficit.

The children ranged in age from 6 to 16 years. As reviewed elsewhere (Kraus *et al.*, in press), there is no documented effect of age on either the psychophysical or neurophysiologic measures used in this study. Moreover, IQ does not vary with these psychophysical measures at any age.

### B. Stimuli

Two /da-ga/ place-of-articulation continua were created using the Klatt cascade-parallel formant synthesizer (Klatt, 1980). All stimuli in both continua consisted of a formant transition period followed by a 60 ms steady-state period. There was no release burst. For all stimuli in both continua, the first and second formant onset frequencies were 220 and 1700 Hz, respectively. The third formant onset frequency varied from 2580 (/da/) to 2180 Hz (/ga/) in 40 steps of 10 Hz each. In the first continuum (daga40), these three formants reached the vowel steady-state frequencies of 720, 1240, and 2500 Hz over a period of 40 ms, giving a total stimulus duration of 100 ms. In the second continuum (daga80), this formant transition duration was lengthened to 80 ms, giving a total stimulus duration of 140 ms. For stimuli in both continua, the fourth and fifth formants were held constant across the entire stimulus duration at 3600 and 4500 Hz, respectively. Bandwidths were set as follows:  $F1 = 60$  Hz,  $F2 = 90$  Hz,  $F3 = 150$  Hz,  $F4 = 200$  Hz, and  $F5 = 200$  Hz. Each stimulus started with  $F0$  set at 100 Hz. The fundamental frequency rose linearly to 125 Hz over the first 35% of the stimulus duration, and then fell linearly to 80 Hz over the remainder of the stimulus duration.

Short syllable durations (100 and 140 ms for daga40 and daga80, respectively) were chosen in order to facilitate the use of identical stimuli in the behavioral and electrophysiologic tests. The use of short stimuli in the electrophysiologic test optimized the recording session by increasing the number of responses it was possible to record in a given time period. Pilot behavioral testing confirmed that these stimuli were appropriate for use with the target subject population. Additionally, a very small step size was chosen for this  $F3$  onset frequency continuum in order to maximize the chance that significant perceptual differences along this continuum between the WNL and LP subject groups would be revealed.

As a task control condition, a stop-glide continuum going from /ba/ to /wa/ was also created. It was expected that discrimination along this continuum would be relatively easy for impaired subjects (Kraus *et al.*, 1996), and therefore that equivalent performance in this condition across subject groups would establish that any group differences on the /da-ga/ continua were not due to general task-related difficulties. Stimuli in this continuum consisted of a formant transition period followed by a steady-state period, with no release burst. For all stimuli, the first and second formant onset frequencies were set at 234 and 616 Hz, respectively. These two formants then rose linearly to reach their steady state values of 769 and 1232 Hz, respectively. The duration of this formant transition period varied from 10 (the /ba/ end of the continuum) to 40 ms (the /wa/ end of the continuum) in 30 steps of 1 ms each. For all stimuli in this continuum the total stimulus duration was 100 ms. The third, fourth, and fifth formants were held constant across the entire stimulus



duration at 2862, 3600, and 4500 Hz, respectively. Bandwidths were set as follows:  $F1=60$  Hz,  $F2=90$  Hz,  $F3=150$  Hz,  $F4=200$  Hz, and  $F5=200$  Hz. Each stimulus started with  $F0$  set at 100 Hz. The fundamental frequency rose linearly to 120 Hz over the first 11 ms of the stimulus, and then fell linearly to 80 Hz over the remainder of the stimulus duration.

## C. Procedure

### 1. Behavioral discrimination thresholds

All subjects participated in both behavioral and neurophysiologic tests of their ability to discriminate stimuli along the two /da-ga/ continua (test conditions) and the /ba-wa/ continuum (control condition). Behavioral discrimination thresholds were determined using an adaptive tracking algorithm (Parameter Estimation by Sequential Tracking) with a four-interval AX discrimination task (4IAX). On each trial in this task, the subject was presented with two pairs of stimuli from a given continuum. In one pair, the two stimuli were identical; in the other pair, the stimuli were different. The subject responded by indicating on a two-button response box which pair was different. Feedback was provided for each response. One end of the stimulus continuum was designated the “anchor” end, and the “same pair” was always two presentations of this anchor stimulus. In the “different pair” the anchor stimulus and another stimulus from the continuum were presented. The PEST algorithm is designed so that the stimuli in the “different pair” get increasingly harder to discriminate (i.e., closer together on the continuum), and the subject’s just-noticeable-difference score (jnd) is determined by the distance between stimuli in the “different pair” that is required to achieve a performance level of 69% correct (Carrell *et al.*, 1999). In order to ensure an accurate jnd score, at least two measurements were taken. If the two scores differed widely, then a third measurement was taken. The final jnd score was then calculated as the average of the two best scores.

In the present study, both the daga40 and daga80 continua consisted of stimuli whose  $F3$  onset frequency varied from 2580 to 2180 Hz in 40 steps of 10 Hz. Thus, assuming the subject could reliably discriminate between the anchor stimulus (stimulus 1) and the stimulus at the other extreme of the continuum (stimulus 41), the range of possible jnd scores in these test conditions was 1–40, with each step representing an  $F3$  onset difference of 10 Hz. In both cases, stimulus 1 ( $F3$  onset=2580 Hz) was designated as the anchor stimulus. Thus, for example, a jnd of 5 indicated that the subject could discriminate stimulus 1 from stimulus 5 with 69% accuracy, that is, the subject needed an  $F3$  onset frequency difference of 40 Hz ( $F3=2580$  Hz versus  $F3=2540$  Hz) in order to tell the stimuli apart 69% of the time.

In the control condition (bawa), the transition duration varied from 10 to 40 ms in 30 steps of 1 ms each. Thus, the range of possible jnd scores was 1–30 with each step representing a transition duration difference of 1 ms. In this condition, the stimulus with the 40-ms transition duration was designated as the anchor stimulus. A jnd of 7, for example,

indicated that the subject could discriminate stimuli with transition durations of 40 and 33 ms with 69% accuracy.

The total test time required for obtaining an average jnd score for each stimulus continuum was approximately 30 min per subject. This included time for instructions to the subject and breaks between individual measurements during which subjects selected “prizes” such as candy, stickers, pencils, etc.

### 2. Neurophysiology

In order to probe the representation of the daga40 and daga80 stimulus contrasts at a preattentive neural level, mismatch negativity (MMN) responses were recorded from all subjects to stimulus pairs from each of the two place of articulation continua, as well as from the stop-glide continuum. The mismatch negativity is a passively elicited cortical evoked potential that is known to reflect the brain’s response to an acoustic change (Näätänen, 1995). It is elicited with an “oddball” stimulus delivery paradigm, in which a “deviant” stimulus is presented embedded in a string of “standard” stimuli. The MMN is seen as a negative deflection around approximately 200 ms after stimulus presentation. It is most clearly noticeable in the difference between the wave in response to the stimulus when presented in a string of identical stimuli and the wave in response to the same stimulus when presented as a deviant stimulus in a string of contrasting standard stimuli. During recording, the subject is engaged in an unrelated task such as watching a video. Thus, it provides a neurophysiologic index of the brain’s representation of an acoustic contrast at a preattentive neural level. For extensive discussion of MMN characteristics, generators, and research applications, see Näätänen *et al.* (1978) and Näätänen and Kraus (1995).

Stimulus pairs for the present study were selected based on pilot behavioral testing which indicated that an  $F3$  onset difference of 80 Hz between the standard and deviant stimuli would constitute a MMN recording between stimuli near the behavioral discrimination threshold along a /da-ga/ continuum for normal adult listeners. Accordingly, for both the daga40 and the daga80 continua, the /da/ endpoint stimulus ( $F3$  onset=2580 Hz) and the stimulus eight steps away on the continuum ( $F3$  onset=2500 Hz) were selected as the deviant and standard stimuli, respectively. The same selection criterion was applied to the control (bawa) condition. In this case, the standard and deviant stimuli had transition durations of 35 and 40 ms, respectively.

A detailed description of procedures used to record the MMN in school-aged children is described elsewhere (Kraus *et al.*, 1993, 1996), therefore here we provide only an abbreviated description of the recording procedures. Subjects were seated in a comfortable reclining chair in a sound-treated booth. During data collection they watched a video of their own choosing. Stimuli were presented to the subject’s right ear at approximately 75 dB SPL through an insert earphone. Subjects listened to the movie soundtrack through the left ear (free field, no earphones). In both the daga40 and the daga80 conditions the stimulus onset asynchrony was 590 ms; in the bawa condition the stimulus onset asynchrony was 710 ms. In all conditions, deviant stimuli were randomly interspersed

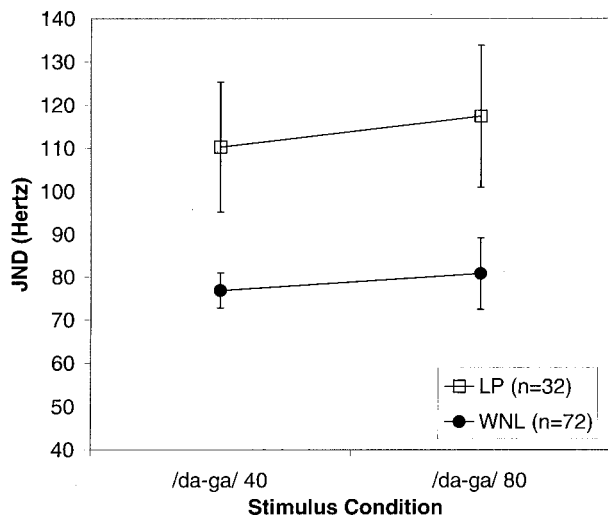


FIG. 1. Discrimination thresholds for the normal children (WNL) and children with learning problems (LP) along the daga40 (short transition) and daga80 (long transition) stimulus continua.

between standard stimuli with a minimum of three standard stimuli between deviants. Frequency of occurrence of the deviant stimulus was 10%. In each recording session, responses to approximately 200–250 deviant stimuli were recorded. In addition to the oddball presentation paradigm, responses were recorded to approximately 1000 presentations of the deviant stimulus presented alone. The wave of interest is the average “deviant” wave minus the average “alone” wave. That is, we were interested in the comparison between the response to the deviant stimulus when presented embedded in stream of standard stimuli versus the response to the same stimulus when presented alone. Neurophysiologic representation of the difference between the standard and the deviant stimulus is manifested in this difference wave as a negativity around approximately 200 ms after stimulus presentation. Data were recorded from nine active scalp electrode sites in accordance with the 10–20 recording system (American Electroencephalographic Society, 1991). Previous studies have shown the MMN to be robust at the *Fz* (frontal, midline) location, therefore in this paper we present only recordings from that site. [For additional information regarding MMN data analysis techniques, see McGee *et al.* (1997).]

## II. RESULTS

### A. Behavioral discrimination thresholds

Figure 1 shows the discrimination thresholds for the normal children (WNL) and children with learning problems (LP) along the daga40 (short transition) and daga80 (long transition) stimulus continua. In both conditions, the normal children required an *F3* onset frequency difference of approximately 75–80 Hz in order to discriminate the stimuli with 69% accuracy. The children with learning problems required an *F3* onset frequency difference of approximately 110–117 Hz in order to perform at this same level. For both groups, there was no difference in their discrimination thresholds along the daga40 and daga80 continua. A two-factor repeated measures ANOVA with Group (WNL vs.

LP) and Stimulus (daga40 vs. daga80) as factors showed a significant main effect of Group [ $F(1,176) = 7.56, p < 0.05$ ]. The main effect of Stimulus was not significant, nor was the Group by Stimulus interaction. In other words, in both conditions the LP subjects had elevated discrimination thresholds relative to the WNL subjects; and, for both groups, the lengthened transition duration had no effect on discrimination threshold.

In the /ba-wa/ task control condition, both subject groups had jnd's of approximately 7, indicating that they needed transition duration differences of approximately 7 ms in order to tell the stimuli apart. Importantly, there was no difference in performance in this condition across the two groups. This equivalent performance established that the group difference observed in the daga40 and daga80 conditions was not due to a general task-related difficulty on the part of the impaired subjects. Rather, the elevated discrimination thresholds in the /da-ga/ conditions were stimulus-related (see also Kraus *et al.*, 1996).

### B. Neurophysiology

Figure 2 shows grand averaged waveforms from the normal children (WNL,  $n = 72$ ) and children with learning problems (LP,  $n = 32$ ) to the daga40 and daga80 stimulus pairs selected for the MMN recordings. Recall that these stimuli were selected such that their *F3* onset frequencies differ by 80 Hz, and the waves shown in the figure come from the *Fz* (frontal, midline) recording site. In each quadrant of the figure the upper two waveforms show the electrophysiologic response to the deviant stimulus (*F3* onset frequency = 2580 Hz) when presented as a rare stimulus in a string of standard stimuli (*F3* onset frequency = 2500 Hz) and when presented alone. The lower waveform represents the difference between these two waveforms at each point in the recording window, with the horizontal line at zero. In all cases, positive is up. The boxes below the difference wave indicate the points at which the difference wave is significantly different from zero at the  $p < 0.01$  and  $p < 0.001$  levels (by one-group, two-tailed *t*-tests). The height of these boxes differentiates the two levels of significance (see labels in the lower right quadrant). These grand averaged waveforms allow us to gain initial insight into the overall response patterns across groups, and as such are useful indicators of intergroup central tendency variation.

In all four cases shown in Fig. 2, we see evidence of a MMN response. That is, in all cases there is a period of negativity starting at around 200 ms after stimulus onset. However, the overall magnitude of this negative deflection appears to vary across groups and stimuli. These variations in MMN responses are seen most clearly in Fig. 2 by the duration of the highly significant negative periods, that is, by the extent of the boxes along the bottom of the plots. These grand averaged waveforms show that the LP group had diminished responses relative to the normal group in the short transition duration condition (daga40). However, the impaired and normal groups had more similar responses in the long transition duration condition (daga80). Similarly, in the LP group, the lengthened transition duration stimuli resulted in an enhanced neurophysiologic response relative to the

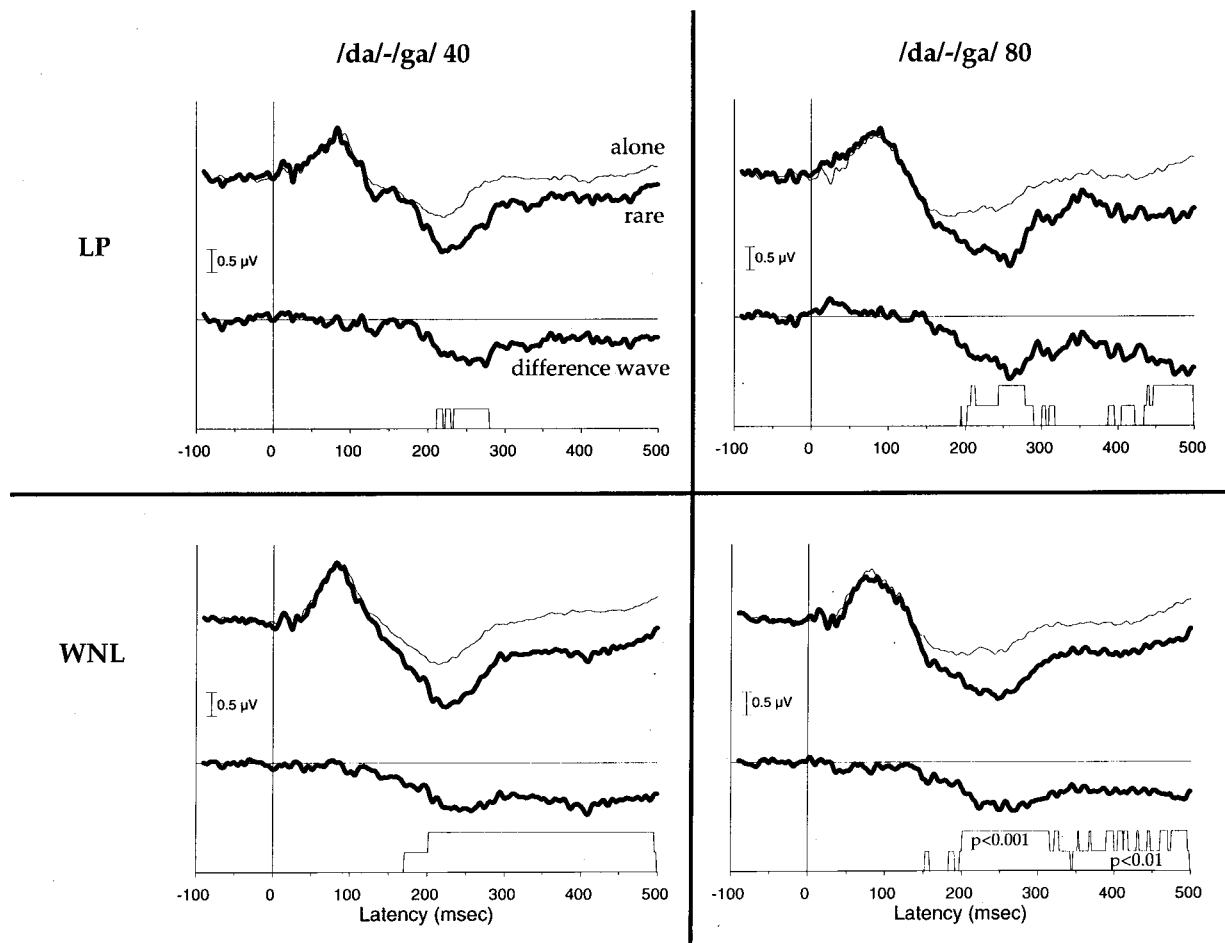


FIG. 2. Grand averaged waveforms from the normal children (WNL,  $n=72$ ) and children with learning problems (LP,  $n=32$ ) to the daga40 and daga80 stimulus pairs selected for the MMN recordings. Waves shown in the figure come from the  $Fz$  (frontal, midline) recording site. In each quadrant, the upper two waveforms show the electrophysiologic response to the deviant stimulus ( $F3$  onset frequency=2580 Hz) when presented as a rare stimulus in a string of standard stimuli ( $F3$  onset frequency=2500 Hz) and when presented alone. The lower waveform, represents the difference between these two waveforms at each point in the recording window, with the horizontal line at zero. In all cases positive is up. The boxes below the difference wave indicate the points at which the difference wave is significantly different from zero at the  $p < 0.01$  and the  $p < 0.001$  levels (by one-group, two-tailed  $t$ -tests). The height of these boxes differentiates the two levels of significance (see labels in the lower right quadrant).

short transition duration stimuli. In the normal group, the lengthened transition duration had no effect on neurophysiologic responses. Thus, the overall impression of the grand averaged waveforms is that the MMN in the daga40 condition for the group of children with learning impairments (Fig. 2, top left panel) is diminished relative to the other three cases.

In order to quantify this general impression from the grand averaged waveforms, the area of negativity was calculated for each individual subject's difference wave. This measure was calculated as the area between the difference wave and the zero line, and took into account both the duration and the amplitude of the MMN response. A large area (e.g., greater than  $450 \mu V \times ms$ ) indicated a robust MMN response, whereas a small area (e.g., less than  $100 \mu V \times ms$ ) indicated a diminished MMN response. Figure 3 shows the percentage of subjects from each group with large and small MMN responses (as defined above) for each stimulus condition. By focusing our analyses on the extreme ends of the MMN area distributions for the two subject groups we could observe the relative robustness of MMN response across the two subject populations. In this figure, as in the grand aver-

aged waveforms shown in Fig. 2, we see a different pattern of results for the impaired subjects in the daga40 condition (open squares) in comparison to the pattern of responses for the other three cases. In the daga40 condition for the impaired subjects, the distribution of MMN areas was skewed towards small areas. In contrast, the distributions of MMN areas were skewed towards large areas in each of the other three cases. It is important to note that, in all four cases, the majority of individual subjects showed MMN responses with areas between 100 and  $450 \mu V \times ms$ . However, group and stimulus differences were revealed when we examined the edges of the MMN area distributions.

Specifically, in the impaired subject group for the daga40 condition, 22% of the individual subjects had very small ( $< 100 \mu V \times ms$ ) MMN areas, whereas only 9% had very large MMN areas ( $> 450 \mu V \times ms$ ). In contrast, only 11% of the normal subjects had very small MMN areas ( $< 100 \mu V \times ms$ ), whereas 28% had very large MMN areas ( $> 450 \mu V \times ms$ ) in this stimulus condition. A significant chi-square statistic [chi-square (1) = 5.29,  $p = 0.021$ ] indicated that the distribution of MMN areas for the impaired subjects was significantly different from that of the normal subjects

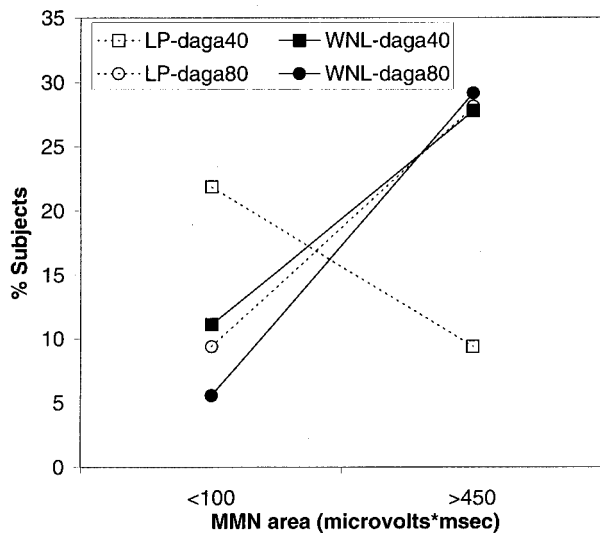


FIG. 3. Percentage of subjects from each group with very large and very small MMN areas for each stimulus condition.

for the daga40 condition. In the daga80 condition, the distribution of MMN areas was similar across the two subject groups. A relatively small percentage of both subject groups in this condition had very small MMN areas: 9% and 6% for the impaired and normal groups, respectively. A considerably larger percentage of both subject groups had large MMN areas in this condition: 28% and 29% for the impaired and normal groups, respectively. In this case, the chi-square statistic showed that the MMN area distributions were not significantly different across the two subject groups.

An examination of the MMN area distributions within each group of subjects across the two stimulus conditions showed a significant difference in MMN area distributions between the daga40 and daga80 conditions in the impaired group [chi square (1) = 6.89,  $p = 0.032$ ], but not in the normal group. In the impaired group, a considerably larger percentage of subjects had small MMN areas than large MMN areas (22% versus 9%) for the daga40 condition, whereas, in the daga80 condition, a considerably larger percentage of subjects had large MMN areas than small MMN areas (28% versus 11%). For the normal subjects, in both conditions the distributions were skewed towards larger MMN areas. Finally, in the control condition with the /ba-wa/ contrast, both the grand averaged waveforms and the MMN area distributions showed equivalent responses in the impaired and normal subject groups.

In summary, an examination of the MMN responses from the same children who participated in the behavioral tests described above showed that responses from the group of impaired subjects in the short transition duration condition tended to be diminished relative to their responses in the long transition duration condition, as well as relative to the responses from the group of normal children in both stimulus conditions. This pattern of MMN responses was evident in the grand averaged waveforms, which provided an indication of the responses across the whole group of subjects. Additional support for this response pattern was obtained by examining the distribution of MMN area measures in indi-

vidual subjects within each subject group in each stimulus condition. In this analysis, we found that small MMN areas were more frequent than large MMN areas in the impaired subject group in the daga40 condition. However, in the daga80 condition, large MMN areas were more frequent than small MMN areas for this same group of subjects. For the normal subjects, in both stimulus conditions, large MMN areas were more frequent than small MMN areas.

Given that in all cases a high proportion of subjects (approximately 61%–69%) showed intermediate MMN areas (between 100 and 450  $\mu\text{V ms}$ ), these neurophysiologic data need to be interpreted with caution. That is, there was a high degree of overlap across groups and stimulus conditions, indicating that MMN response magnitude on its own is not a good basis for subject or stimulus categorization. Nevertheless, these groupwise neurophysiologic data, in combination with the behavioral discrimination data, can provide meaningful initial insights into the nature of the underlying auditory-perceptual deficit that many children with learning problems contend with.

### III. GENERAL DISCUSSION

Taken together, these behavioral and neurophysiologic data suggest that the source of the underlying perceptual deficit may be a combination of faulty stimulus representation at the neural level as well as deficient perception at an acoustic-phonetic level. Specifically, the group of children with diagnosed learning problems exhibited a behavioral deficit in discriminating minimally different stimuli along a /da-ga/ continuum regardless of whether the critical formant transition was relatively short or long in duration. The neurophysiologic data indicated that, in the short transition duration condition, a stimulus contrast below the impaired group's mean discrimination threshold was not well represented at a preattentive neural level in this subject group. Thus, in this case, the neurophysiologic data paralleled the behavioral data, suggesting a biological basis for the impaired behavioral perception (see also Kraus *et al.*, 1996). In the long transition duration condition, the neurophysiologic data indicated that the F3 contrast was more robustly represented in the impaired subject group at the neural level than the short transition duration stimuli; however, the behavioral data showed that their impaired perception of this contrast persisted despite the lengthened transition duration.

While the behavioral and physiologic responses are related in that they reflect auditory pathway representation of acoustic events, it must be remembered that these are inherently *different* responses which represent different levels of processing. That is, the neurophysiologic response is a preattentive neural representation of acoustic change, whereas the behavioral response also involves focused attention and later phonetic processing. Thus, the convergent behavioral and neurophysiologic findings in the short transition duration condition, in combination with the more divergent behavioral and neurophysiologic findings in the long transition duration condition, suggest that robust stimulus representations at both levels of processing are necessary for fine-grained discrimination of the /da-ga/ contrast. In the short transition duration condition, faulty stimulus encoding at the neural



level appeared to underlie the observed elevated discrimination threshold in the LP subject group. In the long transition duration condition, regardless of the relatively robust stimulus representation at the neural level, perceptual discrimination was still impaired in this subject group, thus implicating a breakdown at higher levels of processing.

The present behavioral results replicate previous findings reported in the literature showing impaired place-of-articulation perception for disabled children relative to normal children (see Table I and references therein). Furthermore, based on these behavioral findings, it appears that simply lengthening the formant transition duration from 40 to 80 ms is not effective in enhancing discrimination between stimuli along a /da-ga/ continuum. This behavioral result seems to stand in contrast to the earlier finding of Tallal and Piercy (1975) who showed enhanced perception by children with developmental aphasia of a synthetic /ba-da/ contrast when the formant transition duration was lengthened from 43 to 95 ms. However, there are several subject-, task-, and stimulus-related factors that differentiate the two studies. The impaired subjects in the present study had diagnoses of learning disability, attention deficit disorder, or both, whereas in the Tallal and Piercy study the subjects were diagnosed as developmental aphasics. In the present study, our task determined discrimination thresholds along a /da-ga/ continuum. In contrast, Tallal and Piercy (1975) examined the identification, serial ordering, discrimination, and serial recall of /ba/ and /da/ stimuli that represented good exemplars of the /b/ and /d/ categories. Finally, in the present study, the duration of the steady-state portion of the stimuli remained constant across the daga40 and daga80 conditions; whereas in the Tallal and Piercy study the duration of the steady-state portion was shorter in the long formant transition condition than in the short formant transition condition. In proportional terms, the transition durations of the stimuli in the present study occupied 40% and 57% of the daga40 and daga80 stimuli, respectively. In the Tallal and Piercy stimuli, the transition durations occupied 17% and 38% of the short and long transition duration stimuli, respectively. These subject, task, and stimulus differences between the two studies may well account for the different results.

Nevertheless, the present finding of no perceptual discrimination enhancement with lengthened formant transition duration suggests that a revision of the general interpretation offered in the earlier studies by Tallal and colleagues is called for. Specifically, Tallal and Piercy (1975) concluded that "... it is the brevity not the transitional character of this component of synthesized consonants which results in the impaired perception..." (p. 73). This conclusion was based on their finding of impaired perception of synthetic /ba/ and /da/ with brief formant transition durations and /ɛɪ/ and /æɪ/ with brief initial vowels, but intact perception of these same CV syllables with extended formant transition durations and long steady-state /ɛ/ and /æ/. However, the present study (with subjects with very similar perceptual deficits to the Tallal and Piercy subjects) suggests that this earlier finding fails to generalize to other stimuli (from /ba-da/ to /da-ga/), and also fails to transfer to a novel task (from endpoint identification to discrimination along a continuum). Thus, there

appear to be stimulus- and task-specific (and perhaps also subject-specific) factors that interact with signal manipulations that attempt to enhance perception of synthetic speech sounds. The origin of these interactions remains to be identified. However, by examining the neurophysiologic encoding of the /da-ga/ contrast for both short and long formant transition duration stimuli, the present neurophysiologic data provided some insight into the relationship between the pre-attentive representation of this contrast at the neural level and its perceptual discrimination in a task requiring focused attention to the stimuli.

The present data raise the following, critical question: What bearing do these findings have on the design of remediation programs for children who exhibit speech perception deficits? In response to this question we offer the following speculations based largely on previous findings reported in the literature. Major generator sources for the MMN (obtained from animal models and modeled from scalp-recorded data in humans) include the extralemniscal auditory thalamocortical pathway (reviewed in Alho, 1995; Kraus *et al.*, 1994). Because the MMN has nonprimary, extralemniscal auditory CNS origins (Kraus *et al.*, 1994; Scherg and Picton, 1990) and these pathways inherently exhibit plasticity (Edelin and Weinberger, 1991; Kraus and Disterhoft, 1982), it is amenable to training. Moreover, during speech sound training, learning-associated neurophysiologic changes can be evident before learning is manifested behaviorally (Tremblay *et al.*, 1998). Consequently, the better neurophysiologic representation of the longer duration stimuli may underlie the success of training strategies which employ stimuli with lengthened formant transition stimuli (e.g., Tallal *et al.*, 1996; Merzenich *et al.*, 1996). Brief duration stimuli, which are poorly represented physiologically by the auditory CNS, may be difficult for children to access for training purposes. However, given that the present data showed that lengthening the critical formant transition duration on its own did not result in improved discrimination thresholds along the /da-ga/ continua, it is also likely that any effective training procedure will also need to include direct means of enhancing perception at a phonetic level.

The conclusions and interpretations of the present study are necessarily tentative due to various practical limitations that are imposed on any study of this scale with this subject population. One such constraint is the range of testing paradigms that can be included in the experimental protocol. Ideally, the behavioral data set would have included both discrimination and labeling functions from each individual subject for each stimulus condition. However, in order to avoid excessively long, tedious, and numerous testing sessions with the children in the study, performance on only one kind of task was included in the protocol. The discrimination task was selected because it directly reflects the subject's ability to perceive fine-grained acoustic contrasts, and therefore more closely parallels the task requirements of the odd-ball paradigm used to elicit the MMN than a labeling task. Clearly, the availability of corresponding labeling data would have made it possible to examine in greater detail the relationship between the present data and the Tallal and Piercy (1975) data.

A second limitation of the present data set is related to the auditory nature of the long transition duration stimuli. As a direct consequence of the lengthened transition durations for all of the first three formants, these stimuli lose their stoplike quality and sound somewhat glidelike. The construction of the stimuli for the present study was inspired by the Tallal and Piercy stimuli (which also had lengthened  $F1$ ,  $F2$ , and  $F3$ ); however, it is important to note that other studies have maintained the stoplike quality of this type of CV stimulus by lengthening only the higher formants while maintaining relatively short first formant transitions (Keating and Blumstein, 1978; Sussman and Carney, 1989). The effect of this manner-class shift between the short and long transition stimuli therefore remains an issue for future research.

As a final word, we would like to stress that a critical feature of the present study was the combined behavioral-neurophysiologic approach. The combination of the two sources of data provided insight into the underlying perceptual processes that neither technique could have provided alone. Furthermore, in combination with other studies of auditory-perceptual plasticity, we believe that we will be able to identify critical elements of training procedures that can eventually be incorporated into efficient, effective, and scientifically grounded intervention programs for children with speech perception deficits.

## ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD (Grant Nos. DC 01510 and DC 03762) and by the Foundation for Hearing and Speech Rehabilitation. We gratefully acknowledge the data collection and processing assistance of Audrey Cameron, Beth Perez, and Brad Wible. We also thank two anonymous reviewers for insightful and helpful suggestions for improving the manuscript.

Alho, K. (1995). "Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes," *Ear Hear.* **16**, 38–50.

American Electroencephalographic Society (1991). "Guidelines for standard electrode position nomenclature," *J. Clin. Neurophysiol.* **8**, 200–202.

Brandt, J., and Rosen, J. J. (1980). "Auditory phonemic perception in dyslexia: Categorical identification and discrimination of stop consonants," *Brain Language* **9**, 324–337.

Carrell, T. D., Bradlow, A. R., Nicol, T. G., Koch, D. B., and Kraus, N. (1999). "Interactive software for evaluating auditory discrimination," *Ear Hear.* **20**, 175–176.

De Weirdt, W. (1988). "Speech perception and frequency discrimination in good and poor readers," *Appl. Psycholinguist.* **9**, 163–183.

Edeline, J. M., and Weinberger, N. M. (1991). "Subcortical adaptive filtering in the auditory system: Associative receptive field plasticity in the dorsal medial geniculate body," *Behav. Neurosci.* **105**, 154–175.

Elliott, L. L., and Hammer, M. (1988). "Longitudinal changes in auditory discrimination in normal children and children with language learning problems," *J. Speech Hear. Disord.* **53**, 467–474.

Elliott, L. L., Hammer, M. A., and Scholl, M. E. (1989). "Fine-grained auditory discrimination in normal children and children with language-learning problems," *J. Speech Hear. Res.* **32**, 112–119.

Frumkin, B., and Rapin, I. (1980). "Perception of vowels and consonant-vowels of varying duration in language impaired children," *Neuropsychologia* **18**, 443–454.

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., and Knox, C. M. (1981).

"Performance of dyslexic children on speech perception tests," *J. Exp. Child Psychol.* **32**, 401–424.

Keating, P. A., and Blumstein, S. E. (1978). "The effects of transition length on the perception of stop consonants," *J. Acoust. Soc. Am.* **64**, 57–64.

Klatt, D. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.

Kraus, N., and Disterhoft, J. (1982). "Response plasticity of single neurons in rabbit auditory association cortex during tone-signaled learning," *Brain Res.* **246**, 205–215.

Kraus, N., McGee, T., Littman, T., Nicol, T., and King, C. (1994). "Encoding of acoustic change in non-primary auditory thalamus," *J. Neurophysiol.* **72**, 1270–1277.

Kraus, N., Koch, D. B., McGee, T. J., Nicol, T. G., and Cunningham, J. (in press). "Speech-sound discrimination in school-age children: psychophysical and neurophysiologic measures," *J. Speech Language Hear. Res.*

Kraus, N., McGee, T. J., Carrell, T. D., King, C., Tremblay, K., and Nicol, T. G. (1995). "Central auditory system plasticity associated with speech discrimination training," *J. Cogn. Neurosci.* **7**, 27–34.

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., and Koch, D. B. (1996). "Auditory neurophysiologic responses and discrimination deficits in children with learning problems," *Science* **273**, 971–973.

Kraus, N., McGee, T. J., Micco, A., Carrell, T. D., Sharma, A., and Nicol, T. G. (1993). "Mismatch negativity in school-age children to speech stimuli that are just perceptibly different," *Electroencephalogr. Clin. Neurophysiol.* **88**, 123–130.

Leonard, L. B., McGregor, K. K., and Allen, G. D. (1992). "Grammatical morphology and speech perception in children with specific language impairment," *J. Speech Hear. Res.* **35**, 1076–1085.

McGee, T. J., Kraus, N., and Nicol, T. G. (1997). "Is it really a mismatch negativity? An assessment of methods for determining response validity in individual subjects," *Electroencephalogr. Clin. Neurophysiol.* **104**, 359–368.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., and Tallal, P. (1996). "Temporal processing deficits of language-learning impaired children ameliorated by training," *Science* **271**, 77–81.

Mody, M., Studdert-Kennedy, M., and Brady, S. (1997). "Speech perception deficits in poor readers: Auditory processing or phonological coding?" *J. Exp. Child Psychol.* **64**, 199–231.

Näätänen, R. (1995). "The mismatch negativity: A powerful tool for cognitive neuroscience," *Ear Hear.* **16**, 6–19.

Näätänen, R., and Kraus, N. (editors) (1995). "Mismatch negativity as an index of central auditory function," *Ear Hear.* **16**, Special Issue.

Näätänen, R., Gaillard, A., and Mantysalo, S. (1978). "Early selective-attention effect on evoked potential reinterpreted," *Acta Psychol.* **42**, 313–329.

Reed, M. A. (1989). "Speech perception and the discrimination of brief auditory cues in reading disabled children," *J. Exp. Child Psychol.* **48**, 270–292.

Reidel, K., and Studdert-Kennedy, M. (1985). "Extending formant transitions may not improve aphasics' perception of stop consonant place of articulation," *Brain Language* **24**, 223–232.

Scherg, M., and Picton, T. (1990). "Brain electric source analysis of mismatch negativity," in *Psychophysiological Brain Research, Vol. 1*, edited by C. Brunia, A. Gaillard, and A. Kok (Tilberg U.P., Tilberg) pp. 94–98.

Stark, R. E., and Heinz, J. M. (1996a). "Perception of stop consonants in children with expressive and receptive-expressive language impairments," *J. Speech Hear. Res.* **39**, 676–686.

Stark, R. E., and Heinz, J. M. (1996b). "Vowel perception in children with and without language impairment," *J. Speech Hear. Res.* **39**, 860–869.

Sussman, J. E. (1993). "Perception of formant transition cues to place of articulation in children with language impairments," *J. Speech Hear. Res.* **36**, 1286–1299.

Sussman, J. E., and Carney, A. E. (1989). "Effects of transition length on the perception of stop consonants by children and adults," *J. Speech Hear. Res.* **32**, 151–160.

Tallal, P. (1980). "Auditory temporal perception, phonics, and reading disabilities in children," *Brain and Language* **9**, 182–198.

- Tallal, P., and Piercy, M. (1974). "Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception," *Neuropsychologia* **12**, 83–93.
- Tallal, P., and Piercy, M. (1975). "Developmental aphasia: The perception of brief vowels and extended stop consonants," *Neuropsychologia* **13**, 69–74.
- Tallal, P., and Stark, R. E. (1981). "Speech acoustic-cue discrimination abilities of normally developing and language-impaired children," *J. Acoust. Soc. Am.* **69**, 568–574.
- Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., and Merzenish, M. M. (1996). "Language comprehension in language-learning impaired children with acoustically modified speech," *Science* **271**, 81–84.
- Tremblay, K. M., Kraus, N., Carrell, T., and McGee, T. J. (1998). "The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training?" *Neuroreport* **9**, 3557–3560.
- Werker, J. F., and Tees, R. C. (1987). "Speech perception in severely disabled and average reading children," *Can. J. Psychol.* **41**, 48–61.

# On the number of channels needed to understand speech

Philipos C. Loizou<sup>a)</sup>

Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

Michael Dorman

Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287

Zhemín Tu

Department of Applied Science, University of Arkansas at Little Rock, Little Rock, Arkansas 72204-1099

(Received 5 December 1998; revised 7 April 1999; accepted 21 May 1999)

Recent studies have shown that high levels of speech understanding could be achieved when the speech spectrum was divided into four channels and then reconstructed as a sum of four noise bands or sine waves with frequencies equal to the center frequencies of the channels. In these studies speech understanding was assessed using sentences produced by a single male talker. The aim of experiment 1 was to assess the number of channels necessary for a high level of speech understanding when sentences were produced by multiple talkers. In experiment 1, sentences produced by 135 different talkers were processed through  $n$  ( $2 \leq n \leq 16$ ) number of channels, synthesized as a sum of  $n$  sine waves with frequencies equal to the center frequencies of the filters, and presented to normal-hearing listeners for identification. A minimum of five channels was needed to achieve a high level (90%) of speech understanding. Asymptotic performance was achieved with eight channels, at least for the speech material used in this study. The outcome of experiment 1 demonstrated that the number of channels needed to reach asymptotic performance varies as a function of the recognition task and/or need for listeners to attend to fine phonetic detail. In experiment 2, sentences were processed through 6 and 16 channels and quantized into a small number of steps. The purpose of this experiment was to investigate whether listeners use across-channel differences in amplitude to code frequency information, particularly when speech is processed through a small number of channels. For sentences processed through six channels there was a significant reduction in speech understanding when the spectral amplitudes were quantized into a small number ( $< 8$ ) of steps. High levels (92%) of speech understanding were maintained for sentences processed through 16 channels and quantized into only 2 steps. The findings of experiment 2 suggest an inverse relationship between the importance of spectral amplitude resolution (number of steps) and spectral resolution (number of channels). © 1999 Acoustical Society of America. [S0001-4966(99)01810-X]

PACS numbers: 43.72.Ar, 43.71.Es [JMH]

## INTRODUCTION

Dudley (1939) provided one of the earliest demonstrations that speech understanding does not require a highly detailed spectral representation of the speech signal. After bandpass filtering the speech signal into ten spectral bands, Dudley (1939) estimated the envelopes of the bandpassed waveforms using rectification and low-pass filtering (20-Hz cutoff). Speech was synthesized by filtering an excitation signal (either buzz or hiss) through the same bandpass filters, and amplitude modulating the outputs of the filters by the envelopes of the bandpassed waveforms. The resulting speech was highly intelligible. Dudley (1939) concluded that much of the information in the speech spectrum is redundant. The channel vocoder approach, pioneered by Dudley, was later exploited for efficient transmission of speech over telephone channels (see review by Schroeder, 1966; Flanagan, 1972).

In the 1950s, researchers at Haskins Laboratories used a 50-component sine wave synthesizer to investigate the mini-

mal cues necessary for the recognition of speech. Investigators showed that speech could be recognized with a high degree of accuracy when sine waves specifying only the first two or three formants of the signal were presented (e.g., Delattre *et al.*, 1952). In these experiments as few as four or six sine wave components (out of 50) were sufficient to create intelligible speech, if the sine wave components specified harmonics at or near the formant frequencies of the signal. Remez *et al.* (1981), elaborating on earlier work on syllable recognition by Cutting (1974) and Bailey *et al.* (1976), carried the minimal cues approach to one extreme by replacing the rich harmonic structure of speech with only three sine waves at the formant frequencies of the consonants and vowels in the words of sentences. Most listeners were able to identify the words with high accuracy.

The aforementioned studies, and many others (e.g., Hill *et al.*, 1968), provide overwhelming evidence that speech recognition does not require the fine spectral detail present in naturally produced utterances. This fortunate circumstance has proved essential in restoring speech understanding to deaf individuals fitted with cochlear implants, because it is not currently possible to provide fine spectral detail to im-

<sup>a)</sup>Electronic mail: loizou@utdallas.edu



plant patients. However, in the context of signal processing for cochlear implants, it is still unclear as to how little or how much spectral detail is necessary to allow speech understanding at a high level.

In the work cited above, high levels of speech understanding were obtained if signals were filtered into a reasonably large number of frequency bands and/or a small number of sine waves were output at or near the formant frequencies. Such a strategy is implemented in one of the two current signal processing strategies used for cochlear implants (McDermott *et al.*, 1992; Loizou, 1998). The other signal processing strategy used for cochlear implants divides the speech spectrum into a small number of bands, 4 to 12 depending on the device, and, instead of picking high amplitude channels, transmits the energy in all of the bands. This strategy is the focus of this article. At issue is how many channels of stimulation are necessary to achieve a high level of speech understanding in quiet.

Shannon *et al.* (1995) showed that high levels of speech understanding (e.g., 90% correct for sentences) could be achieved using as few as four spectral bands. In Shannon *et al.* (1995) envelopes of the speech signal were extracted from a small number (1–4) of frequency bands, and used to modulate noise of the same bandwidth. The noise-modulated bands preserved the temporal cues within each band but eliminated the spectral details within each band. Dorman *et al.* (1997) synthesized speech as a sum of a small number of sine waves rather than noise bands. As in Shannon *et al.* (1995), sentence recognition using four channels was found to be 90% correct.

In Shannon *et al.* (1995) and Dorman *et al.* (1997) speech understanding was assessed using sentences produced by a single male speaker. It is very likely that the use of a single speaker overestimates the speech perception abilities of listeners in real-world situations because the use of a single speaker eliminates the need for listeners to accommodate to variability in the acoustic signal (e.g., Mullenix *et al.*, 1989; Sommers *et al.*, 1997). Variability in the acoustic signal arises from differences in the size and shape of vocal tracts, differences in phonetic realization (e.g., pronunciation), and differences in speaking rate. The aim of experiment 1 was to determine the number of channels of stimulation necessary to allow a high level of sentence understanding when speech was produced by 135 talkers, half of whom were female.

The aim of experiment 2 was to assess the intelligibility of speech processed through 6 and 16 channels and quantized into a small number of steps. The purpose of this experiment was to assess the importance of amplitude resolution for speech understanding when signals are processed into a relatively small, and a relatively large, number of channels. Our hypothesis was that a relatively high degree of amplitude resolution is a necessary condition for speech understanding when signals are processed into a small number of channels because, with a small number of channels, listeners must use differences in signal levels across channels to infer the location of formant frequencies (Dorman *et al.*, 1997; Loizou *et al.*, 1998). In contrast, when speech is processed into a large number of channels, a high level of spec-

tral amplitude resolution is not necessary because the location of frequencies in the input spectrum are well specified by the channels which contain energy. The outcome of experiment 2 is of interest because a recent experiment by Nelson *et al.* (1996) with cochlear implant subjects showed that the total number of discriminable intensity steps varied from a low of 6 to a high of 45. If a high degree of amplitude resolution is necessary for frequency analysis when speech is processed into a small number of channels, then it is possible that speech perception in some cochlear implant subjects is constrained by a limited ability to resolve differences in signal level across channels.

## I. EXPERIMENT 1

### A. Method

#### 1. Subjects

Nine graduate students from the Applied Science Department, UALR, served as subjects. All of the subjects were native speakers of American English and had normal hearing. The subjects were paid for their participation.

#### 2. Sentence material

The multi-talker TIMIT database (Garofolo *et al.*, 1993) was used for testing. The TIMIT database contains speech from 630 speakers, representing 8 major dialect divisions of American English, each speaking 10 phonetically rich sentences. Some of the sentences were designed to provide a good coverage of pairs of phones with extra occurrences of difficult phonetic contexts and some of the sentences were designed to maximize the variety of allophonic contexts (Lamel *et al.*, 1986).

A total of 135 sentences were randomly selected from the TIMIT database from the DR3 (north midland) dialect region. The sentences were produced by an equal number of female and male speakers—one sentence per speaker. The 135 sentences were divided into 9 lists (1 list per channel condition), with 15 sentences in each list. Fifteen sentences were used for the first channel condition, 15 different sentences were used for the second channel condition, etc. There were eight sentences spoken by eight different male speakers and seven sentences spoken by seven different female speakers within each list. Each sentence contained, on the average, 7 words, and the 15 sentences in each list contained, on the average, a total of 100 words. Each subject listened to a total of 135 sentences (=15 sentences/condition $\times$ 9 channel conditions).

#### 3. Signal processing

Signals were first processed through a pre-emphasis filter (2000-Hz cutoff), with a 3-dB/octave rolloff, and then bandpassed into  $n$  frequency bands ( $n=2,3,4,5,6,8,10,12,16$ ) using sixth-order Butterworth filters. Logarithmic filter spacing was used for  $n<8$  and mel spacing<sup>1</sup> was used for  $n\geq 8$ . Logarithmic and semi-logarithmic (mel) filter spacing was used because: (1) the filter bandwidths can be computed systematically; and (2) it is the type of filter spacing used in current cochlear implant devices (e.g., Zierhofer *et al.*, 1994; Loizou, 1998).

TABLE I. The center frequencies (Hz) of the filters.

No. of Channels	Channel															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	792	3392														
3	545	1438	3793													
4	460	953	1971	4078												
5	418	748	1339	2396	4287											
6	393	639	1037	1685	2736	4444										
8	394	692	1064	1528	2109	2834	3740	4871								
10	322	546	814	1137	1524	1988	2545	3213	4014	4976						
12	274	453	662	905	1190	1521	1908	2359	2885	3499	4215	5050				
16	216	343	486	647	828	1031	1260	1518	1808	2134	2501	2914	3378	3901	4489	5150

The center frequencies and the 3-dB bandwidths of the filters are given in Tables I and II, respectively. The envelope of the signal was extracted by full-wave rectification, and low-pass filtering (second-order Butterworth) with a 400-Hz cutoff frequency. Sinusoids were generated with amplitudes equal to the root-mean-square (rms) energy of the envelopes (computed every 4 ms) and frequencies equal to the center frequencies of the bandpass filters. The phases of the sinusoids were estimated from the FFT of the speech segment<sup>2</sup> (McAulay and Quatieri, 1986). The sinusoids of each band were finally summed and the level of the synthesized speech segment was adjusted to have the same rms value as the original speech segment.

#### 4. Procedure

The experiment was performed on a PC equipped with a Creative Labs SoundBlaster 16 soundcard. The subjects listened to the sentences via closed ear-cushion headphones at a comfortable level set by the subject. A graphical interface was used that allowed the subjects to type the words they heard. After listening to each sentence, subjects were asked to type in as many words as they could understand.

Before each channel condition, subjects were given a practice session with examples of ten sentences processed through the same number of channels in that condition. None of the sentences used in the practice was used in the test. A sequential test order, starting with sentences processed through a large number of channels ( $n=16$ ) and continuing to sentences processed through a small number of channels ( $n=2$ ), was employed. We chose this sequential test design

to give the subjects time to adapt to listening to the altered speech signals. There is no doubt a “warm-up” effect when listening to sine wave speech of any kind.

#### B. Results and discussion

The subject’s responses were scored as percentage of words correct. The results are shown in Fig. 1. A repeated measures analysis of variance indicated a main effect [ $F(8,64)=261.94, p<0.0001$ ] for number of channels. *Post hoc* tests according to Scheffe showed no statistically significant differences in scores when the number of channels was increased beyond eight. There was a significant difference ( $p=0.001$ ) between the scores obtained with six and eight channels. There was no significant difference between the scores obtained with five and six channels. Speech recognition performance with four channels was 63% correct. This score was significantly lower than the score (90%) reported by Shannon *et al.* (1995) and the score (90%) reported by Dorman *et al.* (1997) using sentences from the H.I.N.T. database produced by a single male talker. This outcome, as well as others, demonstrates that the number of channels necessary to reach asymptotic performance varies as a function of the task and/or need for a listener to attend to acoustic/phonetic detail.

In our study, the task was recognition of speech produced by multiple speakers. Four channels did not seem to be sufficient for achieving high level of sentence understanding. To see why consider, in Fig. 2(a), the channel spectrum of the vowel [ε] (‘head’), spoken by a male talker, and

TABLE II. The 3-dB bandwidths (Hz) of the filters.

No. of Channels	Channel															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	984	4215														
3	491	1295	3414													
4	321	664	1373	2842												
5	237	423	758	1356	2426											
6	187	304	493	801	1301	2113										
8	265	331	431	516	645	805	1006	1257								
10	204	244	293	352	422	506	607	729	874	1049						
12	165	193	225	262	306	357	416	486	567	661	771	900				
16	120	135	151	170	192	216	242	273	307	345	389	437	492	553	622	700

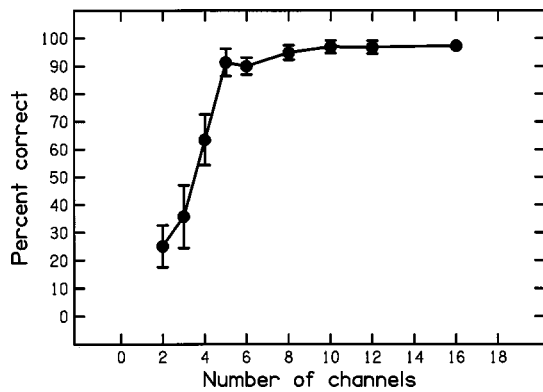


FIG. 1. Sentence understanding (percent correct) as a function of number of channels. Error bars indicate  $\pm 1$  standard deviation.

processed through four channels. Four channels are sufficient to code the frequency of  $F1$  and  $F2$ . The  $F1$  of  $[\epsilon]$  is coded by a high-amplitude in channel one, and a low-amplitude in channel two. The  $F2$  of  $[\epsilon]$  is coded by a high amplitude in channel three, and a low amplitude in channels two and four. Now, consider the four-channel spectrum [Fig. 2(b)] of the vowel  $[\epsilon]$  produced by a female talker. In this case, four channels are not sufficient for coding  $F2$  information, since channel three is no longer a peak in the spectrum. Figure 2(c) and 2(d) shows the channel spectra of the same vowels processed through five channels. The  $F2$  information is coded adequately for both male and female vowels. The  $F2$  is coded by a high amplitude in channel four, and a low amplitude in channels three and five [see Fig. 2(c)]. Most generally, four-channel processors use two channels (channels three and four) for coding  $F2$  and the other high-frequency information needed for consonant recognition, while five-channel processors use three channels (channels three, four, and five). Overall, our results suggest that a minimum of three channels is needed to code  $F2$  and/or high-frequency information for multi-talker speech recognition.

It is possible that four channels might yield higher levels of speech understanding if the filter spacing were optimized. Shannon *et al.* (1998) showed that there was a significant difference in sentence recognition scores as a function of three filter spacings (linear, logarithmic, and intermediate) of

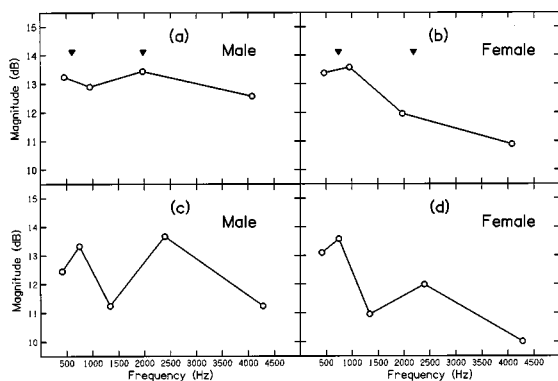


FIG. 2. The channel spectra of the vowel  $[\epsilon]$  ("head") produced by a male and a female talker. The spectra in (a) and (b) were generated using a four-channel processor, and the spectra in (c) and (d) were generated using a five-channel processor. The filled triangles indicate the formant frequencies of the vowels.

a four-channel processor. This filter optimization, however, can only be tailored for a particular speaker, e.g., a particular female or male, and is therefore not practical for real-world situations where multiple talkers must be accommodated.

Although five channels achieved high levels ( $>90\%$ ) of intelligibility, asymptotic performance was not achieved until eight channels were used. Increasing the number of channels beyond eight did not improve speech intelligibility, but did improve the subjective quality of speech. The finding that eight channels are needed to reach asymptotic performance is consistent with the study by Dorman *et al.* (1997) who showed that eight channels were needed to reach asymptote for multi-talker vowel recognition.

Training (i.e., practice) is a factor that needs to be taken into account when interpreting the above results, since the normal-hearing listeners were not accustomed to listening to speech containing limited spectral/temporal information. The order of the test conditions was purposely confounded with the amount of experience in listening to the altered speech signals because it was felt that giving listeners additional practice before encountering signals with the least spectral information would maximize performance in the most difficult listening situations.

## II. EXPERIMENT 2

Experiment 1 showed that a high level (90%) of intelligibility can be achieved using processors with five or more channels of stimulation. This finding is surprising given that the processors did not track or follow formant frequencies, like the pattern playback or the Remez *et al.* sine wave synthesizer. In the Remez *et al.* synthesizer, for instance, three sine waves trace out three formant frequencies in each update cycle. In contrast, the processors used in experiment 1 generated sine waves in each cycle (4 ms) at fixed frequencies (Table I). The only parameter that varied from cycle to cycle was the amplitudes of the sine waves. The frequencies of the sine waves coincided with the formant frequencies of speech only by chance and only rarely. This circumstance raises the question, "How is information coded in the frequency domain with processors that do not track formant frequencies?" As pointed out by Dorman *et al.* (1997), the relative differences in across-channel amplitudes must be used to code frequency information. On this view, if amplitude resolution were to be distorted, then speech recognition ought to decline. This hypothesis was tested in experiment 2 where the channel amplitudes of a six-channel processor were quantized to a finite number (2, 4, 8, 16) of steps. At issue was how many discriminable steps are needed to maintain high levels of speech intelligibility when speech is processed through a small number of channels. The answer to that question is of interest because it could provide some insight into whether the speech perception abilities of some cochlear implant patients are limited by electrode dynamic range or the number of discriminable intensity steps within the dynamic range (Nelson *et al.*, 1998).

It is reasonable to expect that the number of steps used to code amplitude information within a channel will be less important when speech is processed through a large number of channels than when processed through a small number of

channels. This is because in the case of a large number of channels, signal frequency will be indicated by the channel or channels with significant energy. To test this hypothesis we processed speech through 16 channels, and quantized the channel amplitudes into 2–16 steps. At issue was whether the same number of steps are needed to maintain high levels of speech intelligibility for speech processed through a large number (16) of channels and through a small number (6) of channels.

## A. Method

### 1. Subjects

The same subjects as in experiment 1 were used.

### 2. Sentence material

One hundred and fifty new sentences from the TIMIT database, produced by an equal number of female and male speakers, were randomly selected. Seventy-five sentences were used for the 6-channel processor and 75 sentences for the 16-channel processor. The 75 sentences used in each experiment were divided into five lists with 15 sentences in each list—one list was used for each of the four quantized conditions ( $Q=2,4,8,16$  levels), and one list was used for the unquantized condition. The subjects listened to a total of 150 sentences, 75 sentences processed through 6 channels, and 75 sentences processed through 16 channels.

### 3. Quantization and signal processing

The envelope dynamic range of speech processed through a finite number of channels differs from channel to channel. For that reason, different quantization step sizes are needed for each channel. We first determined the amplitude dynamic range of each channel by computing envelope histograms of 100 TIMIT sentences. [The TIMIT sentences were scaled so that all sentences had the same peak amplitude.] The maximum envelope amplitude in each channel, denoted as  $X_{\max}^i$  where  $i$  is the channel number, was chosen to include 99% of all amplitude counts in that channel. The minimum envelope amplitude ( $X_{\min}^i$ ) was set 0.5 dB above the rms value of the noise floor. The  $X_{\max}^i$  and  $X_{\min}^i$  values were then used to estimate the quantization step size,  $\Delta_i$ , of each channel as follows:

$$\Delta_i = \frac{X_{\max}^i - X_{\min}^i}{Q - 1} \quad i = 1, 2, \dots, N,$$

where  $Q$  is the number of quantization levels or steps, and  $N$  is the number of channels (6 or 16 in our case). Note that each channel had a different value for  $X_{\max}^i$  and  $X_{\min}^i$  since the envelope dynamic range of each channel was different. Consequently, the step sizes  $\Delta_i$  were different in each channel.

The quantized version of the six-channel sine wave processor was implemented as follows. Six envelope amplitudes were computed as before by pre-emphasizing the signal, bandpass filtering the signal into six logarithmic frequency bands (Table I), full-wave rectifying the bandpassed waveforms, and low-pass filtering (400 Hz) the rectified waveforms. The envelope amplitudes were then uniformly quan-

tized to  $Q$  discrete levels ( $Q=2,4,8,16$ ). Sine waves were generated with amplitudes equal to the quantized envelope amplitudes, and frequencies equal to the center frequencies of the bandpass filters. The phases of the sinusoids were estimated from the FFT of the speech segment (McAulay and Quatieri, 1986). The sinusoids of each band were finally summed and the level of the synthesized speech segment was adjusted to have the same rms value as the original speech segment.

The quantized version of the 16-channel sine wave processor was implemented as follows. Sixteen envelope amplitudes were computed as before by pre-emphasizing the signal, bandpass filtering the signal into 16 frequency bands (Table I), full-wave rectifying the bandpassed waveforms, and low-pass filtering (400 Hz) the rectified waveforms. Of the 16 envelopes computed, the six envelopes with the largest amplitude were selected in each 4-ms cycle.<sup>3</sup> The six selected envelope amplitudes were then uniformly quantized to  $Q$  discrete levels ( $Q=2,4,8,16$ ). Sine waves were generated with amplitudes equal to the quantized envelope amplitudes, and frequencies equal to the center frequencies of the selected bandpass filters. The phases of the sinusoids were estimated from the FFT of the speech segment. The sinusoids of the six selected bands were finally summed and the level of the synthesized speech segment was adjusted to have the same rms value as the original speech segment.

### 4. Procedure

The experiment was run in two independent 1½-h sessions. In the first session, the listeners were presented with a list of 75 sentences processed through the 6-channel processor, 60 quantized sentences (15 for each of the 4 conditions) and 15 unquantized sentences. In the second session, the listeners were presented with a list of 75 sentences processed through the 16-channel processor, 60 quantized sentences (15 for each of the 4 conditions) and 15 unquantized sentences. The quantized and the unquantized sentences, in both experiments, were completely randomized. A practice session preceded each test session, in which the listeners were presented with ten examples of sentences from each quantized condition. None of the sentences used in the practice session were used in the test session.

## B. Results and discussion

The results for the 6- and 16-channel processors are shown in Fig. 3. A repeated measures analysis of variance on the data for the six-channel processor indicated a main effect [ $F(4,32) = 112.54, p < 0.0001$ ] for the number of quantization steps. The mean scores were 41% correct for the 2-step condition, 52% correct for the 4-step condition, 80% correct for the 8-step condition, 83% correct for the 16-step condition, and 92% correct for the unquantized condition. *Post hoc* tests indicated that 4 steps allowed better performance than 2, 8 allowed better performance than 4 steps, 8 and 12 steps produced scores which did not differ, and the unquantized signal allowed better scores than the signal processed into 16 steps. Relatively high levels of intelligibility were achieved using 8 levels (mean score=80% correct) and 16



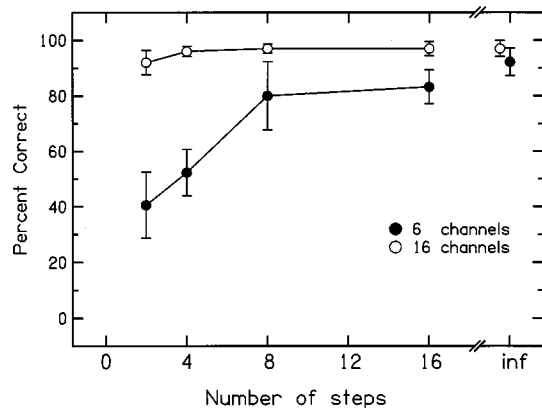


FIG. 3. Speech recognition with 6-channel (filled circles) and 16-channel (empty circles) processors as a function of the number of the steps used to quantize the spectral amplitudes. "Inf" refers to the condition in which the spectral amplitudes were not quantized. Error bars indicate  $\pm 1$  standard deviations.

levels (mean score=83% correct). These results are similar to the results found with early-model ten-channel vocoders (David, 1956), e.g., 82% correct with six levels of intensity quantization.

A repeated measures analysis of variance on the data for the 16-channel processor indicated a main effect [ $F(4,32) = 7.67, p < 0.0001$ ] for the number of quantization steps. *Post hoc* tests according to Scheffe showed that there was a significant difference ( $p = 0.002$ ) between the scores obtained with two and four steps, 92% correct and 96% correct, respectively. There was no statistically significant difference between the scores obtained with four steps and greater number of steps. Thus two steps were sufficient for achieving a high level (92%) of performance. This outcome is consistent with the findings of Drullman *et al.* (1995) that reported nearly perfect intelligibility when speech was processed through 24  $\frac{1}{4}$ -octave bands, and the amplitude envelopes were quantized into two levels. Our results and those of Drullman *et al.* (1995) suggest that poor amplitude resolution (defined in terms of the number of steps) does not have a large effect on intelligibility when speech is processed through a large number of channels.

In contrast, when speech was processed into a small number (six) of channels, performance was poor (<55% correct) when the number of levels was smaller than eight. This outcome can be accounted for by the view that that listeners must rely on relative amplitude differences across channels to infer frequency information when speech is processed into a small number of channels. If amplitude differences are distorted, then recognition accuracy will suffer. On this view, cochlear implant patients who are able to use only a few channels of stimulation, and who are able to discriminate only a small number of intensity differences on each channel (Nelson *et al.*, 1996), should find speech recognition relatively difficult.

### III. GENERAL DISCUSSION

#### A. Number of channels

The results in experiment 1 showed that five channels are needed to achieve high levels of sentence understanding

and eight channels are needed to reach asymptotic performance. The task at hand was recognition of TIMIT sentences produced by multiple speakers. It is very likely that the number of channels needed to reach asymptotic performance as well as the shape of the performance-channels function will depend on the speech material and whether listeners will be required to rely on phonetic detail. A different asymptote would be expected, for instance, if the task were nonsense-syllable recognition since the listeners will need to attend to fine acoustic/phonetic detail in order to understand what was being said. The results of experiment 1 do not support a general conclusion that eight channels are needed for all types of speech material, but rather for recognition of syntactically well-formed and meaningful sentences produced by multiple speakers.

Other factors that could affect the number of channels needed to achieve a high level of sentence intelligibility include speaking rate, speaking style (conversational versus clear) and background noise. Higher speaking rates are often associated with reduced sentence understanding, and speaking clearly is associated with improved sentence understanding in noise for normal-hearing listeners (Tolhurst, 1955) and improved sentence understanding in quiet for hearing impaired listeners (Picheny *et al.*, 1985). Both speaking style and speaking rate deserve further study in the context of the number of channels necessary for speech understanding. Speech understanding in noise has been studied by Dorman *et al.* (1998) and by Fu *et al.* (1998). More channels are needed in noise than in quiet to achieve high levels of speech understanding.

#### B. Number of steps and number of channels

The findings obtained in experiment 2 with the 6- and 16-channel-processors suggest an inverse relationship between the importance of spectral amplitude resolution and spectral resolution (defined in terms of the number of spectral channels available). Two levels of amplitude resolution were sufficient for nearly perfect intelligibility (92%) when speech was processed through 16 channels. However, eight or more levels were needed for high intelligibility when speech was processed through six channels. We have only investigated the effect of quantization on two extreme cases, i.e., a small number of channels and a large number of channels. Further studies are needed to complete our understanding of the effects of spectral amplitude resolution and spectral resolution on speech understanding.

### IV. CONCLUSIONS

These studies have provided yet another demonstration that speech understanding does not require a detailed spectral representation of the speech signal. In experiment 1 we found that five channels of fixed frequency stimulation allowed 90% identification accuracy for sentences produced by multiple speakers. Asymptotic performance was achieved with eight channels. In experiment 2 we found that the number of levels used to code spectral amplitude information has a significant effect on speech understanding. If speech is processed into a large number of channels, two levels of ampli-

tude resolution are sufficient to achieve a high level of speech understanding. However, when speech is processed into a small number of channels, eight or more levels are necessary. Thus the number of channels of stimulation and the resolution of amplitude information within those channels trade off in determining the level of speech understanding allowed by signal processors which reduce the speech signal to a relatively small number of fixed-frequency channels.

## ACKNOWLEDGMENTS

The authors would like to thank James Hillenbrand, Robert Shannon, and Steve Greenberg for providing valuable suggestions on earlier drafts of this paper. This research was supported by a Shannon award (R55 DC03421) from the National Institute of Deafness and other Communication Disorders, NIH.

<sup>1</sup>For  $n \geq 8$ , the filter bandwidths were computed according to the equation:  $1100 \log(f/800 + 1)$ , where  $f$  indicates the frequency in Hz. This is similar to the technical mel scale of Fant (1973), which is a variant of the critical band scale. As shown in Table II, the channel filter bandwidths, for  $n \geq 8$ , are approximately 1/4 of an octave wide, which is roughly the bandwidth of the critical band. Logarithmic spacing was used for  $n < 8$  to conform with the spacing used in current cochlear implant devices (e.g., Zierhofer *et al.*, 1994).

<sup>2</sup>The phases of the sinusoids were computed from the FFT of the speech segment as follows. Let  $\phi(k)$  be the phases of the FFT of a (4-ms) speech segment. The phases  $\theta(j)$  of the  $N$  sinusoids,  $N$  being the number of channels, were set equal to the phases of the FFT spectrum evaluated at frequencies closest to the center frequencies of the bandpass filters, i.e.,

$$\theta(j) = \phi\left(\left\lfloor \frac{f_j}{r} \right\rfloor\right), \quad j=1,2,\dots,N,$$

where  $f_j$  is the center frequency (Hz) of the  $j$ th bandpass filter (Table I),  $r$  is the FFT resolution ( $r$ =sampling frequency/FFT length) in Hz, and  $\lfloor \cdot \rfloor$  denotes the nearest integer. Due to the limited FFT resolution, the above equation only provides a rough estimate of the underlying sinewave phases. This estimate seems to be sufficient in our case, however, since we are only concerned with speech intelligibility rather than speech quality (see McAulay and Quatieri, 1995, for a discussion on alternative sinewave phase representations).

<sup>3</sup>This spectral-maximum implementation was chosen to mimic the signal processing used in the Nucleus 22 cochlear implant processor (McDermott *et al.*, 1992). In this processor, speech is processed through 16 channels, and the 6-channel amplitudes with the largest energy are selected in each cycle for electrical stimulation.

Bailey, P., Summerfield, Q., and Dorman, M. (1977). "On the identification of sine-wave analogues of certain speech sounds," Haskins Laboratories Status Report on Speech Perception, **SR 51-52**, 1-26.

Cutting, J. (1974). "Two left-hemisphere mechanisms in speech perception," *Percept. Psychophys.* **16**, 601-612.

David, E. (1956). "Naturalness and distortion in speech-processing devices," *J. Acoust. Soc. Am.* **28**, 586-589.

Delattre, F., Liberman, A., Cooper, F., and Gerstman, L. (1952). "An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic displays," *Word* **8**, 195-210.

Dorman, M., Loizou, P., and Rainey, D. (1997). "Speech intelligibility as a function of the number of channels of stimulation for signal processors

using sine-wave and noise-band outputs," *J. Acoust. Soc. Am.* **102**, 2403-2411.

Drullman, R. (1995). "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.* **97**, 585-592.

Dudley, H. (1939). "Remaking speech," *J. Acoust. Soc. Am.* **11**, 169-177.

Fant, G. (1973). *Speech Sounds And Features* (MIT Press, Boston).

Flanagan, J. (1972). *Speech Analysis, Synthesis And Perception* (Springer Verlag, New York).

Fu, Q.-J., Shannon, R., 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.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (1993). "DARPA TIMIT: Acoustic-phonetic continuous speech corpus," NIST Technical Report (distributed with the TIMIT CD-ROM).

Hill, F., McRae, L., and McClellan, R. (1968). "Speech recognition as a function of channel capacity in a discrete set of channels," *J. Acoust. Soc. Am.* **44**, 13-18.

Lamel, L., Kassel, R., and Seneff, S. (1986). "Speech database development: Design and analysis of the acoustic-phonetic corpus," Proc. of the DARPA Speech Recognition Workshop, Report No. SAIC-86/1546.

Loizou, P. (1998). "Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants," *IEEE Signal Process. Mag.* **15**, 101-130.

Loizou, P., Dorman, M., and Powell, V. (1998). "The recognition of vowels produced by men, women, boys and girls by cochlear implant patients using a six-channel CIS processor," *J. Acoust. Soc. Am.* **103**, 1141-1149.

McAulay, R., and Quatieri, T. (1986). "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-34**, 744-754.

McAulay, R., and Quatieri, T. (1995). "Sinusoidal coding," in *Speech Coding and Synthesis*, edited by W. Kleijn and K. Paliwal (Elsevier Science, New York).

McDermott, H., McKay, C., and Vandali, A. (1992). "A new portable sound processor for the University of Melbourne/Nucleus Limited multi-electrode cochlear implant," *J. Acoust. Soc. Am.* **91**, 3367-3371.

Mullenix, J., Pisoni, D., and Martin, C. (1989). "Some effects of talker variability on spoken word recognition," *J. Acoust. Soc. Am.* **85**, 365-378.

Nelson, D., Schmitz, J., Donaldson, G., Viemester, N., and Javel, E. (1996). "Intensity discrimination as a function of stimulus level with electric stimulation," *J. Acoust. Soc. Am.* **100**, 2393-2414.

Picheny, M., Durlach, N., and Braida, L. (1985). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Speech Hear. Res.* **28**, 96-103.

Remez, R., Rubin, P., Pisoni, D., and Carrell, T. (1981). "Speech perception without traditional cues," *Science* **212**, 947-950.

Schroeder, M. (1966). "Vocoders: Analysis and synthesis of speech," *Proc. IEEE* **54**, 720-734.

Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303-304.

Shannon, R., Zeng, F.-G., and Wygonski, J. (1998). "Speech recognition with altered spectral distribution of envelope cues," *J. Acoust. Soc. Am.* **104**, 2467-2476.

Sommers, M., Kirk, K., and Pisoni, D. (1997). "Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format," *Ear Hear.* **18**, 89-99.

Tolhurst, G. (1955). "The effect of intelligibility scores of specific instructions regarding talking," USAM Report No. NM 001 064 01 35 (Naval Air Station, Pensacola, FL).

Zierhofer, C., Peter, O., Bril, S., Pohl, P., Hochmair-Desoyer, I., and Hochmair, E. (1994). "A multichannel cochlear implant system for high-rate pulsatile stimulation strategies," in *Advances in Cochlear Implants*, edited by I. Hochmair Desoyer and E. Hochmair (International Interscience Seminars, Vienna), pp. 204-207.

# Subharmonic backscattering from ultrasound contrast agents

P. M. Shankar

*Department of Electrical and Computer Engineering, and School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, Pennsylvania 19104*

P. D. Krishna

*School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, Pennsylvania 19104*

V. L. Newhouse

*Department of Electrical and Computer Engineering, and School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, Pennsylvania 19104*

(Received 18 January 1999; accepted for publication 14 June 1999)

The ultrasonic contrast of blood in tissue, which is needed for ultrasonic estimation of tissue perfusion, can be increased by injecting the blood with bubbles or hollow microspheres. It has been shown that an even greater improvement in contrast can be obtained by using the *subharmonic* generated by irradiated microspheres. By obtaining analytical solutions to the modified RPNNP equation for a coated microbubble, the relationship between the physical parameters of the encapsulated bubble and the threshold pressure is established. The observed increase in the resonance frequency of a coated microsphere is explained by introducing the concept of "acoustic radius" of the encapsulated bubble. It is predicted that subharmonic generation in contrast agents requires a threshold insonifying pressure, and should be a minimum when microspheres are insonated at twice their resonance frequency. Experiments confirm the existence of this optimum incident frequency and of a reasonably low threshold pressure for the generation of the subharmonic. The existence of the low threshold pressures for subharmonic generation in contrast agents may prove to be very valuable in ultrasonic contrast imaging. © 1999 Acoustical Society of America. [S0001-4966(99)01510-6]

PACS numbers: 43.80.Vj [FD]

## INTRODUCTION

The study of contrast agents has attracted a great deal of attention recently, since their backscattered echo provides a significant enhancement of the signal-to-clutter ratio in flow measurements as well as in imaging.<sup>1-9</sup> In addition to the enhanced echo from contrast agents at the excitation frequency, it has been established that the second harmonic signal generated by the contrast agents provides a further improvement in the signal-to-clutter ratio.<sup>2-6</sup> The second harmonic signal at twice the excitation frequency shows a peak when the excitation frequency matches the resonance frequency of the bubble. This second harmonic behavior results from the nonlinearity of the bubble pulsations when subjected to external perturbation.

Since the nonlinear response of tissue may be weaker than the nonlinearity of bubbles or contrast agents, second harmonic imaging provides an excellent separation between the echoes from the blood vessels containing contrast agents and the tissue (vessel walls). These agents typically have a thin shell encapsulating air (or gas) bubbles, which may or may not have surfactant properties. The contrast agents manufactured by different manufacturers include Alunex®, Optison™ (Manufactured by Molecular Biosystems, Inc., San Diego, CA), Levovist® (Manufactured by Schering AG, Berlin, Germany), etc.<sup>6-9</sup> These agents consist of bubbles or microspheres, which typically have a range of diameters (up to about 12  $\mu\text{m}$ ), and have resonance frequencies in the range of about 1–5 MHz.

The second harmonic properties of contrast agents have been studied in detail by a number of researchers.<sup>2-12</sup> Specifically the relationships between different physical parameters of contrast agents and their backscattering cross sections at the second harmonic have been explored theoretically as well as experimentally to optimize second harmonic generation. Second harmonic imaging using contrast agents can, however, suffer from two major drawbacks; the second harmonic signal may undergo higher attenuation than the fundamental, and the tissue itself may produce a significant second harmonic signal, thereby reducing the contrast of the ultrasound echo from the bubbles or contrast agents.<sup>13-15</sup> These problems could be overcome if it were possible to use other harmonics generated by pulsating bubbles. In addition to the second harmonic at  $2f$ , bubbles insonified at frequency  $f$  have been shown to produce<sup>12</sup> higher order harmonics at  $3f$ ,  $4f$ ,  $5f$ , etc., ultraharmonics at  $(3/2)f$ ,  $(5/2)f$ , ..., etc., as well as subharmonics at  $f/2$ ,  $f/3$ , ..., etc. Of these harmonics, the subharmonics at  $f/2$  have been studied in detail in connection with acoustic cavitation.<sup>16</sup> Subharmonics typically have a threshold for generation, and this threshold is related to the physical parameters of the bubbles.<sup>16-18</sup> Some preliminary results<sup>13-15</sup> show that the subharmonic signal from contrast agents appears to have a better clutter-to-noise ratio than the second harmonic signal. In fact, the bubble subharmonic can even exceed the strength of the backscattered tissue echo at the fundamental,<sup>13,17</sup> thereby providing possibility for improvement in ultrasonic imaging.

Subharmonic signals at a frequency of  $(f/2)$  are typically generated by bubbles or contrast agents when insonified at a frequency of  $f$ . From extensive studies on free bubbles, it has been established that the subharmonic signal can be excited easily<sup>12,17</sup> if the incident frequency ( $f$ ) is equal to twice the resonance frequency of the bubble ( $f_0$ ). Subharmonic generation, however, requires that the applied pressure exceed a threshold. For free bubbles, this threshold pressure increases as the size of the bubble decreases. The threshold also increases with damping.<sup>12,17-21</sup> As far as the values of threshold pressures for subharmonic generation from contrast agents are concerned, there is some confusion as to the relationship between the observed threshold values and the threshold values expected from models currently available.<sup>21-24</sup> This confusion may arise from the notion that encapsulated bubbles must have a higher threshold than free bubbles of the same physical size and that threshold values for encapsulated bubbles will be equal to those for free bubbles if the shell encapsulating the bubble breaks under the applied pressure.<sup>24</sup> Many published results show, however, that the threshold pressures for encapsulated bubbles may even be lower than those expected for free bubbles.<sup>10,13-15,24</sup>

This apparent contradiction will be explored here. Expressions for the threshold pressure for subharmonic generation will be derived by solving the equation of motion for an encapsulated bubble. Experiments conducted on the contrast agent, *Optison*, will be correlated with the theoretical results.

## THEORY OF SUBHARMONIC GENERATION

Subharmonic generation by free bubbles has been analyzed by a number of researchers.<sup>12,17,19</sup> The starting point of these analyses is the well-known RPNNP (Rayleigh, Plesset, Noltingk, Neppiras, and Poritsky) equation of motion for bubbles. This equation for free bubbles can be modified to include the effects of encapsulation of the bubbles, as is the case for many contrast agents.<sup>8,9,22</sup> The modified RPNNP equation for a bubble having a shell can be expressed as<sup>8,9</sup>

$$\rho R R'' + \frac{3}{2} \rho (R')^2 = P_g \left( \frac{R_0}{R} \right)^{3\Gamma} + \left( p_v - p_{lo} - \frac{2\sigma}{R} \right) - 2S_p \left( \frac{1}{R_0} - \frac{1}{R} \right) - \delta_t \omega \rho R R' - p_{ac}(t), \quad (1)$$

where  $R$  is the instantaneous radius of the bubble,  $R_0$  is the radius at rest, and  $p_{ac}(t) = P \cos(\omega t)$  is the excitation signal at a frequency of  $\omega$  rad/s. Equation (1) includes the effect of the shell through the shell elasticity parameter  $S_p$  (N/m). The quantity  $\delta_t$  is the damping constant that includes contributions due to viscous damping, thermal damping, radiation damping, and damping due to internal friction in the shell.<sup>9,22,25-27</sup> The other terms in Eq. (1) are given in the Appendix. The solution to Eq. (1) will lead to the conditions necessary for the existence of the subharmonic component.

The equation of motion for bubbles given in Eq. (1) can be solved numerically.<sup>12-14</sup> As mentioned earlier, these results indicate that the subharmonic strength is large when the

excitation frequency is twice the resonance frequency of the bubble. In order to understand the existence of a threshold pressure, and the relationship between the threshold and the bubble parameters, an analytical solution will be more appropriate than a numerical solution. Toward this end, one can follow the approach of Prosperetti<sup>19-21</sup> and assume a solution of the form

$$R = R_0 x, \quad (2)$$

where  $x$  is given by

$$x = 1 + A_1 \cos(\omega t + \phi_1) + A_2 \cos(0.5\omega t + \phi_2) + \dots \quad (3)$$

The quantities  $A_1$  and  $A_2$  are the amplitudes of the fundamental ( $\omega = 2\pi f$ ) and of the subharmonic ( $\omega/2$ ) component, respectively. Quantities  $\phi_1$  and  $\phi_2$  are the phases associated with fundamental and subharmonic signals, respectively. Substituting Eqs. (2) and (3) in Eq. (1), the solutions for  $A_1$ ,  $A_2$ , and the resonance frequency ( $\omega_s = 2\pi f_s$ ) of the encapsulated bubble can be obtained. Treating  $A_2$  as zero and making a small amplitude approximation, the amplitude at the fundamental,  $A_1$ , can be expressed as

$$A_1 = \frac{P}{(\rho R_0^2) \sqrt{(S\omega)^2 + [\omega^2 + \chi - LG]^2}}, \quad (4)$$

where  $S$ ,  $\chi$ ,  $L$ , and  $G$  are given by

$$S = \delta_t \omega, \quad (5a)$$

$$\chi = \frac{2(\sigma - S_p)}{\rho R_0^3}, \quad (5b)$$

$$L = \frac{P_g}{\rho R_0^2}, \quad (5c)$$

$$G = 3\Gamma. \quad (5d)$$

Neglecting the damping term  $S$ , Eq. (5a), the resonance frequency,  $f_s$ , of the encapsulated bubble can be expressed as

$$f_s = \left( \frac{1}{2\pi} \right) \sqrt{\frac{3\Gamma P_g}{\rho R_0^2} - \frac{2(\sigma - S_p)}{\rho R_0^3}}. \quad (6)$$

Thus the resonance frequency,  $f_s$ , of the encapsulated bubbles will be *higher* than the resonance frequency  $f_0$  of a free bubble<sup>8,22</sup> of identical physical dimensions, since typical values of  $S_p$  are larger<sup>8,9</sup> than  $\sigma$ . The resonance frequency  $f_0$  of a free bubble is given by

$$f_0 = \left( \frac{1}{2\pi} \right) \sqrt{\frac{3\Gamma P_g}{\rho R_0^2} - \frac{2\sigma}{\rho R_0^3}}. \quad (7)$$

An encapsulated bubble can thus be said to have an ‘‘acoustic radius’’ smaller than its physical radius,<sup>22</sup> where the acoustic radius is defined as the radius of a free bubble having the resonance frequency,  $f_s$ . The acoustic radius,  $R_{ac}$ , is given by

$$R_{ac} = R_0 \sqrt{\frac{\frac{3\Gamma P_g}{\rho R_0^2} - \frac{2\sigma}{\rho R_0^3}}{\frac{3\Gamma P_g}{\rho R_0^2} - \frac{2(\sigma - S_p)}{\rho R_0^3}}}. \quad (8)$$



Note that the acoustic radius,  $R_{ac}$ , and the physical radius,  $R_0$ , of the bubble become equal if the shell is nonexistent. Equation (8) is approximate due to the cubic dependence of  $R_0$ .

Knowing  $A_1$ , the amplitude of the component at the fundamental frequency, it is possible to write an expression for the amplitude of the subharmonic,  $A_2$ ,

$$A_2^2 = \left(\frac{1}{F}\right) \left\{ -B - 2FA_1^2 \pm \sqrt{A_1^2 \left( D^2 + \frac{S^2 \omega^2}{16} \right) - \frac{S^2 \omega^2}{4}} \right\}, \quad (9)$$

where  $B$ ,  $F$ , and  $D$  are given by

$$B = \frac{\omega^2}{4} - LG + \chi, \quad (10a)$$

$$F = \left(\frac{3}{4}\right) \left[ \chi - \left(\frac{1}{6}\right) LG(G+1)(G+2) \right] \quad (10b)$$

$$D = \frac{\omega^2}{8} + \chi - \left(\frac{1}{2}\right) LG(G+1). \quad (10c)$$

For the subharmonic to exist, the quantity in the square root in Eq. (9) must be positive. This leads to the frequency dependent threshold pressure,  $P_t$ , for the generation of subharmonic<sup>17,19</sup> given by

$$P_t = \frac{\delta_t \omega^2 \rho R_0^2}{2} \sqrt{\left\{ \frac{(S\omega)^2 + [\omega^2 - \omega_s^2]}{\sqrt{\left[ D^2 + \left(\frac{S\omega}{4}\right)^2 \right]}} \right\}}. \quad (11)$$

It is clear from Eq. (9) that the threshold pressure increases with damping. The instability threshold pressure,<sup>19</sup>  $P_{ins}$ , for the subharmonic generation results when  $A_2$  is greater than zero in Eq. (9), with the negative sign outside the square root. Once again solving for  $P_{ins}$ ,

$$P_{ins} = \rho R_0^2 Z \sqrt{(S\omega)^2 + [\omega^2 - \omega_s^2]^2}, \quad (12)$$

where  $Z$  is given by

$$Z = \sqrt{\frac{D^2 + \left(\frac{S^2 \omega^2}{16}\right) - 4FB - Y}{8F^2}}. \quad (13)$$

The quantity  $Y$  is given by

$$Y = \sqrt{D^4 - D^2 \left( 8FB - \frac{S^2 \omega^2}{8} \right) - \frac{FBS^2 \omega^2}{2} + \frac{S^4 \omega^4}{256} - 4F^2 S^2 \omega^2}. \quad (14)$$

Note that in the absence of a shell, Eqs. (11) and (12) reduce to the corresponding equations derived by Prosperetti.<sup>19</sup> The various attributes of these two thresholds have been discussed in detail in a series of papers by Prosperetti,<sup>19-21</sup> and will not be discussed here. It suffices to mention that the condition that the applied pressure must exceed the frequency dependent threshold  $P_t$  of Eq. (11) is necessary but not sufficient. If the applied pressure is between  $P_t$  and the instability pressure  $P_{ins}$  defined in Eq. (12), subharmonics may or may not be generated depending on the type of excitation involved.

Instead of calculating the damping constants,  $\delta_t$ , and then computing the thresholds, the threshold pressures can be

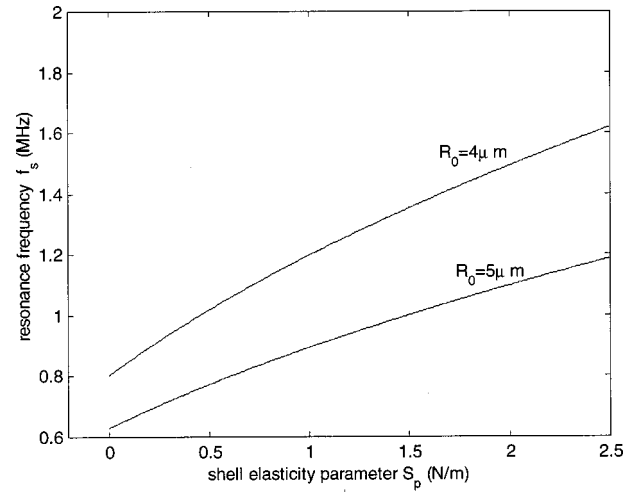


FIG. 1. Resonance frequency as a function of the shell elasticity parameter,  $S_p$  for two values of radii ( $4 \mu\text{m}$  and  $5 \mu\text{m}$ ).

calculated for various values of  $\delta_t$  and shell elasticity parameter,  $S_p$ . Most of the currently available values of  $\delta_t$  are based on simple models<sup>25-27</sup> and do not take into account the nonlinear effects of the presence of the shell as well as effects of having an organic compound such as a surface coating on the shell. This may affect the surface tension which in turn can modify the viscous as well as thermal damping coefficients.<sup>23,26,28-30</sup> For these reasons, the damping constant will not be calculated here. Damping constant may, however, be extracted from the analysis of experimental as well as theoretical results. These aspects will be discussed later.

The resonance frequency given in Eq. (6) was computed and is shown in Fig. 1. Two values of the radii ( $R_0 = 4 \mu\text{m}$  and  $5 \mu\text{m}$ ) of the contrast agent were used in all the computations. These sizes were within the size range of the contrast agent, *Optison*, used in our experiment.<sup>6,13,14</sup> The physical constants used in the computation are given in the Appendix. The effect of the shell is seen to increase the resonance frequency, thereby reducing the ‘‘acoustic radius’’ of the bubble. Thus the encapsulated bubble will have a resonance frequency higher than a free bubble of the same physical dimensions. Note that frequency calculations were done by ignoring the damping.

Threshold pressures calculated for different values of the damping constant,  $\delta_t$ , and for a set of values of the shell elasticity parameter  $S_p$ , are shown in Fig. 2 ( $R_0 = 4 \mu\text{m}$ ) and Fig. 3 ( $R_0 = 5 \mu\text{m}$ ). While the pressure values (the frequency dependent threshold,  $P_t$ , and the instability threshold,  $P_{ins}$ ) are normalized to the atmospheric pressure (100 kPa), and the frequencies are normalized to the resonance frequency of the encapsulated bubble ( $f_s$ ). It is clear that a minimum threshold exists around an excitation frequency of about  $2f_s$ . This threshold value rises slowly with the value of  $S_p$ ; it rises quickly, however, if the damping,  $\delta_t$ , increases. Threshold pressures reach about 2–3 atmospheres for  $S_p = 1$  and  $\delta_t = 0.05$ , demonstrating the steep rise in threshold pressures with damping. These results are similar to those reported by Prosperetti.<sup>19-21</sup>

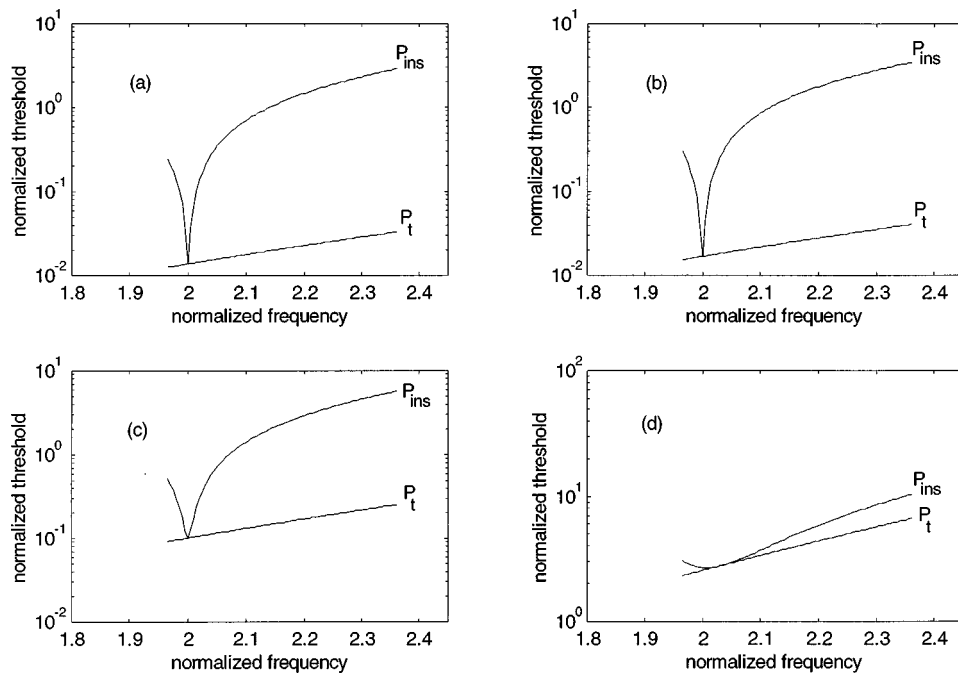


FIG. 2. Threshold pressures (frequency dependent threshold,  $P_t$  and instability threshold,  $P_{ins}$ ), for subharmonic excitation (normalized to the ambient pressure of 100 kPa) for an encapsulated bubble versus excitation frequency (normalized to the resonance frequency  $f_s$ ).  $R_0 = 4 \mu\text{m}$ ; (a)  $S_p = 0.0$ ;  $\delta_i = 0.001$ ;  $f_s = 800 \text{ KHz}$ . (b)  $S_p = 0.1$ ;  $\delta_i = 0.001$ ;  $f_s = 850 \text{ KHz}$ . (c)  $S_p = 0.45$ ;  $\delta_i = 0.0035$ ;  $f_s = 1 \text{ MHz}$ . (d)  $S_p = 1.0$ ;  $\delta_i = 0.05$ ;  $f_s = 1.2 \text{ MHz}$ .

## EXPERIMENTAL PROCEDURE AND RESULTS

Experiments using the contrast agent *Optison* were conducted in our laboratory to explore some of the theoretical aspects described in the previous section. The resonance frequency<sup>13,14</sup> of the batch of *Optison* used was approximately 1 MHz. About 50  $\mu\text{l}$  of this material was drawn into a Wheaton Socorex 50–200  $\mu\text{l}$  variable micropipette. This was mixed with 300 ml of saline. The container was gently agitated to ensure uniform mixing. This solution was then poured into the test tank (shown in Fig. 4), and readings taken. Fresh solution was prepared for each measurement set. The solution was gently stirred between individual measurements to ensure proper mixing.

The electronic setup (Fig. 4) consisted of a Hewlett-

Packard 3326A two-channel synthesizer, used as a signal generator, feeding an ENI A150 power amplifier which drove the transmit transducer (unfocused)—a 2.25-MHz center frequency, 0.25-in. diameter, Panametrics v323 type unit, with 63.1% bandwidth at  $-6 \text{ dB}$ . The transmit and receive transducers were fitted on to a specially designed test tank (Fig. 4). The receive transducer (unfocused)—a 1-MHz, 0.5-in., Panametrics v303, with 61.2% bandwidth at  $-6 \text{ dB}$ , was connected to a Hewlett-Packard type 3585A spectrum analyzer. The transducers were calibrated using a Force instruments needle type PVDF hydrophone. All components of the setup were first individually tested to rule out nonlinearities at the signal levels being used. When making measurements, the amplitude at the frequency of interest was determined

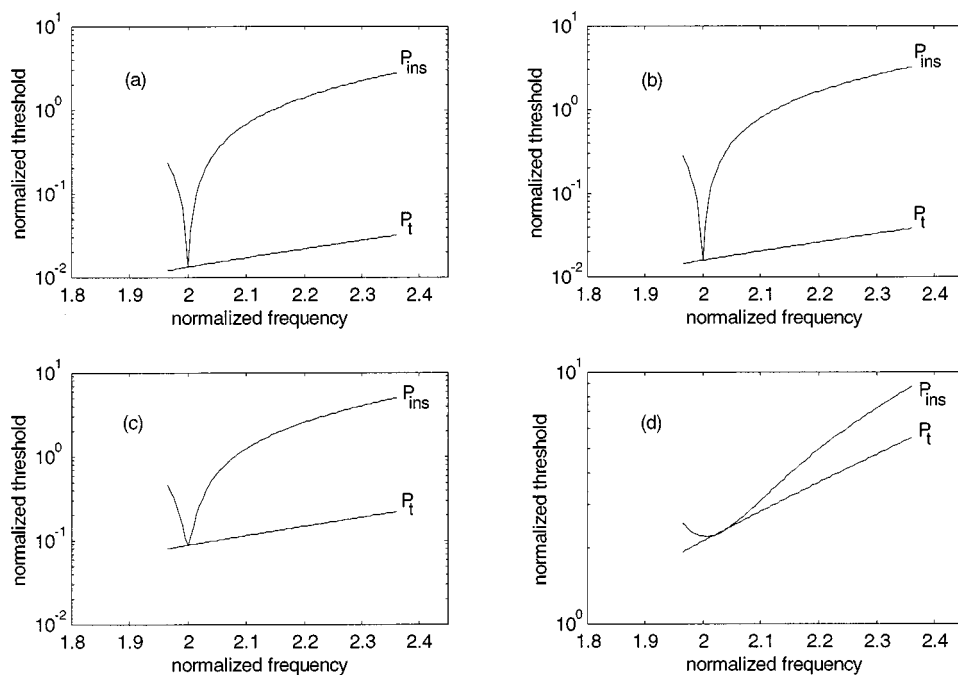


FIG. 3. Threshold pressures (frequency dependent threshold,  $P_t$  and instability threshold,  $P_{ins}$ ), for subharmonic excitation (normalized to the ambient pressure of 100 kPa) for an encapsulated bubble versus excitation frequency (normalized to the resonance frequency  $f_s$ ).  $R_0 = 5 \mu\text{m}$ ; (a)  $S_p = 0.0$ ;  $\delta_i = 0.001$ ;  $f_s = 628 \text{ KHz}$ . (b)  $S_p = 0.1$ ;  $\delta_i = 0.001$ ;  $f_s = 659 \text{ KHz}$ . (c)  $S_p = 0.45$ ;  $\delta_i = 0.0035$ ;  $f_s = 773 \text{ KHz}$ . (d)  $S_p = 1.0$ ;  $\delta_i = 0.05$ ;  $f_s = 895 \text{ KHz}$ .

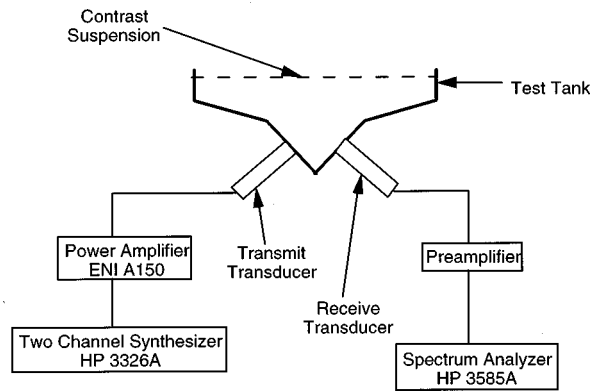


FIG. 4. The tank setup with the measurement system.

from the spectrum analyzer, with its “resolution bandwidth” set at 3 Hz and its “video bandwidth” set at 1 Hz.

The transducers were held at 90° to each other with their beams intersecting at the center of the test tank. The design of the testing tank was such that it did not allow the transmitted signal to reach the receive transducer in the absence of a scatterer in the contained solution. Hence a linear reflector (stainless steel plate) was suspended in the tank within the volume of interest (about 2.25 cm from the transducer end face) to measure transmission nonlinearities in saline. The measured signal (at the fundamental) with this linear reflector was approximately 100 dB above the signal level in the absence of the bubbles or contrast agent. To normalize for transmission nonlinearities as well as the transducer transfer functions, the ratio of the backscattered signal from the contrast with respect to the backscattered signal from a linear reflector (suspended in saline) was obtained.

Experiments were conducted to measure the subharmonic signal as a function of the excitation frequency. (All the data points were averaged from at least eight separate measurements except for the data for saline, in which case averaging was performed on four separate measurements.) The aim here was to establish the existence of a maximum at an excitation of twice the resonance frequency of the contrast agent. The transmit transducer was calibrated as described earlier such that a constant pressure of 10 kPa was obtained at the location of the linear reflector at the various insonation frequencies. The contrast suspension was insonated at frequencies ranging from 1.95 MHz to 3.0 MHz at a constant pressure of 10 kPa (Fig. 5) in the continuous wave (CW) mode. A dilution of 1 in 2500 was used. The scattered signal at the subharmonic shows a distinct peak (2.00 MHz) twice the resonance frequency of the contrast agent. The subharmonic signal is seen to be weaker on either side of this frequency.

Experiments were also conducted to measure the subharmonic signal as a function of the applied pressure (Fig. 6). The existence of a threshold should be observable from these measurements. The contrast suspension was insonated with 2.00-MHz continuous wave (CW) at various acoustic pressures and the backscattered subharmonic measured. A dilution of 1 in 6000 was used. As the pressure is increased to about 10 kPa, the strength of the subharmonic is seen to go up sharply. The onset of the subharmonic thus takes place at

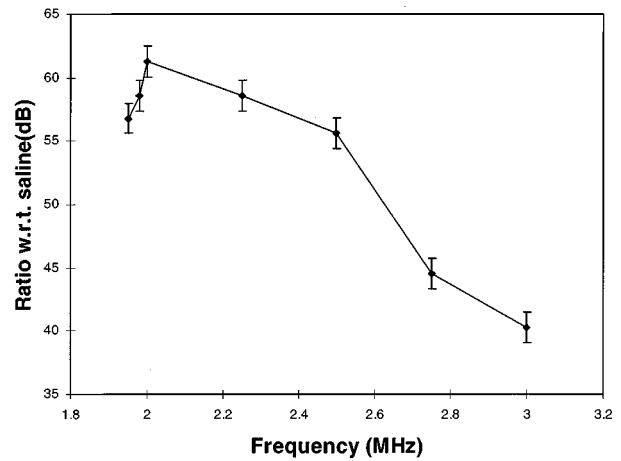


FIG. 5. The measured subharmonic (normalized w.r.t. saline) for different excitation frequencies (acoustic pressure=10 kPa). A distinct peak at twice the resonance frequency of the contrast agent is seen.

a very low “threshold pressure,” in agreement with results reported by other research groups.<sup>3,24</sup>

## DISCUSSION

Expressions derived in this paper showed that the resonance frequency of encapsulated bubbles increases with the shell elasticity parameter. It is possible to characterize this effect in terms of an “acoustic radius” of the encapsulated bubble,  $R_{ac}$ , which will be less than the physical radius,  $R_0$ , of the free bubble. In addition, the relationships among bubble size, shell elasticity parameter, and damping were investigated theoretically, and experiments were conducted to demonstrate the existence of a minimum threshold at approximately twice the resonance frequency of the bubble.

It is clear from our experimental results that subharmonics are generated at much lower pressures than would be expected on the basis of the predicted damping constants for free as well as encapsulated bubbles.<sup>22,25,26</sup> Similar results have also been observed by other research groups.<sup>3,24</sup> If these damping constants (typically ~0.1 for the size of bubbles investigated here) were used in the computation of threshold pressures, much higher pressures than those found experi-

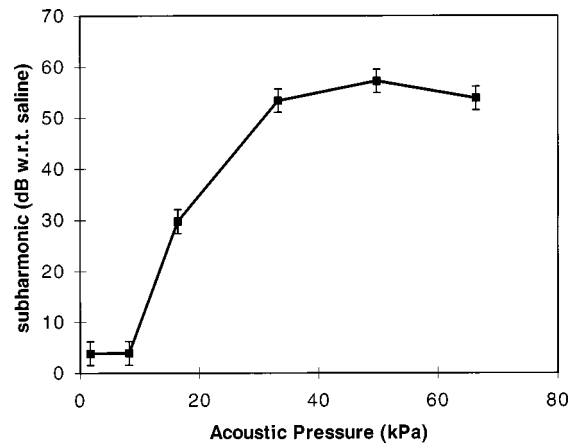


FIG. 6. The measured subharmonic (normalized w.r.t. saline) for different excitation pressures (insonating frequency=2.00 MHz). Threshold pressure of 10 kPa is seen.

mentally here would be needed to observe subharmonics.<sup>17,19</sup> (The predicted damping constants<sup>22,25,26</sup> were obtained on the basis of both linear and nonlinear models; however, it is possible that not all chemical and physical effects of coatings were taken into account.)

One of the possible explanations for the lowering of the thresholds demonstrated above, was provided by Lotsberg *et al.*,<sup>24</sup> who suggested that the shell of contrast agent microspheres could break up as a result of which they would then behave like free bubbles. If one examines the values used by Lotsberg *et al.*,<sup>24</sup> threshold pressures should be higher than 50 kPa, even for free bubbles in the range of 4  $\mu\text{m}$  while the reported pressures were about 36 kPa or less. Thus the shell breakage explanation may not in fact provide the low thresholds seen. The possible explanation may lie elsewhere. The fact that encapsulated bubbles are coated with organic materials may have an impact on the damping constants. The damping constants for encapsulated bubbles are calculated on the basis of the shell being "elastic solid."<sup>18,22</sup> Since the thickness of the shell is very small and since shell materials are not "hard," it is possible that thermal and viscous damping factors are reduced due to the presence of the soft shells. In addition, organic coatings may provide some surfactant properties, again affecting the damping factors. Taking these effects into account, studies<sup>23,29,30</sup> have shown that the damping constants for encapsulated bubbles at high frequencies can be lower than those of free bubbles. Another possibility is that the various damping factors are not independent and any correlation between various damping mechanisms can lead to reduction of the total damping factor. This could explain the existence of the lower thresholds observed for encapsulated bubbles. These aspects were explored first by computing the damping constants from the measured threshold pressure and then correlating with the theoretical results.

For a radius of 4  $\mu\text{m}$  and a resonance frequency of 1 MHz, the value of  $S_p$  for *Optison* was calculated to be equal to 0.452 from Eq. (6). Using a threshold value of 10 kPa, the damping factor  $\delta_t$  was calculated to be equal to 0.0035 from Eq. (11). This corresponds to the theoretical curves in Fig. 2(c). It therefore appears that the results reported by other researchers and the results reported here in terms of very low threshold pressures can be explained in terms of actual damping factors being much smaller than those predicted on the basis of linear models for damping factors. In fact, it was suggested by Prosperetti<sup>19</sup> that subharmonic generation *might* be a better way to experimentally measure damping constants due to the direct relationship existing between damping constants and thresholds. The results obtained here *strongly* support this concept.

The experimental and theoretical results on thresholds presented above also demonstrate potential applications of subharmonic imaging. Since tissue produces weak subharmonics,<sup>13</sup> it is possible to increase the contrast of the echo from blood carrying encapsulated bubbles with respect to tissue, through the use of subharmonic signals. By insonifying at twice the resonance frequency of the bubbles, the pressures required to generate subharmonics are in the range of 10–20 kPa which are reasonably low from the standpoint of clinical usage. (Note, however, that in actual experiments

*in vivo*, additional damping provided by the blood and tissue is likely to increase the threshold pressures.) Since subharmonic imaging will be done at the bubble resonance frequency, the resolution of the imaging system will be no worse than what one would expect for insonation at the resonance frequency.

## ACKNOWLEDGMENTS

The authors are grateful to Professor Peter Lewin for his assistance in calibrating the hydrophone. The authors are indebted to Molecular Biosystems, Inc. for the provision of contrast agents. The authors also wish to thank the anonymous reviewers for their valuable suggestions. This work is supported by NIH-NCI Grant No. CA 52823.

## APPENDIX

Terms used in Eq. (1) along with the pertinent values for the analytical solutions:

- $\rho$  = density of surrounding medium (998 Kg/m<sup>3</sup>)
- $\Gamma$  = polytropic exponent of gas (1.09 for C<sub>3</sub>F<sub>8</sub>, the gas in Optison)
- $p_v$  = vapor pressure ( $2.33 \times 10^3$  N/m<sup>2</sup>)
- $p_{lo}$  = hydrostatic pressure ( $1.01 \times 10^5$  N/m<sup>2</sup>)
- $\sigma$  = surface tension of water (0.072 N/m)
- $P_g$  = initial gas pressure inside the bubble ( $(2\sigma/R0) + p_{lo} - p_v$ )

- <sup>1</sup>R. Gramiak and P. M. Shah, *Invest. Radiol.* **3**, 356–366 (1968).
- <sup>2</sup>B. Schrope, V. L. Newhouse, and V. Uhlendorf, *Ultrason. Imaging* **14**, 134–158 (1992).
- <sup>3</sup>P. H. Chang, K. K. Shung, and H. B. Levene, *Ultrasound Med. Biol.* **22**, 1205–1214 (1996).
- <sup>4</sup>F. Forsberg, B. B. Goldberg, J. B. Liu, D. A. Merton, and N. M. Rawool, *J. Ultrasound Med.* **15**, 853–860 (1996).
- <sup>5</sup>F. Forsberg, B. B. Goldberg, Y. Q. Wu, J. B. Liu, D. A. Merton, and N. M. Rawool, *Int. J. Imaging Syst. Technol.* **8**, 69–81 (1997).
- <sup>6</sup>P. D. Krishna and V. L. Newhouse, *Ultrasound Med. Biol.* **23**, 453–459 (1997).
- <sup>7</sup>V. Uhlendorf and C. Hoffmann, "Nonlinear acoustical response of coated microbubbles in diagnostic ultrasound," *Proc. IEEE Ultrason. Symp.*, pp. 1559–1562 (1994).
- <sup>8</sup>N. de Jong, "Acoustic properties of ultrasound contrast agents," Ph.D. dissertation, Erasmus University, Rotterdam, 1993.
- <sup>9</sup>N. de Jong, L. Hoff, T. Skotland, and N. Bom, *Ultrasonics* **30**, 95–103 (1992).
- <sup>10</sup>P. H. Chang, K. K. Shung, S. Wu, and H. B. Levene, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **42**, 1020–1026 (1995).
- <sup>11</sup>D. L. Miller, *Ultrasonics* **19**, 217–224 (1981).
- <sup>12</sup>W. Lauterborn, *J. Acoust. Soc. Am.* **59**, 283–293 (1976).
- <sup>13</sup>P. M. Shankar, P. D. Krishna, and V. L. Newhouse, *Ultrasound Med. Biol.* **24**, 395–399 (1998).
- <sup>14</sup>P. D. Krishna, P. M. Shankar, and V. L. Newhouse, *Phys. Med. Biol.* **44**, 681–694 (1999).
- <sup>15</sup>W. T. Shi, F. Forsberg, and B. B. Goldberg, (abstract) *J. Acoust. Soc. Am.* **101**, 3139 (1997).
- <sup>16</sup>W. T. Coakley and L. N. Nyborg, Cavitation: Dynamics of gas bubbles; Applications, in *Ultrasound: Its Applications in Medicine and Biology: Part One*, edited by F. J. Frey (Elsevier Scientific, New York, 1978), Chap. II, pp. 77–159.
- <sup>17</sup>A. Eller and H. G. Flynn, *J. Acoust. Soc. Am.* **46**, 722–727 (1969).
- <sup>18</sup>E. A. Neppiras, *J. Acoust. Soc. Am.* **46**, 587–601 (1969).
- <sup>19</sup>A. Prosperetti, *J. Acoust. Soc. Am.* **61**, 11–16 (1977).
- <sup>20</sup>A. Prosperetti, *J. Acoust. Soc. Am.* **61**, 17–27 (1977).
- <sup>21</sup>A. Prosperetti, *J. Acoust. Soc. Am.* **56**, 878–885 (1971).
- <sup>22</sup>C. C. Church, *J. Acoust. Soc. Am.* **97**, 1510–1521 (1995).
- <sup>23</sup>T. G. Leighton, *The Acoustic Bubble* (Harcourt Brace, San Diego, 1994), pp. 181–191 and 370–379.



- <sup>24</sup>O. Lotsberg, J. M. Hovem, and B. Aksum, *J. Acoust. Soc. Am.* **99**, 1366–1369 (1996).
- <sup>25</sup>C. Devin, *J. Acoust. Soc. Am.* **31**, 1654–1667 (1959).
- <sup>26</sup>B. C. Eatock, R. Y. Shi, and G. W. Johnston, *J. Acoust. Soc. Am.* **77**, 1692–1701 (1985).
- <sup>27</sup>A. Eller, *J. Acoust. Soc. Am.* **47**, 1469–1470 (1970).
- <sup>28</sup>C. C. Church, “On the interactions between sound waves and microbubbles,” *The Third Annual International Symposium on Contrast Agents in Diagnostic Ultrasound*, Atlantic City, New Jersey, May 13, 1997.
- <sup>29</sup>R. E. Glazman, *J. Acoust. Soc. Am.* **74**, 980–985 (1983).
- <sup>30</sup>R. E. Glazman, *J. Acoust. Soc. Am.* **76**, 890–896 (1984).

**Erratum: “Monosyllabic word recognition at higher-than-normal speech and noise levels” [J. Acoust. Soc. Am. 105, 2431–2444 (1999)]**

Gerald A. Studebaker and Robert L. Sherbecoe

*Memphis Speech and Hearing Center, The University of Memphis, 807 Jefferson Avenue, Memphis, Tennessee 38105*

D. Michael McDaniel

*Communication Disorders, Arkansas State University, P.O. Box 910, State University, Arkansas 72467*

Catherine Gwaltney

*Memphis Speech and Hearing Center, The University of Memphis, 807 Jefferson Avenue, Memphis, Tennessee 38105*

(Received 8 June 1999; accepted for publication 24 June 1999)

[S0001-4966(99)02110-4]

PACS numbers: 43.66.Ba, 43.66.Sr, 43.66.Ts, 99.10.+g [SPB]

Table VII incorrectly reported that the data of four studies showed slopes (dB/dB) of 0. The correct values in these four cases should have been 1.0. The corrected table is reproduced below.

TABLE VII. The change in S/N ratio needed to maintain a constant speech-recognition score, based on data from 12 studies. Also reported are the speech and noise types used in each study. The slope values indicate the amount of masking, in dB, per dB of applied noise.

Investigation	Speech	Noise	Slope (dB/dB)
Kryter (1946)	PAL PB words	Engine spectrum	1.12
French and Steinberg (1947)	Nonsense syllables	Quiet	1.15 to 1.4
Hawkins and Stevens (1950)	Continuous discourse	White	1.0
Hirsh and Bowman (1953)	Spondees	White	1.0
Hirsh <i>et al.</i> (1954)	Various	White	1.0
Pickett and Pollack (1958)	PAL PB words	−12 dB/oct	1.2
Dirks <i>et al.</i> (1982)	NU#6 words	12-voice babble	1.14 to 1.17
Duquesnoy and Plomp (1983)	Dutch sentences	Talker spectrum	1.0
Dubno <i>et al.</i> (1984)	SPIN-PL	12-voice babble	1.11
Stelmachowicz <i>et al.</i> (1985)	NU#6 words	Broadband	1.18
Walker and Byrne (1985)	Connected discourse	Talker spectrum	1.05
Current study	NU#6 words	Talker spectrum	1.34

## PROGRAM OF

### The 138th Meeting of the Acoustical Society of America Hyatt Regency Columbus • Columbus, Ohio • 1–5 November 1999

1a MON. AM

**NOTE:** All Journal articles and Letters to the Editor are peer reviewed before publication. Program abstracts, however, are not reviewed before publication, since we are prohibited by time and schedule.

MONDAY MORNING, 1 NOVEMBER 1999

MARION ROOM, 8:10 TO 11:15 A.M.

#### Session 1aAO

#### Acoustical Oceanography: Surf Zone Processes and Internal Wave Oceanography

James C. Preisig, Cochair

*Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution,  
Woods Hole, Massachusetts 02543*

Grant B. Deane, Cochair

*Marine Physical Laboratory, Scripps Institute of Oceanography, University of California, San Diego,  
La Jolla, California 92093-0238*

Chair's Introduction—8:10

#### Contributed Papers

8:15

**1aAO1. Introduction to the NSANE experiment.** Dale Green (Datasonics, Inc., P.O. Box 8, Cataumet, MA 02534, dgreen@datasonics.com)

The NearShore Acoustic Network Experiment (NSANE) surf zone characterization experiment recently conducted on behalf of ONR at the Scripps pier in San Diego was a major logistical enterprise, as well as a challenging scientific and engineering exercise. The goal of the experiment was to ascertain the feasibility of acoustic communications and navigation in this adverse environment. Test equipment was provided by a consortium of organizations, all of which required coordinated installation over a full week, occasionally under rather adverse conditions. As a prelude to several other presentations on technical and scientific results, a video recording of much of the installation and demobilization will be shown highlighting the NSANE nature of the experiment.

8:30

**1aAO2. Signaling characteristics of the surf zone acoustic communications channel.** James C. Preisig, James W. Partan, and Mark P. Johnson (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The Near Shore Acoustic Network Experiment (NSANE) was conducted at the Scripps Pier in May 1999 to ascertain the physical properties of the very shallow water and surf zone environments which enable or constrain acoustic communications. Results of the analysis of the experimental data will be presented focusing on the properties of the environment which are of particular relevance to the performance of acoustic communications systems. Communications channel characteristics such as

multipath spread, coherence time, and channel outage and signal attenuation characteristics will be examined. These characteristics will be related to observable environmental conditions and their effect on the performance of acoustic communications systems will be discussed.

8:45

**1aAO3. Bubble distributions and high-frequency propagation in the surf zone.** David M. Farmer and Svein Vagle (Inst. of Ocean Sci., P.O. Box 6000, Sidney, BC V8L 4B2, Canada, farmerd@dfm-mpo.gc.ca)

The masking and scattering effects of dense bubble populations are a distinguishing feature of high-frequency acoustic propagation in the surf zone. The high-quality factor of bubbles at resonance ensures that at higher frequencies bubble size distribution is a crucial factor determining propagation characteristics. In May 1999 a collaborative study near Scripps Pier led to the simultaneous acquisition of bubble size distributions within the surf, together with acoustic back and forward scatter at frequencies of 12 and 100 kHz. Vertical sonars were also used to monitor the development of bubble clouds beneath breaking waves, while horizontal imaging and Doppler sonars tracked the evolution of the larger scale bubble field. Bubble field and surface wave measurements provide the environmental background that must be incorporated in realistic propagation models. Bubble evolution models must in turn account for the currents and wave-induced turbulence, along with the loss of larger bubbles through buoyancy and effects of gas dissolution on smaller bubbles. We present a preliminary assessment of the observations and indicate the direction of required modeling efforts. [Work supported by ONR.]

9:00

**1aAO4. Acoustic propagation measurements in the surf.** Kerry W. Commander and John S. Stroud (Coastal Systems Station, Code R21, 6703 W. Hwy. 98, Panama City, FL 32407-7001, CommanderKW@ncsc.navy.mil)

As part of the Near Shore Acoustic Network Experiment (NSANE), measurements were conducted in a very shallow water (VSW)/surf zone (SZ) region to investigate acoustic propagation and coherence in coastal areas. These measurements utilized an acoustical array of 16 hydrophones (supplied by Datasonics, Inc.) deployed over many tens of meters, and a broadband source. The data to be presented were acquired from hydrophones on two vertical arrays, separated by 240 m, which comprised half of the total array. The broadband projector was operated in the 3.5–28 kHz range using a variety of pulse types. The projector was aimed seaward, as it was located closer to shore than any of the receivers. Data were acquired in conjunction with environmental data by other NSANE team members. The data were analyzed using coherence techniques. Results of this acoustic measurement will be reported. [Work supported by ONR Code 3210E.]

9:15

**1aAO5. The acoustic environment of the surf zone.** Grant B. Deane and M. Dale Stokes (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238)

The topics of this presentation are the factors that determine acoustic transmission and ambient noise in the surf zone. The transmission and reception of acoustic signals in the surf zone are complicated by a rich variety of oceanographical phenomena. The bubble field created by breaking surf strongly absorbs and scatters sound, and the reception of acoustic signals transmitted through the breaking region is quite sensitive to the sizes and spatial distribution of bubbles in both the horizontal and vertical directions. Because wave-induced bubbles are advected by prevailing long-shore and off-shore currents, a deterministic solution to the transmission problem requires a coupled oceanographic/acoustic model, and in addition must include the effects of bubble absorption, surface wave scattering and multi-path propagation in the wedgelike, near-shore waveguide. Temperature gradients caused by shoaling internal waves and bubble-induced sound speed profiles may also play a role in signal transmission. Some examples of these phenomena and their effect on acoustic transmission will be given. Ambient noise, which also forms part of the surf zone acoustic environment, will be discussed; the dominant sources in the surf zone are wave breaking noise, biological sounds and, in populated regions, manmade noise. [Work supported by ONR.]

9:30

**1aAO6. Low-frequency acoustic resonances in the bubble plume formed by a plunging water jet.** Thomas K. Berger, Thomas R. Hahn, and Michael J. Buckingham (Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, tberger@ucsd.edu)

As a jet of water penetrates a free water surface, air is entrained in the form of bubbles which produce sound by various mechanisms, including both individual and collective oscillations. The low-frequency acoustic emissions of the bubble clouds entrained and dispersed by vertical continuous and transient fresh water jets impinging from varying heights into a fresh water receiving pool in a large laboratory tank were studied both theoretically and experimentally for jet velocities ranging to about 10 m/s, and jet diameters of millimeter order. The acoustic signals were measured with hydrophones in different positions, amplified, and digitally sampled. The measured power spectra revealed well-defined, nonuniformly spaced resonance peaks at frequencies below about 1 kHz. A theory based on a conical bubble plume, with a nonuniform sound speed profile, acting as a resonant cavity predicts the frequencies of the resonance peaks and shows remarkable agreement with the experimentally measured values. Further, time-frequency analysis gives bubble size distributions and shows the

driving source for the resonant cavity to be large secondary bubbles formed through coalescence, which become acoustically active through deformation or break-up as they are re-entrained by the downstream flow in a failed attempt to rise to the surface.

9:45–10:00 Break

10:00

**1aAO7. Decay mechanisms for large underwater bubble oscillations.** B. Edward McDonald and Charles Holland (Saclant Undersea Res. Ctr., 19138 La Spezia, Italy)

Acoustic signatures from centimeter-scale underwater bubble oscillations (primarily from lightbulb implosions) have been found to exhibit much higher decay rates than can be explained using viscous, thermal, or radiative mechanisms which apply to microbubbles. Where does the energy go? If one assumes energy transfer to shape oscillations (surface capillary waves) of large amplitude in subharmonic resonance with the breathing mode [M. S. Longuet-Higgins, *J. Acoust. Soc. Am.* **91**, 1414 (1992)], the shape oscillations can drive fluid motions outside the bubble capable of exciting turbulent instabilities. Application of an appropriate eddy viscosity from mixing length theory to the viscous decay mechanism appears to offer a credible mechanism to explain the observed large decay rates. We give an analysis to show that energy is transferred from the breathing mode to surface capillaries fast enough to make the proposed decay mechanism viable. [Work supported by Saclantcen.]

10:15

**1aAO8. Planning for the first Asian Sea International Acoustics Experiment (ASIAEX).** Jeffrey Simmen (Office of Naval Research, 800 N. Quincy St., Arlington, VA 22217-5660) and Warren W. Denner (EOS Res. Assoc., Monterey, CA 93940)

Organization and planning has been ongoing for the first comprehensive international shallow water acoustic experiment in Asia. Four workshops have been held to reach the final scientific plan: December 1997 (San Francisco), June 1998 (Seattle), April 1999 (Singapore), and July 1999 (Anchorage). This paper will describe ASIAEX, which will consist of two components—volume interaction and bottom interaction. The sites for these components are expected to be a shelf break region in the South China Sea and a shallow water region in the East China Sea, respectively. ASIAEX may involve scientists from the United States, China, Taiwan, South Korea, Singapore, and Russia.

10:30

**1aAO9. Acoustic monitoring of generation and evolution of shallow-water solitary internal waves.** Altan Turgut, Stephen Wolf, Marshall Orr (Naval Res. Lab., Acoust. Div., Washington, DC 20375), and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

The feasibility of acoustic monitoring of generation and evolution of shallow-water solitary internal waves is investigated by using the acoustic and environmental data collected during the 1995 Shallow Water Acoustics in a Random Medium (SWARM) experiment. Spatial and temporal variability of acoustic signals (60 to 450 Hz) were measured by using two well-populated vertical receiving arrays at 30- and 40-km ranges from two moored sources and at 20-km cross-range from two ship-deployed sources. Environmental data were collected by using CTDs, thermistor strings, a high-frequency acoustic backscattering imaging system, and a chirp sonar system. Correlation between cross-range propagation and oceanographic measurements during the generation and evolution of a solitary internal wave packet indicates the possibility of acoustic monitoring of this oceanographic process. Results of a parametric study using broadband numerical simulations are also presented to develop the best monitoring strategies for future experiments. [Work supported by ONR.]



**1aAO10. Horizontal refraction and sound fluctuations caused by internal waves packets in shallow water.** Boris Katsnelson and Serguey Pereselkov (Phys. Dept., Voronezh Univ., Universitetskaya sq.1, Voronezh, 394693 Russia, katz@mph.vsu.ru)

The subject of the paper is to consider the propagation of a low-frequency sound field in shallow water in the presence of solitonlike internal waves (SLIW). It is assumed that SLIW spread approximately across the stationary acoustic trace. The problem of sound propagation is considered within the framework of 3-D model of the shallow water. The theory of "horizontal rays and vertical modes" is applied to take into account influence of horizontal refraction due to SLIW. As shown in the paper SLIW can lead (periodically in time) to "focusing" and "defocusing" of the horizontal rays or, in other words, to the periodic appearance of "horizontal waveguide." Estimations of the sound fluctuations due to SLIW demonstrate that the amplitude of fluctuations can reach up to 10–20 dB (between moments of the "focusing" and "defocusing"). The next effect conditioned by internal waves is horizontal filtering of vertical modes, where we have separate focusing and defocusing rays, corresponding to different vertical modes at different points of horizontal waveguide. The results presented in the paper can be used as the basis of remote sensing of the internal waves packets in shallow water. [Work supported by RFBR, Grant 97-05-64878.]

**1aAO11. Internal waves and sound-speed distributions in the Yellow Sea.** Tao Wang, Tian-fu Gao, Yun-peng Zhang, and De-jun Jiang (Inst. of Acoust., Academia Sinica, 17 Zhong Guan Cun Rd., Beijing 100080, PROC)

The theory of resonant scattering of sound by internal solitons in the Yellow Sea is well known [Zhou *et al.*, *J. Acoust. Soc. Am.* **90**, 2042–2054 (1991)]. However, the oceanographic measurements made were rather sparse. In this paper, the time dependencies and the vertical structures of the internal waves and the resulting sound speeds over a period of 49 h 49 min in a strong seasonal thermocline of the Yellow Sea are presented. A temperature chain, sampling once per 6.42 s and consisting of 32 sensors spaced 0.4 m apart, provides the data. It is found that the volume sound speeds are significantly perturbed in the presence of the internal waves. At a fixed depth, the sound-speed perturbations may exceed  $45 \text{ ms}^{-1}$ . The mean sound-speed gradient is  $-8.3 \text{ s}^{-1}$ , but the gradient becomes  $-25 \text{ s}^{-1}$  when the thermocline is the strongest. The sound-speed distributions show the diurnal periodicity of the internal wave activity. The 6-h component, with high-frequency fluctuations riding on top of it, is also clearly seen. However, the data give no evidence of the presence of solitons although the internal waves are indeed nonlinear. [Work supported by NSFC.]

MONDAY MORNING, 1 NOVEMBER 1999

MORROW ROOM, 8:30 TO 11:45 A.M.

### Session 1aSA

## Structural Acoustics and Vibration: Mid-Frequency Modeling in Structural Acoustics I

Courtney B. Burroughs, Chair

*Applied Research Laboratory, Pennsylvania State University, Box 30, State College, Pennsylvania 16801*

### Invited Papers

8:30

**1aSA1. A zero-order energy finite-element method (EFEM) and its similarity to SEA.** Robert J. Bernhard and Shuo Wang (1077 Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN 47907-1077, bernhard@ecn.purdue.edu)

In this investigation it was found that a simple zero-order EFEM procedure using piecewise constant elements has essentially the same form as an SEA approximation. Thus, a relationship between SEA and EFEM can be established and the approximations of the two methods compared. The derivation of the zero-order EFEM approximation also indicates potential new formulations of coupling loss factors for damped and strongly coupled systems. These coupling loss factor formulations were used and verified against analytical solutions with both EFEM and SEA software. The predicted results converge to the analytical result as more elements are used.

9:00

**1aSA2. A hybrid FEA/SEA technique for the analysis of complex systems: Application to plate and beam structures.** Robin S. Langley (Dept. of Eng., Univ. of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK), Jeremy R. D. Smith, and Paul G. Bremner (Vibro-Acoust. Sci., Inc., San Diego, CA 92130)

The authors have previously developed a method for the low- to high-frequency vibration analysis of complex systems which is based on wave number partitioning. The long wavelength deformation of the system is modeled by using the finite element method (FEA), while the short wavelength deformation is modeled by using statistical energy analysis (SEA). The coarse mesh FEA model which is used to capture the long wavelength deformation of the system is enhanced by the addition of "fuzzy" mass and damping terms; this compensates for the "missing" short wavelength dynamics. The response yielded by the FEA model is then used in conjunction with SEA to give the short wavelength response. The method is noniterative, and it has previously been shown to produce good results for structures constructed from rod and beam elements. In this paper the method is extended to two-dimensional structural components such as plates. The method is applied to an example beam/plate assembly, and a comparison is made with the results yielded by SEA alone, and by a very detailed FEA model. The present approach is found to give significant improvements over SEA, while being very efficient compared to the detailed FEA model.

**1aSA3. Main difficulties in the midfrequency range and reduced matrix models for structural-acoustic problems.** Christian Soize (Structural Dynam. and Coupled Systems Dept., ONERA, 92322 Chatillon, France)

In the MF range, the modal density of 3-D structures increases with respect to the LF range, but is not yet high enough to be able to apply SEA, which is adapted to the HF range. This means that finite-element models have to be used to model the master system constituted of the master structure coupled with the internal and external acoustic fluids. The main difficulties in the MF range are due to the mechanical model and the solving methods. Concerning the mechanical model, the difficulties are induced by the structural complexity for which the fuzzy structure theory can be used; we present recent results concerning the validation of the fuzzy structure theory for continuous junctions between the master structure and the fuzzy substructures whose structural complexity is due to local modes of continuous structures on which are attached many discrete dynamical subsystems. Concerning analysis and solving methods in the MF range, finite-element models of master systems have a large number of DOF and modal analysis is not efficient for constructing a reduced model. We then present recent developments and applications concerning a reduced model adapted to structural-acoustic systems in the MF range.

10:00–10:15 Break

### Contributed Papers

10:15

**1aSA4. Models of uncertainty in structural dynamics.** Robin S. Langley (Dept. of Eng., Univ. of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK)

At medium to high frequencies small changes in the properties of a structure can have a significant effect on the vibrational response. The theoretical methods which are currently available to assess the influence of structural uncertainties are either probabilistic (for example, Monte Carlo simulations, FORM and SORM reliability methods, and asymptotic analysis of probability integrals) or possibilistic (for example, convex modeling, interval analysis, and fuzzy set theory). The interrelationships between these methods are far from clear, and research in each area tends to progress independently of the others. It is shown here that there is a very close relationship between asymptotic analysis, FORM, SORM, and the full range of possibilistic techniques: in fact, a single algorithm can encompass all the methods. This means that software developed for FORM and SORM can be applied to convex modeling (for example) providing the input variables and output quantities are reinterpreted in an appropriate way. It then becomes straightforward to select and implement the analysis method most suited to the available data. The theoretical developments are illustrated here by application to an example dynamic system.

10:30

**1aSA5. A comparison between an indirect Trefftz method and the finite-element method for solving coupled vibro-acoustic problems.** Wim Desmet, Paul Sas, and Dirk Vandepitte (K. U. Leuven, Dept. of Mech. Eng., Div. PMA, Celestijnenlaan 300 B, B-3001 Leuven, Heverlee, Belgium, wim.desmet@mech.kuleuven.ac.be)

The finite-element (FE) method is most commonly used for solving time-harmonic structural acoustic problems in bounded domains. Since the dynamic variables within the element domains are expanded in terms of simple, approximating shape functions, a large amount of elements is required to get a reasonable prediction accuracy and this amount increases with frequency. This results in large, computationally expensive models, so that the use of FE models is practically restricted to low-frequency applications. In a newly developed, indirect Trefftz method, the dynamic field variables are expanded in terms of wavefunctions, which exactly satisfy the governing dynamic equations. The wavefunction contributions to these expansions result from a weighted residual formulation of the boundary conditions. Since an approximation is only involved with the boundary conditions, the resulting models are substantially smaller than FE models. However, the small wave models are fully populated, com-

plex, and frequency dependent. The performances of the indirect Trefftz method and the FE method are compared for some validation examples. It is illustrated that the new method provides a high prediction accuracy with substantially smaller computational efforts. In this way, the indirect Trefftz method can be applied up to much higher frequencies.

10:45

**1aSA6. Midfrequency acoustical modeling of a coated plate.** Andrew Norris (Rutgers Univ., Mech. and Aerosp. Eng., 98 Brett Rd., Piscataway, NJ 08854-8058)

A steel plate with an attached thick viscoelastic layer exhibits “mid-frequency” characteristics at relatively low frequencies, certainly below 1 kHz for a 1-in. steel plate with thicker coating. The low-shear stiffness of the coating leads to a cut-on frequency associated with a shear thickness resonance. Classical plate theory or composite plate theory cannot capture this phenomenon. Although it is not difficult to compute the exact dispersion and modal properties of this system, an approximate plate theory is required in order to examine how the coated plate interacts dynamically with uncoated regions (reflection, transmission), and for modeling scattering from structure incorporating these elements. A hybrid plate theory is presented, based on a classical Kirchhoff assumption for the steel, and a Timoshenko/Mindlin model for the coating. This leads to a three degrees of freedom model, which exhibits the same dispersive properties as the exact theory. Comparisons and differences are discussed. The effects of fluid loading are also presented. [Work supported by ONR.]

11:00

**1aSA7. Basic development of a hybrid finite element method for mid-frequency computations of structural vibrations.** Xi Zhao and Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109-2145)

In the past it has been demonstrated that energy finite element analysis (EFEA) allows computation of high-frequency structural vibration in complex marine structures and produces results similar to statistical energy analysis (SEA) predictions [Vlahopoulos *et al.*, “Numerical implementation, validation, and marine applications of an energy finite element formulation,” accepted by the Journal of Ship Research]. In this paper a basic theoretical development of a hybrid finite element method is presented. Conventional finite element analysis (FEA) is combined with

EFEA in order to achieve solution to mid-frequency vibrations. The new formulation is based on deriving appropriate interface conditions at the joints between members modeled by the FEA and the EFEA methods. Effects of strong coupling, power re-radiation, and power re-injection can be accounted in the analysis. Examples of systems of co-linear beams will be presented. Analytical solutions are compared to the hybrid solution in order to validate the theoretical developments. [Research supported by Automated Analysis Corporation.]

11:15

**1aSA8. Structural acoustic noise modeling of a supercavitating high-speed vehicle using statistical energy analysis.** Robert M. Koch (Adv. Technol. Div., Naval Undersea Warfare Ctr., Code 8232, Bldg. 1302, 1176 Howell St., Newport, RI 02841-1708, kochrm@npt.nuwc.navy.mil)

The simulation of propagation of structureborne vibratory and acoustic energy within a supercavitating high-speed underwater vehicle is studied herein. The supercavitating vehicle is rocket propelled and rides entirely within a cavity in order to reduce drag and allow for higher underwater speeds. A ventilation system exists within the vehicle to ensure the cavity completely encloses the vehicle. With the exception of isolated control surfaces, the only wetted portion of the vehicle is the cavitating sonar array at the vehicle nose. This paper focuses on the use of statistical energy analysis (SEA) to quantify and parametrize the high-frequency transmission of vibro-acoustic energy within the complex, multiply connected marine vehicle, as well as including contributions from the reverberant cavity and seawater acoustic paths. Additionally, an examination of the many potential noise sources is made, including a discussion on han-

dling the effects of surface backscattering and volume reverberation within an SEA framework. Numerical case studies are performed to exercise the model in order to predict the effects of specific design modifications such as structural design variations, isolation mounting scenarios, etc.

11:30

**1aSA9. Energy flow modeling in the mid-frequency range.** Philip J. Shorter (Vibro-Acoust. Sci., Inc., 12555 High Bluff Dr., Ste. 310, San Diego, CA 92130) and Brian R. Mace (Univ. of Auckland, Auckland, New Zealand)

This paper describes how uncertainties in the properties of a structure may be included in energy flow models derived from finite element analysis. Current approaches to stochastic dynamics either adopt a detailed probabilistic approach, as in the stochastic finite element method, or a broad brush approach as in statistical energy analysis (SEA). In the former, detailed information about the joint probability density function of a large number of physical properties is required in order to define an ensemble. In the latter, the ensemble is defined implicitly by assuming some distribution of the local dynamic properties of a structure (typically one that is uniform in some dynamic sense). In the mid-frequency range neither approach is entirely satisfactory. This paper presents an alternative definition of the ensemble, defined in terms of the statistics of the local natural frequencies of a structure. A numerical technique, which combines component mode synthesis with a Taylor series expansion and a Monte Carlo simulation, is then described and used to obtain a stochastic energy flow model of a structure.

MONDAY AFTERNOON, 1 NOVEMBER 1999

MARION ROOM, 2:45 TO 4:45 P.M.

### Session 1pAO

## Acoustical Oceanography: Current and Long Range Tomography

Mohsen Badiey, Chair

*College of Marine Studies, University of Delaware, Robinson Hall, Newark, Delaware 19716*

### Contributed Papers

2:45

**1pAO1. Current tomography in very shallow water.** M. Badiey (Univ. of Delaware, Col. of Marine Studies, Newark, DE 19716)

Shallow and estuarine environments share a common acoustical property in that multi-path arrival structure of a pulse propagating through these regions generally occurs. For very shallow water depths, the direct-path becomes scattered path in a short range between the source and the receiver. The received acoustic signal becomes a superposition of many arrivals following all the possible acoustic paths between source and receiver. Most of these paths interact with sea surface and sea bottom several times. The impact of tidal currents and the changes in temperature and salinity profiles are most easily observed by focusing on the direct-path component of the received signal. Because the direct-path is not easily obtained in these regions of interest, the question of how and to what extent one can use the latter arrivals for current inversion is legitimately addressable. In this paper, we show the results in the application of current tomography in water depths of approximately 15 m. Also, the possibility

of obtaining oceanographic parameters such as surface roughness by using broadband reciprocal transmission of a local source and receiver as well as a far source-receiver configuration is discussed.

3:00

**1pAO2. Multi-frequency scintillation method for ocean flow measurement.** Joseph Fuks and Konstantin Naugolnykh (Univ. of Colorado, CIRES/NOAA, Environ. Technol. Lab., 325 Broadway, Boulder, CO 80303, kan@etl.noaa.gov)

A multi-frequency scintillation method for ocean flow measurement is presented. The transverse flow of inhomogeneous fluid produces fluctuation of the acoustic signal passing through it. These fluctuations vary with the cw signal frequency change due to variation of the Fresnel zone size and diffraction angles. Respectively, the fluctuations of signals at two different frequencies are coherent at low frequency and noncoherent in the high-frequency band. The frequency cutoff of the correlation function of signals fluctuation depends upon the flow velocity at given fine structure of the flow. The measurement of the cutoff frequency allows one therefore

to determine the flow velocity of the fluid [J. Fuks, Radiotekhn. Elektron. (in Russian), **3**, 515–524 (1975)]. This method can be considered as the “frequency-domain” version of the conventional scintillation approach to the current velocity registration based on measurement of correlation of the signals transmitted from the source to two separated in space receivers and can be, respectively, considered as space-domain scintillation method [S. Clifford and D. Farmer, J. Acoust. Soc. Am. **74**, 1826–1832 (1983)]. The features of the multi-frequency method of the ocean current measurement including the optimal frequency dependence upon the distance and the ocean environment are considered in the paper.

3:15

**1pAO3. Preliminary result of Central Equatorial Pacific Ocean Tomography Experiment.** Hidetoshi Fujimori, Iwao Nakano, Toshiaki Nakamura, Tomoyuki Kanaizumi (JAMSTEC, 2-15 Natsushima-cho, Yokosuka, Kanagawa, 237-0061 Japan, fujimori@jamstec.go.jp), Bruce M. Howe (Univ. of Washington, Seattle, WA 98105), Takashi Kamoshida (OKI Electric Industry Co. Ltd., 4-11-12 Shibaura Minato-ku Tokyo, Japan), and Gang Yuan (Univ. of Hawaii, Honolulu, HI 96822)

JAMSTEC (Japan Marine Science and Technology Center) and APL (Applied Physics Laboratory, University of Washington) started a joint program to directly measure part of the circulation cell of the Central Equatorial Pacific Ocean using acoustic tomography. The tropical and subtropical Pacific Ocean are suggested to connect via a shallow meridional circulation cell by Gu and Philander (1997). This subtropical cell (STC) has been hypothesized to the El Nino-Southern Oscillation phenomena in the tropics, which then, in turn, can affect the subtropical ocean via rapid atmospheric forcing. But no direct measurements exist and indirect measurements and modeling support the hypothesis. The observational centerpiece is tomography systems located in the Central Equatorial Pacific Region. JAMSTEC began the field work by deploying five acoustic transceivers in the observational region and this work will continue until January 2001 (two-year plan). Measurement data were sent to JAMSTEC in realtime using satellite telemetry system. These data were analyzed and compared with CTD observations. The time series fluctuation of the temperature was solved by this analysis.

3:30

**1pAO4. Effects of bathymetry on long-range acoustic transmissions.** Kevin Heaney, Brian Sperry, and Christopher Tiemann (Sci. Applications Intl. Corp., 888 Prospect St., Ste. 201, La Jolla, CA 92037)

Long-range acoustic propagation has been used recently to perform basin scale inversions of the temperature field of the oceans. The inversion technique involves measuring acoustic travel times and inverting them for sound-speed structure using a forward acoustics model. Ocean acoustic tomography has generally been formulated and applied to problems with little bottom interaction where ray theory is a suitable forward model. Power and reduced source motion requirements have led the global acoustics community to use near-seafloor moored sources off the coast of islands or continents. We examine the effects that scattering with the seafloor near the source and near the receiver has on broadband acoustic signals. Comparisons are made between range-dependent oases, the parabolic equation, and ray codes for several bottom types and bathymetry contours. The effect of scattering on travel time structure in time/depth space as well as mode amplitudes will be addressed. The implications of this work to future plans for global scale acoustic thermometry and internal wave inversion will be presented.

3:45

**1pAO5. Acoustic thermometry data compared with ocean models: The importance of ENSO in modifying the ocean interior.** John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 3440 Market St., Ste. 400, Philadelphia, PA 19104-3325, johnsr@sas.upenn.edu), Harley E. Hurlburt (Naval Res. Lab., Stennis Space Center, MS 39529-5004), Mark Johnson (Univ. of Alaska–Fairbanks, Fairbanks, AK 99775-1080), Mark Keller (Pennsylvania State Univ., State College, PA 16804), Steven Meyers (Univ. of South Florida, St. Petersburg, FL 33701), and James J. O’Brien (Florida State Univ., Tallahassee, FL 32306-3041)

The difficulties in detecting anthropogenic changes in ocean temperature due to the industrial revolution are discussed using acoustic thermometry data taken between Oahu and seven SOSUS receivers at distances of 3000–4000 km. Measurements were made in late 1983, and over two 5-month intervals between 1987 and 1989. The travel times are dominated by interannual fluctuations. Two hydrodynamic ocean models are used to identify plausible oceanic features that could cause these variations. Modeled El Niños and La Niñas exhibit oceanic teleconnections between the Equator and mid-latitudes that lead to Rossby waves that propagate westward at mid-latitudes. Rossby waves are the dominant model features which affect the modeled acoustic travel times, and hence section-averaged temperatures in the eastern North Pacific. Modeled predictions of the travel times are significantly different than some of the data. The magnitude and rates of changes of the data are the same as those which result from the effects of the Rossby waves. The largely unknown large-scale variability of the ocean at interannual and decadal scales is a principal problem for detecting possible changes in temperature due to human activity. Effects from eddies, seasonal cycles, and internal waves are of second order.

4:00

**1pAO6. Advantages of mapping climatic temperature changes in the ocean with mobile receivers.** John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 3440 Market St., Ste. 400, Philadelphia, PA 19104-3325, johnsr@sas.upenn.edu), Anatoly L. Fabrikant (Therm-Wave, Fremont, CA 94539), Anisim A. Silivra (Pennsylvania State Univ., State College, PA 16804), and Harley E. Hurlburt (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

An acoustic tomography simulation is carried out in the North Pacific ocean to assess whether climate trends are better detected and mapped with mobile or fixed receivers. Natural variability of the sound-speed field is simulated with the Naval Research Laboratory (NRL) layered-ocean model. A Kalman filter is used to estimate the sound speed field, where the *a priori* error covariance matrix of the parameters is estimated from the model. A spatially homogeneous climate trend is added to the NRL fluctuations, but the trend is not parametrized in the Kalman filter because, in reality, the ocean’s variability is poorly known. The effects of the unparametrized climate trend are projected onto parameters which eventually drift beyond acceptable limits, which triggers the trend’s detection. Mobile and fixed receivers detect the trend at the same time, but the maps from the mobile receivers are more accurate because they provide a synthetic aperture which better samples the variability than the fixed receivers [A. L. Fabrikant *et al.*, IEEE Ocean. Eng. **23**, 20–25 (1998)]. This simulation is one of several which motivates a program designed to show that U.S. submarines are useful for collecting tomographic signals for scientific study. [Work supported by ONR.]

4:15–4:45

Panel Discussion



**Session 1pEA****Engineering Acoustics: Transducer Loss Mechanisms**

Arthur Ballato, Cochair

*U.S. Army Communications, Electronics Command, AMSEL-RD-CS, Fort Monmouth, New Jersey 07703-5201*

Harold C. Robinson, Cochair

*Naval Undersea Warfare Center, Code 2131, 1176 Howell Street, Newport, Rhode Island 02841-1708***Invited Papers****2:00**

**1pEA1. The use of complex material constants to model losses in piezoelectric.** Binu K. Mukherjee, Guomao Yang (Dept. of Phys., Royal Military College of Canada, Kingston, ON K7K 7B4, Canada), and Stewart G. Sherritt (Jet Propulsion Lab., Pasadena, CA)

Complex material constants are now being used to accurately model the dynamic response of piezoelectric materials. A simple physical model of piezoelectric materials will be used to show how complex material constants can be conveniently used to characterize losses in the dielectric, elastic, and piezoelectric response of piezoelectric materials. The same model is also used to explain why the material constants are also a function of frequency, temperature, and applied signal strength. Theoretical attempts to understand the loss mechanisms and experimental methods to measure the various functional dependencies will be reviewed. [Funding support from the U.S. Office of Naval Research is gratefully acknowledged.]

**2:25**

**1pEA2. Loss mechanisms in piezoelectrics and resonance/antiresonance.** Kenji Uchino, Jiehui Zheng, Yun-Han Chen, Xiaohang Du (Intl. Ctr. for Actuators and Transducers, Mater. Res. Lab., Penn State Univ., University Park, PA 16802), and Seiji Hirose (Yamagata Univ., Yonezawa, 992 Japan)

Losses in piezoelectrics are considered in general to have three different mechanisms: dielectric, mechanical and piezoelectric losses. This paper deals with the phenomenology of losses, first, then how to measure these losses experimentally. We will discuss two methods for measuring the electrical, mechanical, and piezoelectric coupling losses separately: (1) strain versus stress and field and displacement versus stress and field curves, and (2) a resonance/antiresonance technique. Also, one can measure heat generation at an off-resonance frequency to obtain a total loss. By combining the above methods, we can investigate the loss mechanisms in practical piezoelectric materials and their consequences in devices. It will be shown that heat generation is caused mainly by the dielectric loss, not by the mechanical loss. Furthermore, a drastic decrease in mechanical Q with an increase of vibration level is observed in resonant piezoelectric ceramic devices, a result not observable using conventional impedance analyzers at low power. Based on this loss knowledge, we will discuss the differing situations between piezoelectric resonance and antiresonance states. In particular, since the mechanical quality factor at an antiresonance frequency is larger than that at a resonance frequency the antiresonance mode seems to be superior for high-power applications.

**2:50**

**1pEA3. Modeling and control issues concerning smart materials.** Ralph C. Smith (CRSC, Dept. of Mathematics, Box 8205, North Carolina State Univ., Raleigh, NC 27695)

Modeling and control issues concerning certain smart material actuators utilized in nonlinear regimes will be presented. Piezoelectric, electrostrictive, and magnetostrictive materials all exhibit various degrees of hysteresis and nonlinear dynamics at high-drive levels. The accurate and efficient quantification of these effects and their incorporation in control design are necessary to attain the full capabilities of the materials. The models described here are based on the quantification of domain effects inherent to ferroelectric and ferromagnetic materials. These models require a small number of parameters and are robust with respect to a variety of operating conditions. Hence, they are amenable to control design. Linear control methods prove ineffective at high-drive levels and certain nonlinear control laws and inverse compensation techniques for smart material applications will be discussed.

**3:15**

**1pEA4. Large signal dielectric losses in sonar transducers.** Harold C. Robinson (NAVSEA Newport, Code 2131, Newport, RI 02841, robinsonhc@npt.nuwcc.navy.mil) and Mark B. Moffett (Georgia Tech. Res. Inst., Atlanta, GA 30318)

The dielectric loss factor  $\tan \delta$  is a critical parameter in transducer design and performance prediction, as it is directly related to the electrical energy lost to Joule heating. For an ideal linear material,  $\tan \delta$  is a constant irrespective of the operating conditions. However, electrostrictive materials such as lead magnesium niobate/lead titanate (PMN/PT) exhibit significantly nonlinear behavior. It is, therefore, expected that  $\tan \delta$  (as well as the other material constants) will be a strong function of the transducer operating conditions. In this paper we will present three different methods for calculating the loss factor of "quasilinear" materials. It will be shown that, in the limit of small signals about a bias point, the three methods are consistent with the standard linear definition. These

three methods will be applied to material measurements taken at NAVSEA Newport on PMN/PT. The behavior of the loss tangent as a function of the dc bias field, ac drive field, prestress level, and temperature will be examined. Finally, the predicted response of stacks and transducers using  $\tan \delta$  calculated from the material measurements will be compared with experimental results. [Work sponsored by PEO(USW)-P and the Office of Naval Research.]

### *Contributed Paper*

3:40

**1pEA5. The stress dependence of the piezoelectric  $d$  coefficient of piezoelectric and electrostrictive ceramic materials.** Guomao Yang and Binu K. Mukherjee (Dept. of Phys., Royal Military College of Canada, Kingston, ON K7K 7B4, Canada)

An experiment has been developed to measure the piezoelectric  $d$  coefficient as a function of uniaxial stress on a piezoelectric or electrostrictive ceramic material. Measurements on piezoelectric ceramic specimens show that there is a very significant enhancement in the value of the

piezoelectric coefficient as a function of applied stress with the value of the coefficient going up by a factor of about 2.5 over a 100-MPa range of applied stress. The increase in the piezoelectric coefficient continues until the stress is large enough for the ceramic to start depoling. Electrostrictive ceramics also show a similar trend, although their lack of mechanical strength prevents measurements over a large range of stress. Our results suggest that the performance of transducers can be enhanced by ensuring that the active piezoelectric material is under stress. [Funding support from the U.S. Office of Naval Research is gratefully acknowledged.]

MONDAY AFTERNOON, 1 NOVEMBER 1999

MORROW ROOM, 1:15 TO 4:30 P.M.

### **Session 1pSA**

## **Structural Acoustics and Vibration: Mid-Frequency Modeling in Structural Acoustics II**

Courtney B. Burroughs, Chair

*Applied Research Laboratory, Pennsylvania State University, Box 30, State College, Pennsylvania 16801*

### *Invited Papers*

1:15

**1pSA1. Computationally efficient modeling of structural inhomogeneities by analytical/numerical matching.** Donald B. Bliss, Rickard C. Loftman, Christopher D. Park, and Linda P. Franzoni (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Box 90300, Durham, NC 27708)

Solving structural acoustic problems in the mid-frequency range is a challenging task. Computation is expensive due to both the large well-coupled nature of the domain and resolution constraints such as small flexural wavelength. In addition, discrete structural elements such as ribs and stiffeners cause localized high-resolution content that requires much greater resolution than would be otherwise warranted. A method called analytical/numerical matching (ANM) alleviates this additional computational burden. The ANM method captures the particular influence of the inhomogeneity on the structure in an analytically expressed local solution. The governing equation is then used to show that the extraction of this local solution by superposition amounts to replacing the original discrete influence of the inhomogeneity by a smoothed forcing derived from the local solution. The result is a smoother problem that can be computed more efficiently and without the loss of any information. The method has been demonstrated for several configurations of increasing complexity in the process of its development as a useful tool for efficiently solving realistic problems of engineering interest. These problems have shown the usefulness, accuracy, and efficiency of the method. Examples will be presented, including recent work on implementation within a finite element analysis.

1:45

**1pSA2. The fundamental mechanisms and control of power flow to cylindrical shells in the midfrequency range.** J. Gregory McDaniel (Dept. of Aeronaut. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

This presentation will survey the fundamental mechanisms of midfrequency power flow to cylindrical shells from point excitations and, based on these mechanisms, propose approaches to controlling power flow. The mechanisms derive from an analytical expression for the shell velocity that involves a Fourier series about the shell's circumference. Each Fourier coefficient is a sum of four waves that propagate in the axial coordinate. At high frequencies, these four waves resemble the flexural, evanescent, shear, and longitudinal waves that propagate in a flat plate and each has an associated cutoff frequency, below which it becomes spatially attenuated. When a wave is excited near its cutoff frequency, its amplitude rises dramatically and, therefore, the total power delivered to the shell is dominated by that wave. This view suggests an approach to controlling power flow in the midfrequency range by constraining the

shell from responding in circumferential harmonics whose waves are excited near cutoff. In this context, the attachment of a reinforcing structure that is axially coincident with the excitation but circumferentially distributed is explored. Using a genetic algorithm to optimize structural parameters, an example will illustrate that a circumferentially distributed structure controls power much more effectively than a circumferentially compact structure.

2:15

**1pSA3. Summary of some of the tools for mid-frequency analysis.** Joseph M. Cuschieri (Ctr. for Acoust. and Vib., Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, FL 33431)

Over the past few years various attempts have been made to identify solutions/approaches that can address the mid-frequency regime. The mid-frequency regime in this paper is being defined as that frequency range which lies above the frequency of the first few modes (5 to 10), and below the frequency at which the modes start to strongly couple. The definition of strong coupling between modes is somewhat arbitrary and an alternative to define the high-frequency bound would be that, in broadband analysis (one-third octave), there are at least five or more, or a sufficient number of modes, within the one-third octave frequency band, such that the response of the system can be estimated in the one-third octave frequency band, using broadband type tools, to within  $\pm 3$  dB (also may be considered arbitrarily selected). In this presentation, a summary of some techniques and results that have been identified to address the mid-frequency regime, as defined above, will be presented, together with an overview of results that have been generated under a study to address structure-borne analysis techniques.

2:45

**1pSA4. The mid-frequency range finite element (MFR-FE) method: MATLAB implementation and results.** Victor W. Sparrow (Grad. Prog. Acoust., Penn State Univ., 157 Hammond Bldg., University Park, PA 16802, sparrow@helmholtz.acs.psu.edu)

The mid-frequency range finite element (MFR-FE) method was developed by Christian Soize in the early 1980s, and it is well described in Soize's book with Roger Ohayon [*Structural Acoustics and Vibration*, (Academic, New York, 1998), Chap. 7]. The method combines finite element equations and the frequency translation property of the Fourier transform to solve acoustics and vibration problems in the mid-frequency range, at frequencies higher than possible with conventional finite element codes. In an effort to understand the method fully, an implementation of MFR-FE was developed in MATLAB [Mathworks Inc., Natick, MA], a natural environment for prototyping algorithms involving matrices and signal processing. The FE formulation was built upon one already prepared by Kwon and Bang [*The Finite Element Method Using MATLAB*, (CRC, Boca Raton, FL, 1997)], and initial results were recently described [Paper AIAA-99-1855, 5th AIAA/CEAS Aeroacoustics Conference]. The present talk will provide an overview of MFR-FEs and give example results, extending those previously shown. One advantage of the method is that it can be used with other improvements in FE technology, such as hp adaptive formulations. [Work sponsored by NASA Langley Research Center Structural Acoustics Branch.]

3:15–3:30 Break

### Contributed Papers

3:30

**1pSA5. An approximation to the frequency-average radiation efficiency of rectangular flat plates.** Murilo W. Bonilha (United Technologies Res. Ctr., 411 Silver Ln., MS 129/17, East Hartford, CT 06108, bonilhmw@utrc.utc.com) and Frank J. Fahy (Univ. of Southampton, Southampton SO17 1BJ, UK)

The average radiation efficiency of plate modes which are resonant in a given frequency band is termed the frequency-average radiation efficiency and it is an extremely useful quantity in applications of Statistical Energy Analysis to interior or exterior noise control. In the present work, computation of the frequency-average radiation efficiency of flat plates is made on the basis of the vibration field correlation of randomly excited flat plate structures [M. W. Bonilha and F. J. Fahy, *J. Sound Vib.* **214**, 443–467 (1998)]. The advantage of this representation is that the radiation efficiency can be economically computed for simply-supported, free, clamped, guided and spring-supported boundary conditions. In addition, corrections to account for the presence of stiffeners and plate curvature can also be incorporated in the computation. Furthermore, the current representation also allows the use of FEA-derived spatial correlation results for the case of plates with complex boundary conditions and accommodate spatially varying materials or geometric properties. Radiation ef-

iciency results for various boundary conditions are presented. It is shown that results for simply-supported boundary conditions agree well with those obtained by Leppington *et al.* [F. G. Leppington, E. G. Broadbent, and K. H. Heron, *Proc. R. Soc. London, Ser. A* **382**, 245–271 (1982)].

3:45

**1pSA6. Midfrequency structural acoustics by DDSM. I. Beam-like structures.** Charles H. Oppenheimer and Russel D. Miller (A&T/ETC, 2341 Jefferson Davis Hwy., Ste. 1250, Arlington, VA 22202-3809)

The direct dynamic stiffness method (DDSM) allows low and midfrequency structural acoustic analysis of complex structures containing beam- and plate-like members and may be used in conjunction with other models like finite-element (FE) models. The DDSM description of beam-like structures involves closed-form relationships between transverse, axial, and torsion loads and displacements, and is accurate at frequencies below a beam-like structure's fundamental cross section resonance frequency. Transverse shear and rotational inertia effects are included. Modeling of free layer and granular material damping treatment is discussed, and measurements on treated tubes are compared to predictions. A com-

parison is made between DDSM and FE solutions of a truss structure. The predictions are essentially identical, but the DDSM model is 91% smaller and requires 90% less computation.

4:00

**1pSA7. Midfrequency structural acoustics by the dynamic direct stiffness method. II. Plate-like structures.** Russel D. Miller (A&T/ETC, Century Bldg., Ste. 1250, 2341 Jefferson Davis Hwy., Arlington, VA 22202-3809, rdm@atinc.com)

The use of the dynamic direct stiffness method (DDSM) for the harmonic analysis of beam structures is well known. In this paper the development of a harmonic plate element with similar characteristics is discussed. Whereas the finite-element method requires approximately six nodes per wavelength, the harmonic plate element requires nodes only at element connections, terminations, and load points. The governing equations are used in the variational form to derive the direct stiffness matrix. The matrix form of the element derived is compatible with those of the FEM and may be used for the harmonic analysis of general FEM/DDSM structures. The motivation for this development is the need for computationally efficient evaluation of the midfrequency range for general structures where the FEM method requires a large number of degrees of freedom and statistical energy analysis (SEA) does not provide sufficient accuracy. A number of examples are provided to establish verification of the DDSM plate element. Examples include the response of harmonically driven plates, plates with beam stiffeners, and plate/beam truss structures.

4:15

**1pSA8. A single-degree-of-freedom approach to response perturbation due to feature variability.** Kenneth A. Cunefare (School of Mech. Eng., The Georgia Inst. of Technol., Atlanta, GA 30332-0405, ken.cunefare@me.gatech.edu) and Sergio De Rosa (Univ. of Naples, "Federico II," 80125 Naples, Italy)

This paper presents a technique to assess the impact on model predictions due to feature perturbations on harmonically forced fluid-loaded structural acoustic models. The perspective taken is one of knowledge of a reference state, with a desire to determine the impact on the total radiated acoustic power due to perturbations in the reference state. Such perturbations change the predicted resonance frequencies of a structure under consideration, and hence, change the predicted response amplitudes. The method uses a single-degree-of-freedom response model in the local region of each fluid-loaded resonance, coupled with eigenvalue sensitivities, to estimate the perturbation impact. The sensitivity of the eigenvalues to changes in model detail is derived based on variations in the spatial representation of attached features (e.g., point versus distributed attachments). Elements of the analysis method are not necessarily restricted to model perturbations nor acoustic power, rather they may be used to assess the perturbation of any quadratic response quantity of interest due to changes in resonance frequency. The analysis reveals that the bandwidth of response perturbation increases directly with increasing resonance frequency. This "proportional bandwidth" has bearing upon the so-called midfrequency gap in structural acoustics.

MONDAY AFTERNOON, 1 NOVEMBER 1999

KNOX ROOM, 1:00 TO 3:45 P.M.

### Session 1pSP

## Signal Processing in Acoustics and Physical Acoustics: Signal Processing Applications for Low Diffraction Beams

David H. Chambers, Cochair

*Lawrence Livermore National Laboratory, P.O. Box 808, L-333, Livermore, California 94551*

D. Kent Lewis, Cochair

*Advanced Applications, Lawrence Livermore National Laboratory, 7000 East Avenue, L-372, Livermore, California 94551*

Chair's Introduction—1:00

### Invited Papers

1:05

**1pSP1. A review of acoustic localized wave phenomenology.** Richard W. Ziolkowski (Dept. of Elec. and Computer Eng., Univ. of Arizona, Tucson, AZ 85721)

A brief review of the history that led to several classes of exact solutions of scalar wave equations that describe localized transmission of acoustic wave energy will be given. It has been shown that these acoustic localized wave (LW) solutions can be optimized so that they are localized near the direction of propagation and their original amplitude is recovered out to extremely large distances from their initial location. Pulses with these very desirable localized transmission characteristics have a number of potential applications in the areas of directed energy systems, secure communications, and remote sensing. Simulations will be shown for several aspects of LW pulse-driven arrays. Various performance criteria including near-field and far-field sidelobe, intensity, and energy distributions will also be discussed for acoustic LW pulse-driven arrays. Since these LW arrays deal with different ultra-wide bandwidth driving pulses at each element of the array, these criteria are generalizations of the standard continuous wave (CW), phased array results. Recent attempts to reduce the equivalent sidelobe levels will also be discussed. Past acoustic experimental setups and results will be summarized briefly in regards to all of these time domain versus frequency issues.



**1pSP2. Acoustic Bessel bullets: Where did they come from and where are they going.** Peter Stepanishen (Dept. of Ocean Eng., Univ. of Rhode Island, Kingston, RI 02881)

Acoustic Bessel bullets (BB's) are a new class of band-limited transient localized waves. A brief review of the development of acoustic BB's using space-time methods is first presented where the roots are traced back to the well-known plane-wave solution of the 1-D wave equation. Although ideal three-dimensional acoustic BB's which maintain their spatially dependent shape and time history as they propagate in free space can be generated from an infinite planar aperture, edge diffraction effects play an important role in determining the space-time characteristics of the acoustic BB's generated by a finite planar aperture. A generalized modal impulse response approach is proposed to investigate the characteristics of acoustic BB's and other localized wave fields generated from finite planar apertures with radially symmetric space-time source distributions. Numerical results are presented to illustrate the general space-time characteristics of the modal decomposition of the source distributions, the generalized impulse responses, and the associated acoustic BB fields. These results will illustrate the trade-offs between the radial extent of the acoustic BB's and the near-to-far-field transition. Several interesting characteristics of the far field will also be discussed.

2:05

**1pSP3. Limited diffraction beams and their applications on signal processing.** Jian-yu Lu (Ultrasound Lab., Dept. of Bioengineering, The Univ. of Toledo, Toledo, OH 43606, jilu@eng.utoledo.edu)

Limited diffraction beams such as Bessel beams and X waves have been studied recently. These beams have many potential applications in medical imaging, tissue identification, blood flow velocity vector estimation, nondestructive evaluation of industrial materials, optical communications, and other physics related areas such as electromagnetics and optics. In this talk, fundamentals of limited diffraction beams will be presented. Results of applications of X wave theory on high-frame rate medical imaging will be shown. Applications of these beams in other areas, especially, in signal processing will be discussed. [This work was supported in part by grant HL 60301 from the National Institutes of Health.]

2:45–3:00 Break

### Contributed Papers

3:00

**1pSP4. Broadband processing using localized waves in a shallow ocean environment.** David H. Chambers, James V. Candy, and D. Kent Lewis (Lawrence Livermore Natl. Lab., POB 808, Livermore, CA 94551)

In this paper the MPS localized wave pulse is used as the basis for designing a broadband signal processor to discriminate between two point sources. The nonseparable nature of the MPS pulse requires a "best fit" approach to processor design. This approach illustrates the kinds of problems one would encounter when using localized waves in a processor. The processor design is described, then applied to simulated acoustic signals generated in a shallow ocean environment. The results show that the MPS-based processor can discriminate between the two sources at least as well as a more conventional broadband beamformer. No attempt is made to estimate the positions of the sources. [This work was performed under the auspices of the Department of Energy by the Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.]

3:15

**1pSP5. Localized wave beam forming tests.** D. Kent Lewis, James V. Candy (Lawrence Livermore Natl. Lab., 7000 East Ave., Livermore, CA 94550), and Richard W. Ziolkowski (Univ. of Arizona, Dept. of Elec. and Computer Eng., Tucson, AZ 85721)

Much has been reported about work done creating the localized wave (LW) acoustic pulse using active beam creation. This presentation will show work done in passive beam forming using the LW source signals as array filter coefficients. The simple tests which were performed showed

that the LW filter can detect wide bandwidth signals in an idealized medium, and that there is an apparent increase in signal resolution. These results suggest that there may be applications for other low-diffraction beam techniques in passive signal processing. [This work was performed by the Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under Contract No. W-7405-ENG-48.]

3:30

**1pSP6. Acoustical helicoidal wave transducers with applications for the alignment of ultrasonic and underwater systems.** Brian T. Hefner and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

A simple four-panel transducer capable of producing a beam with a screw-dislocation along its axis was constructed and evaluated. A screw-dislocation in a wave front is characterized by a phase dependence about the dislocation axis that varies as  $\exp(-im\phi)$ , where  $m$  is an integer and  $\phi$  is the angle about the axis. At the axis the phase is indeterminate and as a result there is a corresponding null in the pressure magnitude. To generate a wave front with these characteristics, a four-panel 3-1 composite piezoelectric transducer was driven with the appropriate phasing of the panels to create dislocation along the beam axis. As a result the beam does not possess cylindrical symmetry, however, the dislocation is found to exist in both the far and near fields of the transducer. This null then clearly indicates the axis of the beam at all distances and has the potential to be used as an aid in the alignment of objects in sonar experiments or other similar applications. A related transducer was summarized previously [J. Acoust. Soc. Am. **103**, 2971 (1998)] and is also discussed here for the purposes of comparison. [Work supported by the Office of Naval Research.]

## Session 1pUW

## Underwater Acoustics: Matched Field Processing

John S. Perkins, Cochair

Naval Research Laboratory, Code 7145, Washington, DC 20375-5350

Arthur B. Baggeroer, Cochair

Department of Ocean Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 5-204, Cambridge, Massachusetts 02139

## Contributed Papers

1:00

**1pUW1. Conditions for ocean acoustic matched field inversions to be unbiased and attain the Cramer–Rao bound.** Nicholas C. Makris and Eran Naftali (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

In recent years, a wide variety of acoustic techniques have been developed to probe the marine environment. These techniques typically require the nonlinear inversion of acoustic field data measured on a hydrophone array. The data, however, are often randomized by the addition of naturally occurring ambient noise or by fluctuations in the waveguide's refractive index and boundaries. The nonlinear inversion of random data often leads to estimates with biases and variances that are difficult to quantify analytically. It has become popular in recent years to compute limiting bounds on the variance of these nonlinear estimators, since these bounds are usually much easier to obtain. A major problem, however, is that the bounds are only guaranteed to converge to the true variance of the estimator under sufficiently high signal-to-noise ratio (SNR). By applying higher order asymptotics, general expressions are obtained for the first-order bias and second-order variance of a general maximum-likelihood estimate. These expressions are applied in the matched field localization of a source in a shallow-water waveguide. The results quantitatively show that the Cramer–Rao bound provides an unrealistically optimistic estimate of the true variance in many practical scenarios, and the bias can be significant as the SNR passes below 0 dB.

1:15

**1pUW2. Stochastic matched field processing.** Arthur B. Baggeroer (MIT, Cambridge, MA 02139, abb@arctic.mit.edu) and Edward K. Scheer (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Stochastic matched field processing was introduced [Baggeroer and Kuperman, *J. Acoust. Soc. Am.* **91**] as a means to address stochastic variability and signal gain degradation in a replica signal. The approach parallels what is usually termed the “random signal in noise” problem in time series analysis wherein one uses a stochastic model to respond to the signal gain degradation issue when high-fidelity replicas cannot be generated. For stochastic MFP the important issues are: (i) the signal power scatters into a “red” spectra of eigenvalues where only a small number are significant, (ii) the eigenvectors are predictable based upon *a priori* environmental data such as internal wave or signal motion models in addition to nominal sound speed models. For many environments these conditions are met and, importantly, more robust, MFP can be obtained. This presentation will review (i) stochastic matched field, (ii) apply it to deep and shallow water random ocean models, and (iii) assess its performance using Chernoff detection theory bounds for low signal-to-noise ratio scenarios.

1:30

**1pUW3. Source localization in a stochastic ocean using data from the Santa Barbara Channel experiment.** Peter Daly and Arthur Baggeroer (MIT, 77 Massachusetts Ave., Rm. 5-204, Cambridge, MA 02139)

Source localization performance using matched field processing is degraded when: (1) incorrect environmental information is used in replica generation; and (2) the propagation model fails to correctly predict energy transfer from one spatially derived eigenvalue to another. Krolik [*J. Acoust. Soc. Am.* **92**] addressed (1) by developing a robust adaptive matched field processor assuming a stochastic propagation medium and a perturbed sound speed profile. Czenszak [*J. Acoust. Soc. Am.* **101**] extended Krolik's work to include multiple narrow-band frequencies. We approach both (1) and (2) by direct decomposition of the simulated pressure field based on environmental perturbations. This yields information on energy distribution from a single point source on spatially derived eigenvalues in a simulated environment. We will explore the application of both processors and extensions of them to range-dependent shallow water acoustic data taken from the 1998 Santa Barbara Channel experiment. Source localization performance of the stochastic ocean beamformer is contrasted with both Bartlett and White Noise Constrained processors at ranges to 15 km, with emphasis on performance metrics in the face of uncertain environmental knowledge.

1:45

**1pUW4. Broadband adaptive matched field processing with modal filtering.** Lisa M. Zurk, Nigel Lee, and James Ward (MIT Lincoln Lab., Lexington, MA 02420)

This paper will demonstrate 3D localization of a moving, broadband source in a shallow water channel using adaptive matched field processing (AMFP) with modal filtering. The adaptation is desirable to mitigate the high sidelobes of the conventional processor, but it also introduces the requirement for long observation times, which can be difficult to satisfy for moving sources. Rank reduction can be achieved through modal preprocessing by exploiting the sparsity of propagating modes. In addition, the modal decomposition provides a mechanism for filtering the signatures of surface interferers from those of a submerged source. Results will be presented that demonstrate rejection of surface interference as well as quantifying the relationship between localization accuracy and the size of the modal subspace. The acoustic data was obtained during the Santa Barbara Channel Experiment (SBCX) which employed a 150-hydrophone volumetric array consisting of five vertical line arrays (VLAs) instrumented for element localization. The VLAs fully spanned the 209-m water column and allowed accurate modal decomposition. The MFP results will be presented for exercises involving a research vessel, the ACOUSTIC EXPLORER (AX), towing an acoustic source. The results of broadband AMFP show accurate localization of both the towed source and towship.

2:00

**1pUW5. Sub-optimal adaptive matched-field processing of a three-dimensional array in a changing environment.** Yung P. Lee (Sci. Applications Intl. Corp., 1710 Goodridge Dr., McLean, VA 22102, yung@osg.saic.com)

The major problem in adaptively processing data from a large array is inversion of the covariance matrix. The matrix must be well-conditioned to be inverted. The number of independent samples that is needed to form a well-conditioned matrix is at least twice the number of array elements. The integration time to form a well-conditioned large array covariance matrix may then be unreasonably long relative to the coherent time of signal in a changing environment. A sub-optimal subarray processing has been developed. The adaptive weight vectors are calculated for each subarray. These weight vectors are linearly combined to form a full-array adaptive weight vector and are applied to the full-array covariance matrix. The required integration time is reduced to that which is needed to form the well-conditioned subarray covariance matrices. In a recent shallow water experiment, a three-dimensional array had been deployed. The array consisted of 5 VLAs each with 30-element spaced evenly on a circular baseline of a radius of 200 m. The subarray processing approach had been used to localize a moving source that had a fast range/bearing change relative to the baseline.

2:15

**1pUW6. Matched-field processing of two multi-dimensional arrays.** John S. Perkins, Michael Nicholas, Joseph F. Lingeitch, Gregory J. Orris (Naval Res. Lab., Washington, DC 20375), and Garry J. Heard (Defence Res. Establishment Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

Source location and tracking results using matched-field processing are presented for two multi-dimensional arrays. One array was L-shaped with the horizontal leg lying on the bottom. The vertical leg consisted of 32 elements with 2.5-m spacing and the horizontal leg consisted of 32 elements with 5.0-m spacing. The array was deployed in 100-m-deep water near the shelf break southeast of Halifax, Nova Scotia. The second array consisted of five vertical legs (each with 30 elements at 5.7-m spacing) arranged in a pentagon, 136-m on each side. This array was located in 200-m-deep water in the Santa Barbara Channel. For the L-shaped array, the individual legs produced results that were markedly different but extremely complimentary. As a result, their incoherent average often outperformed the fully coherent processing of the entire (64-element) array: the peak locating the source was narrower and the sidelobes were suppressed. This was not the case for the pentagonal array. The ambiguity surfaces from the individual legs were similar so that incoherent averaging broadened the source peak and did not suppress the sidelobes. Processing the entire array coherently produced excellent results. [Work supported by the Office of Naval Research.]

2:30

**1pUW7. Three-dimensional tracking of two blue whales using MFP and a tilted vertical array.** Aaron M. Thode, Gerald L. D'Spain, and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093-0205)

Two methods are explored for obtaining the bearing, range, and depth from three different blue whales, recorded in 1996 by a tilted vertical array deployed from the research platform FLIP. The first method uses the global inversion program SAGA to invert for the range, depth, and projected array tilt along the 2-D matched-field processing (MFP) plane between the animal and FLIP. Using array inclinometer information, the projected tilt

can be converted into a source bearing. A second method exploits the range-dependent bathymetry around FLIP to perform three-dimensional MFP. This method makes fewer assumptions about the array geometry, but is much more computationally intensive than the first approach, and is sensitive to bottom mismatch. DIFAR sonobuoy bearings from two whales provided an independent test of these MFP bearing estimates. It was found that the projected-tilt method could identify the correct azimuthal quadrant. The fully 3-D MFP method selected bearings within 20 deg of the true bearing, but exhibited a bias toward regions of low-range dependence. [Work sponsored by ONR.]

2:45

**1pUW8. Comparison of robust matched-field processing algorithms in a real, shallow water environment.** K. H. Kim and W. S. Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

In recent years, a number of matched-field processing (MFP) algorithms have been proposed to mitigate the ever-present problem of mismatch. This paper compares the source localization performance of six processors—Bartlett, Midpoint, Minimum Variance (MV) Distortionless Response, White Noise Constrained, MV with Neighboring Location Constraints, and MV with Environmental Perturbation Constraints—in a real, shallow water environment. MFP was performed on an 84.4-m aperture, 16-element subset of a MPL vertical line array deployed from the R/P FLIP west of Point Loma in approximately 200-m water. A continuous wave source, emitting ten tonals from 53–197 Hz, was towed over both range-independent and range-dependent tracks at source ranges of 1–7 km and at source depths of 60–130 m. First, MFP results utilizing a detailed geoacoustic model of the region, water column sound speed profiles measured during the experiment, and *a priori* knowledge of array tilt, demonstrate general processor characteristics. Second, MFP results which assume: (1) reasonable though further mismatched geoacoustic parameters; (2) water column sound speed profiles based upon historical data; and (3) no *a priori* knowledge of array tilt, illustrate processor performance under various mismatch conditions likely to be encountered in realistic applications.

3:00

**1pUW9. Adaptive matched beam processing for source localization.** Kwang Yoo and T. C. Yang (Naval Res. Lab., Washington, DC 20375)

Matched beam processing (MBP) has been previously applied to a horizontal line array for localizing a moving source in bearing and range, and for differentiating a submerged target from a surface clutter [T. C. Yang and T. Yates, *J. Acoust. Soc. Am.* **104** (1998)]. MBP is matched field processing in the beam domain. Using the Bartlett processor, bearing and depth resolutions are often inadequate for rejecting surface clutter and other noise sources. High-resolution processors may be required. We note that in situations where many clutters exist, adaptive (plane-wave) beamforming has effectively rejected clutter by either high-angular resolution or by steering a null to the interference source. Motivated by this approach, an adaptive MBP will be investigated in this paper. The new processor will apply adaptive beamforming to the data and replica fields such that interference has already been removed before the matched field correlation. This approach differs from the adaptive matched field processing where interference remains in the data field. The latter approach is known to be sensitive to environmental mismatch, as a result the correlation between the replica and the interference fields may be present as substantial clutter. Initial numerical results will be presented using simulated data with interference sources. [Work supported by the Office of Naval Research.]

1p MON. PM

NOTE: Attendance at this Tutorial requires payment of an additional registration fee.

MONDAY EVENING, 1 NOVEMBER 1999 REGENCY BALLROOM SOUTH, 7:00 TO 9:00 P.M.

**Session 1eID**

**Interdisciplinary: Tutorial Lecture**

Uwe J. Hansen, Chair

*Department of Physics, Indiana State University, Terre Haute, Indiana 47809*

***Invited Paper***

**7:00**

**1eID1. An evening with the art and science of music: The families of musical instruments: Physics and performance.** Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809) and James M. Pyne (Ohio State Univ., Columbus, OH 43210). The Ohio State Univ. String Quintet, Paul Robinson, Coordinator; The Ohio State Univ. Woodwind Quintet, Katherine Borst Jones, Coordinator; The Ohio State Univ. Brass Ensemble, James Akins, Coordinator; special Keyboard Demonstrations, Stephen Glaser, Piano; special Percussion Demonstrations, Michael Bump, Percussion; and The Ohio State Univ. Chamber Orchestra, Marshall Haddock, Conductor.

During this tutorial we hope to communicate some of the enjoyment associated with musical acoustics by presenting musical examples along with scientific concepts. During each segment of the presentation, the relevant musical instrument will be introduced, and at the conclusion of each section, the piece of music performed will feature the appropriate instrument family. Physics principles related to tone generation, instrumental timbre, pitch change, and amplification will be discussed. Musical examples will be performed by The Ohio State University String Quintet, Paul Robinson, Coordinator; Michael Davis, violin; Laurel Butler, violin; Edward Adelson, viola; William Conable, cello; and Paul Robinson, bass; by The Ohio State University Woodwind Quintet, Katherine Borst Jones, coordinator; Katherine Borst Jones, flute; Robert Sorton, oboe; Robert Jones, clarinet; Christopher Weait, bassoon; Heidi Wick, horn; by The Ohio State University Brass Ensemble, James Akins, coordinator; William Campbell, trumpet; Peter Gibbons, trumpet; Joseph Duchi, trombone; James Akins, tuba; by Steven Glaser, piano; and by Michael Bump, percussion. For the gala conclusion, The Ohio State University Chamber Orchestra will bring all instrument families together.



## Session 2aAB

## Animal Bioacoustics, Speech Communication and Signal Processing in Acoustics: Signal Processing Techniques for Animal Bioacoustics

Larry L. Pater, Chair  
USACERL, 2902 Farber Drive, Champaign, Illinois 61821

Chair's Introduction—7:55

### Invited Papers

8:00

**2aAB1. Applications of a high-quality sound manipulation algorithm STRAIGHT for animal voices.** Hideki Kawahara (Dept. of Design Information Sci. Faculty of Systems Eng., Wakayama Univ./ATR/CREST, 930 Sakaedani, Wakayama, 640-8510 Japan)

Applications of a high-quality speech analysis/modification/synthesis algorithm [Kawahara *et al.*, *Speech Commun.* **27**, 187–207] for animal voices are discussed. The proposed algorithm consists of (1)  $F_0$  adaptive spectral smoothing to eliminate periodic interferences on a time-frequency representation, (2)  $F_0$  and source information extractor based on a fixed point analysis of mapping from frequency to instantaneous frequency, and (3) group delay manipulation to control temporal structure of an excitation source. Each component has to be tuned to specifications of animal voices to take full advantage of the method. There are two types of applications: The first method is converting animal voices to sound more like human speech: This modification tries to enhance the researcher's auditory inspection of animal voices by mapping their voice physical parameters onto a region which is ecologically valid for humans. The second method is manipulating specific parameters of an animal voice to test function/contribution of each physical parameter: This modification tries to keep the naturalness of the stimuli while keeping a high degree of control accuracy of specific physical parameters of stimuli. This strategy would enable measurements of behavior responses in the vicinity of natural stimulus conditions.

8:30

**2aAB2. Wavelet transforms for bioacoustic signal processing.** Leon H. Sibul (Penn State Univ., Appl. Res. Lab. and Acoust. Prog., University Park, PA 16802, lhs2@psu.edu)

Wavelet transforms (WT) are natural tools for analysis of bioacoustic signals. In analysis of transient signals and nonstationary stochastic processes it is important to know not only what are the frequency content transients but when did these transient signals occur. The importance of representation of signals on the time-frequency plane, not only on frequency or time axes, is now widely accepted. Linear time-frequency analysis of transient signals (signals with time-varying frequency content) can be based on short-time Fourier transforms (STFT) and generalized Gabor transforms (GT). Wavelet transform analysis is analogous to STFT and GT analysis, except WTs represent signals on time and scale plane. In time-frequency analysis there is a tradeoff between time and frequency resolution. WTs have the advantage that they have constant time-frequency resolution over a wide frequency range. WT represent constant fractional or constant Q frequency analysis, thus efficiently analyzing a wideband of frequencies. Continuous wavelet transforms are linear transforms that can be inverted. In this tutorial paper basic properties of continuous and discrete wavelet transforms are reviewed, basic issues between Fourier transform techniques are discussed and disadvantages and advantages of WT are pointed out. [Work supported by ONR, Code 333.]

9:00

**2aAB3. Time-frequency analysis: A tutorial review.** Patrick J. Loughlin (Dept. of Elec. Eng., Univ. of Pittsburgh, Pittsburgh, PA 15261)

Many signals, such as speech and other animal sounds, FM radio waves, machine vibrations, and sonar and radar echoes, exhibit frequency characteristics that change over time. Standard spectral analysis of such signals provides an incomplete description of the process, because the spectral density reveals only what frequencies existed in the signal, but not when they occurred. Time-frequency analysis, however, shows how the frequencies change over time and hence is a more complete characterization of these signals. This approach to time-varying spectral analysis has become standard and has revealed new physical properties of signals. Perhaps the most common method of time-frequency analysis is the spectrogram, developed over 50 years ago. Many other methods have since been

developed, driven largely by a desire to overcome limitations of the spectrogram. Most of these methods were developed by effective use of Cohen's 1966 formulation for generating time-frequency distributions (TFDs). A review of the spectrogram and these TFDs, including the Wigner, Choi-Williams, and Zhao-Atlas-Marks distributions, among others, is presented. A variety of applications are shown that illustrate the basic ideas and the different methods. [Work supported by ONR grant N00014-98-1-0680.]

9:30

**2aAB4. Speech recognition meets bird song: A comparison of statistics-based and template-based techniques.** Sven E. Anderson (Dept. of Computer Sci., Univ. of North Dakota, Grand Forks, ND 58202-9015, anderson@cs.und.edu)

Pattern recognition technology that has been developed for recognizing units of human speech can often be adapted for both recognition and analysis of animal vocalizations. This paper discusses two types of speech recognition algorithms, template based and statistics based, with respect to their ease of deployment and potential application to the objective, quantitative analysis of animal vocalizations. Implementations of the two types of algorithms have been compared using a large database of song units recorded from two song bird species. The algorithms exhibit different strengths and weaknesses. The template-based dynamic time-warping algorithm provides quantitative sound comparisons that are directly useful to a researcher, but selection of training materials depends on expert knowledge. The statistics-based hidden Markov model algorithm requires more training data, but usually performs better in noisy environments and with more variable vocalizations. While both algorithms are accurate in restricted domains, recognition performance could be improved if it were based on species-specific features extracted from the acoustic input. [Work supported by NIH 1-F32-MH10525 and ARO DACA88-95-C-0016.]

10:00

**2aAB5. Application of current speech recognition technology to nonstandard domains.** Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

The typical input to automatic speech recognition (ASR) algorithms is a word-length or longer acoustic waveform. The typical output consists of names of items identified from stored vocabularies. Most algorithms also use scoring procedures to find the top matches between the input and the stored information. These evaluation scores, probabilities or distance metrics, may also be output. There are many nonstandard applications of ASR technology that combine identification and evaluation scores. Examples to be discussed include speech training to improve intelligibility, spoken language proficiency of non-native speakers, and language training for adults with developmental disabilities. Most current ASR algorithms are fine tuned for species-specific properties of human speech and language. In addition, many incorporate psychophysical properties of the human auditory system in the initial signal processing. However, the underlying pattern recognition algorithms are also applicable to a wide range of animal vocalizations. Some commercial and laboratory systems will be discussed in relation to nonstandard applications of ASR. [Work supported by NIH, RO1 DC-02229, R44 DC-02213, and R43 HD-35425.]

10:30-10:45 Break

### Contributed Papers

10:45

**2aAB6. Technique for the generation and frequency compensation of bandlimited white noise and its application in studies of masked hearing thresholds.** James J. Finneran, Donald A. Carder, Sam H. Ridgway (SPAWAR Systems Ctr. San Diego PL-BS, Div. D35, 49620 Beluga Rd., San Diego, CA 92152-6266, finneran@spawar.navy.mil), and Carolyn E. Schlundt (Sci. Applications Intl. Corp., San Diego, CA 92110)

Masking noise is often used in hearing tests to create a floor effect in the presence of ambient noise or to examine specific features of the auditory system (e.g., the critical bandwidth). One of the chief requirements of the masking noise is that it possess a flat frequency spectrum within some user-defined bandwidth. Generation of suitable masking noise is complicated by the frequency response of the sound projector, which may possess a frequency-dependent transmitting sensitivity and/or exhibit resonances within the desired frequency range. At low frequencies acoustic standing waves may also alter the noise frequency spectrum. To overcome these limitations, a technique has been developed to generate bandlimited noise whose frequency content is compensated in order to flatten peaks or valleys in the measured frequency spectrum. Compensation is performed by passing white noise through a digital filter whose coefficients are determined from previous measurements of the acoustic system frequency response. The system has been implemented using a personal computer with commercial hardware and custom software. The method has been

used to quickly generate bandlimited Gaussian and uniform white noise for studies of masked underwater hearing thresholds in marine mammals. [Work supported by ONR and the NRC Research Associateship Program.]

11:00

**2aAB7. Improved signal processing techniques for measurement of the inter-pulse interval (IPI) of sperm whale clicks.** Thomas J. Hayward (Naval Res. Lab., Washington, DC 20375) and G. Pavan (Università degli Studi di Pavia, Via Taramelli 24, 27100 Pavia, Italy)

Improved signal processing techniques are described for estimating the inter-pulse interval (IPI) associated with the multiple pulses in a sperm whale click. These pulses are believed to result from repeated reflections of the original pulse within the spermaceti organ in the whale's head. Previous estimates of the IPI have been based on time series correlation or on computation of the cepstrum of the received waveform [J. Goold, J. Acoust. Soc. Am. **100**, 3431-3441 (1997)]. The present work describes real-time computation and display of the spectrogram and cepstrogram (time-cepstral distribution) [G. Pavan, Proc. IoA **19** (1997)] and examines the benefits of the simultaneous display of this information for field analysis and for browsing long recordings to identify sperm whale clicks. In addition, a detailed model of the spectral modulation associated with the multiple pulses is developed and leads to improved estimation of the IPI. Examples are presented that indicate the method may be more robust with

respect to additive noise and the effects of surface reflection than the cepstrum-based approaches. [Work of the first author supported by ONR. Work of the second author supported by the University of Pavia, Italy, the Italian Ministry of the Environment, and by the Italian Navy.]

11:15

**2aAB8. Testing the acoustic adaptation hypothesis for eastern and spotted towhee song.** Hitesh Khanna (Dept. of EEOB, Ohio State Univ., Columbus, OH 43210)

The acoustic adaptation hypothesis predicts that signals adapted for long-distance communication in forested habitats should possess low rates of amplitude modulation (slow trills), whereas signals adapted to open habitats should possess high rates of amplitude modulation (fast trills). The songs of two songbirds, the eastern and spotted towhees, were used to test this hypothesis. These two species have only recently been recognized as distinct species. Differences in song are one of the reasons given for splitting the species. Eastern towhees occupy edges of deciduous forests and sing slow trills. Spotted towhees occupying open chaparral habitat sing fast trills. Songs of both species were broadcast in their native habitats, and recorded at several distances from the sound source. Spectrogram cross-correlation was used to quantify sound degradation with distance. The results support the predictions of the acoustic adaptation hypothesis. The slow eastern towhee trill degrades less than the fast spotted towhee trill in a forested habitat, whereas the fast spotted towhee trill degrades less than the slow eastern towhee trill in an open habitat.

11:30

**2aAB9. Non-Gaussian randomness outside and inside the bat's brain.**

Rolf Müller and Roman Kuc (Intelligent Sensors Lab, Elec. Eng. Dept., Yale Univ., New Haven, CT 06520-8284, rolf.mueller@yale.edu)

Reflector placement in the habitats of bats is pronouncedly random. Furthermore, the impulsive nature of reflection sequences generated by facets found, e.g., in foliage can endow the probability density function of the received echo amplitudes with heavy tails. Two useful features for classification of foliages, which may be utilized by bats as ubiquitous landmarks, are the extent to which the amplitude distribution deviates from Gaussian (its "peakiness") and an eventual nonstationary gain of the propagation channel. The signal processing in the bat's auditory system does not have direct access to the target's reflection sequence, however. Even the shortest echolocation pulses seen in bats are chirps with durations, which are nonnegligible compared to the duration of the reflection sequence and the temporal spacing of salient peaks inside of it. On reception, echoes are subjected to bandpass filtering and demodulation. The latter step prohibits recovery of the reflection sequence by means of pulse compression using a matched filter. Consequently, it is investigated how the demodulated filter bank output may be used directly as a substrate for target classification. Specifically, the possible role of encoding the waveforms in sparse spike trains and comparing binaural information is evaluated.

2a TUE. AM

TUESDAY MORNING, 2 NOVEMBER 1999

MCKINLEY ROOM, 7:55 A.M. TO 12:00 NOON

### Session 2aAO

## Acoustical Oceanography: Geoacoustic Propagation and Inversion

Kevin D. Heaney, Chair

*Science Applications International Corporation, 888 Prospect Street, Suite 201, La Jolla, California 92037*

Chair's Introduction—7:55

### Contributed Papers

8:00

**2aAO1. The perfectly matched layer for elastic waves in poroelastic media.** Yanqing Zeng (Dept. of Civil and Environ. Eng., Duke Univ., Durham, NC 27708, yz3@acpub.duke.edu), Jiangqi He, and Qinghuo Liu (Duke Univ., Durham, NC 27708)

The perfectly matched layer (PML) as a material absorbing boundary condition (ABC) was first introduced by Berenger for electromagnetic waves, and later developed by Chew and Liu for elastic waves. In the continuous limit, an interface between a regular medium and a fictitious, lossy PML medium can be made perfectly matched so that there is no reflection from the PML to the regular medium. This property is independent of the incidence angle and the frequency of the incoming waves. Consequently, the PML provides an ideal ABC for the truncation of the computational domain in numerical methods such as the finite-difference, finite-element, and pseudospectral time-domain methods. Numerical experiments show that this ABC can reduce the reflection to several orders of magnitude below the level of the previous ABCs. In this work, the PML is further extended to elastic waves in poroelastic media through the approach of complex coordinates for Biot's equations. This nonphysical ma-

terial is used as an ABC at the computational edge of a finite-difference algorithm to truncate unbounded media. Numerical results show that the PML ABC attenuates the outgoing waves effectively.

8:15

**2aAO2. Frame bulk and shear moduli of air and water saturated glass beads.** Masao Kimura (Dept. of Ocean Eng., Tokai Univ., 3-20-1 Orido, Shimizu, Shizuoka, 424-8610 Japan)

Are the values of the frame bulk and shear moduli of air saturated sediments, and those of water saturated sediments, the same or not the same? This problem is currently under debate. In this study, the frame bulk and shear moduli of air and water saturated glass beads with five different grain sizes are derived from the measured values of the longitudinal and shear wave velocities in these media. The grain sizes are 0.05, 0.1, 0.2, 0.4 and 0.8 mm. The longitudinal wave velocities are measured using piezoelectric transducers with the operating frequency of 11.8 kHz for dry sample, and 500 kHz for wet sample. The shear wave velocities are measured using bimorph type piezoelectric transducers with the operating

frequency of 3.5 kHz for both samples. The relationship between the values of these moduli in air and water saturated glass beads and the effect of grain size are investigated.

8:30

**2aAO3. Geoacoustic and signal gain modeling in a range-dependent, shallow-water environment.** Ilya Rozenfeld (Rensselaer Polytechnic Inst., Troy, NY 12180, rozeni@rpi.edu), William M. Carey (79 Whipoorwill Rd., Old Lyme, CT 06371), Peter G. Cable (BBN Systems and Technologies, New London, CT 06320), and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

The Acoustic Characterization Test III (ACTIII) was conducted in the Strait of Korea in 1995. One of the goals of the experiment was to determine limitations on spatial signal processing due to environmental variabilities. Measurements of transmission loss and array signal gain (ASG) provide quantitative estimates of fluctuations in the acoustic field. Analysis of observations is supported by using grab samples, historical cores, and seismic measurements to develop a geoacoustic model. The geoacoustic model along with sampled bathymetry and sound-speed profiles is used to generate simulations of an effective attenuation coefficient, which measures the rate of change of mean transmission loss with range. The introduction of nonlinear frequency dependence into sediment attenuation profiles produces good agreement between the simulations and measured data. The ASG is directly related to the spatial coherence function. By employing procedures we have developed previously for approximating the coherence function, we use the environmental model to compute ASG. This allows further comparisons with experimental data as well as a determination of validity limits for the procedures. [Work supported by ONR.]

8:45

**2aAO4. Newly observed seismoacoustic wave characteristics in disturbed water-covered sand.** Jacques R. Chamuel (Sonoquest Adv. Ultrason. Res., P.O. Box 81153, Wellesley Hills, MA 02481-0001)

Surprising new experimental results are presented revealing the characteristics of broadband transient seismoacoustic waves in water-covered naturally deposited sand and disturbed sand. In the naturally deposited sand, one compressional wave ( $\sim 1650$  m/s) and one shear wave ( $\sim 140$  m/s) were detected as expected. In the disturbed sand, the high-frequency ( $\sim 100$  kHz) fast compressional wave was highly attenuated, and a slow low-frequency ( $\sim 20$  kHz) compressional wave component was observed. In certain models, the slow low-frequency compressional wave component co-existed with the fast high-frequency compressional wave. The slow compressional wave velocity varied between 285 m/s and 210 m/s. The shear wave velocity and amplitude remained practically unchanged. Results from coarse and fine sand are compared. The findings are important for understanding the penetration and conversion of high-frequency acoustic waves in sandy seafloor critical for the detection of buried objects in littoral regions. [Work supported by ONR.]

9:00

**2aAO5. The dispersion and attenuation of body waves in a transversely isotropic porous medium including the effect of pore shape.** Yinbin Liu (Avadh Bhatia Phys. Lab., Dept. of Phys., Univ. of Alberta, Edmonton, AB T6G 2J1, Canada)

Based on Biot's theory of a general anisotropic porous medium [J. Appl. Phys. **33**, 1482–1498 (1962)] and the dynamic permeability formula given by Johnson *et al.* for an isotropic medium [J. Fluid Mech. **176**, 379–402 (1987)], we derived analytic expressions for the velocities and attenuations as functions of frequency for the quasi-P1 wave, quasi-P2 wave, quasi-SV wave, and SH wave in a transversely isotropic porous medium. Numerical results are given for one set of material parameters and compared with Schmitt's results [J. Acoust. Soc. Am. **86**, 2397–2421 (1989)] based on an alternative theoretical approach. The calculated results show that the phase velocities for quasi-P1 wave, quasi-SV wave,

and SH wave are not sensitive to the pore shape, and the attenuations for four kinds of waves and the phase velocity for quasi-P2 wave are sensitive to the pore shape in the region of near and above threshold frequency.

9:15

**2aAO6. Wave number sampling at short range.** Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

The self-starter is an efficient approach for solving geoacoustic inverse problems involving a vertical array of receivers located on the order of ten wavelengths from a source [R. J. Cederberg and M. D. Collins, J. Acoust. Soc. Am. **22**, 102–109 (1997)]. With this approach, accurate solutions can be obtained by sampling the wave number spectrum at approximately ten points. There have been unsubstantiated claims that the spectral solution provides similar efficiency when implemented with an approach that involves perturbing the integration contour off the real line [F. B. Jensen *et al.*, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994), pp. 231–240]. It is demonstrated that this quadrature scheme breaks down at short ranges and that the self-starter solution is about an order of magnitude faster. These conclusions are based on comparisons of the rational approximations associated with the self-starter and the quadrature scheme, comparisons of acoustic fields for particular problems, and a simple analysis of the spectral integral. These conclusions are not surprising since the rational approximation associated with the self-starter is based on the analytic evaluation of the spectral integral while the other rational approximation is based on numerical integration. [Work supported by ONR.]

9:30

**2aAO7. Geoacoustic inversion techniques applied to field data.** Joseph F. Lingeitch, Michael D. Collins, Dalcio K. Dacol, Michael Nicholas, and John S. Perkins (Naval Res. Lab., Washington, DC 20375)

Reliable and efficient techniques for solving geoacoustic inverse problems have been developed. A coordinate rotation technique can be used to identify the best resolved combinations of parameters for a given frequency and configuration of hardware [M. D. Collins and L. Fishman, J. Acoust. Soc. Am. **98**, 1637–1644 (1995)]. Problems involving a vertical array of receivers at short range can be solved in a few seconds with an implementation of the self starter that is based on the method of undetermined coefficients [D. K. Dacol and M. D. Collins, J. Acoust. Soc. Am. (accepted for publication)]. These techniques are applied to data that were obtained in the Straits of Florida during the KWIX '98 experiment. [Work supported by ONR.]

9:45

**2aAO8. Estimation of anisotropic sediment parameters.** Andrew J. Fredricks, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180), Michael D. Collins, and Joseph F. Lingeitch (Naval Res. Lab., Washington, DC 20375)

Parabolic equation techniques have recently been developed for solving geoacoustic inverse problems [R. J. Cederberg and M. D. Collins, IEEE J. Ocean Eng. **22**, 102–109 (1997)] and for handling propagating problems involving anisotropic elastic layers [A. J. Fredricks, J. Acoust. Soc. Am. **101**, 3182 (1997)]. These techniques are used to investigate inverse problems involving anisotropic elastic layers. Complications arise from the direction dependence of the two sediment sound speeds. For the particular case of transverse isotropy, both sound speeds need to be estimated in two different directions. A coordinate rotation technique [M. D. Collins and L. Fishman, J. Acoust. Soc. Am. **98**, 1637–1644 (1995)] will be employed to estimate the resolvability of the direction-dependent wave speeds for different experimental configurations. [Work supported by ONR.]



10:15

**2aAO9. A cross-relation matched field inversion for geoacoustic parameter estimation in shallow water.** Reza M. Dizaji (Dept. of Elec. and Computer Eng., P.O. Box 3055, Univ. of Victoria, Victoria, BC V8W 3P6, Canada), N. Ross Chapman, and R. Lynn Kirilin (Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

In this paper the application of a matched field (MF) processor for geoacoustic parameter estimation from a signal of opportunity such as the broadband random noise radiated by passing ships is considered. A novel cross-relation (CR) based matched field processor is introduced for the purpose of ocean tomography based on broadband random source. This class of MF processors is based on the cross-relation property of sensor outputs at an array and their corresponding transfer functions from the true source location to the array. The processors are developed for nonstationary (NS), and wide sense stationary (WSS) random signals. For each formulation, two processors are proposed, a self-CR and a cross-CR. The performance of the proposed MF processors for environmental parameter estimation is demonstrated for real ocean environments using data collected by a 16-element vertical line array during one of the experimental tracks from the Pacific Shelf experiment that was carried out in shallow water off the west coast of Vancouver Island in the Northeast Pacific Ocean. The high-resolution property makes the cross-CR processor an excellent candidate for inversion for model parameters of the ocean waveguide. The replica or modeled fields are calculated using the normal mode model, ORCA.

10:30

**2aAO10. Geoacoustic inversions in a very shallow water environment.** K. M. McArthur and W. S. Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr., San Diego, CA 92093-0701)

The genetic algorithm is used to invert the geoacoustic parameters of a near shore, very shallow water region (from 10-m to 20-m water depth). The experimental data are source tows in the frequency range from 70 Hz to 700 Hz recorded on a horizontal line array of a 120-m aperture. The source tows were conducted offshore of Camp Pendleton, north of Ocean-side, CA, along range-independent and range-dependent bathymetry tracks. Generally, a thin layer of sediment overlies a harder subbottom in this region. An echosounding survey, a sediment coring survey, and regional geologic information are used to develop background geoacoustic models. Inversion results from transmission loss versus range data agree well with the results obtained from inverting complex-pressure data across the array for a single range.

10:45

**2aAO11. Results of environment inversion using modes extracted from vertical line array data.** Tracianne B. Neilsen and Evan K. Westwood (Appl. Res. Lab., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, neilsen@arlut.utexas.edu)

The results are presented for geoacoustic inversion using the depth-dependent normal modes extracted from vertical line array (VLA) data. The normal modes for multiple frequencies are obtained by performing a singular value decomposition (SVD) on VLA data recorded as a source moves outward in range. The data-extracted mode functions are then used to invert for the environmental parameters. The Levenberg–Marquardt nonlinear optimization method is used to adjust the environmental parameters to find the best fit between the data-extracted modes and modes modeled by the ORCA normal mode model. The method is useful for determining the sound speed profile and the characteristics of the uppermost bottom layers [T. B. Neilsen and E. K. Westwood, *J. Acoust. Soc. Am.* **104**, 1741 (1998)]. Mode extraction and geoacoustic inversion results for a range-independent track during the ACTII experiment will be pre-

sented. Extension of the theory for mode extraction (and preliminary results for inversion using simulated data) in a range-dependent environment using the adiabatic approximation will also be presented. [Work supported by ONR.]

11:00

**2aAO12. Sediment tomography in the Middle Atlantic Bight.** Gopu R. Potty and James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882)

Sediment properties such as compressional speed and attenuation have been extracted using the modal dispersion of explosive signals in the Middle Atlantic Bight. Data collected during the Shelf Break PRIMER experiment are used to generate sediment sound speed profiles along various propagation paths. Compressional speeds are inverted using traveltimes of modes 1 to 9 from 10 Hz to 225 Hz. Compressional attenuation is inverted using modal amplitude ratios. Gravity cores taken at the experimental site give direct measurements of compressional speed and attenuation for comparison. *A posteriori* estimates of resolution are evaluated. Nonlinear inversion techniques based on a Genetic Algorithm and Levenberg–Marquardt methods are used. Methods for increasing the efficiency of the inversion algorithm are also investigated. [Work supported by ONR.]

11:15

**2aAO13. Geoacoustic parameter estimation using a back-wave propagation technique.** Reza M. Dizaji (Dept. of Electron. and Computer Eng., P.O. Box 3055, Univ. of Victoria, Victoria, BC V8W 3P6, Canada), N. Ross Chapman, and R. Lynn Kirilin (Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

The concept of back-wave propagation is developed as an inversion method to estimate ocean geoacoustic parameters where the source location is known. A phase-regulated technique is introduced to increase the sensitivity of the method for low-sensitive geoacoustic model parameters. In this procedure, a sensitivity factor is varied to enhance the phase changes due to model and environmental mismatch. It is shown that the spatial resolution of signal-energy concentrated at the true source location is increased when the sensitivity factor increases. This leads us to define a criterion based on spatial variance of signal energy around the true source location. This technique is applied to the real data from the Pacific Shelf experiment that was carried out in the shallow water off the West Coast of Vancouver Island in the Northeast Pacific Ocean. The inversion for estimating three ocean parameters including water depth, compressional speed of the first sediment layer, and sediment density is demonstrated. A way to reduce the three-dimensional search to three one-dimensional searches is proposed by exploiting the fact that the pressure field has different sensitivity with respect to these parameters. The replica or modeled fields used in this section are calculated using ORCA.

11:30

**2aAO14. Use of surfseisms for determining near-surf-zone properties.** Gerald L. D’Spain, K. Megan McArthur, Grant B. Deane, and W. Kendall Melville (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

Geophone array recordings of “surfseisms,” i.e., seismic interface waves created by breaking waves in the surf zone, show that at times, a breaking wave creates a pair of arrivals. The first arrival has peak spectral levels between 10 and 15 Hz and rolls off sharply above this peak, and the second surfseism arrival occurs 5–15 s later with broader-band energy concentrated above 20 Hz. The first arrival is created offshore when wave breaking has a plunging character, rather than spilling, whereas the second is present during both breaking types and apparently signifies the final dissipation of the turbulent wave bore on the beach face. The time interval between arrivals measures the distance offshore that plunging occurs, their relative amplitudes is an indicator of the type of wave breaking, and their absolute amplitudes measure the breaking wave forces interacting with the ocean bottom. In addition, the frequency of the first arrival’s spectral peak

is a sensitive function of the thickness of the unconsolidated sand layer overlying wave-cut terrace material, indicating that surfseisms are naturally occurring probes of sand layer thickness. [Work supported by ONR, Code 32.]

11:45

**2aAO15. Estimation of bottom parameters in shallow water using interference pattern in frequency domain.** Boris Katsnelson, Valery Grigoryev (Phys. Dept., Voronezh Univ., Universitetskaya sq.1, Voronezh, 394693 Russia, katz@mph.vsu.ru), and Valery Petnikov (General Phys. Inst., Vavilova str.38, Moscow, 117333 Russia, petniko@az.gpi.ru)

A method of data inversion for geoacoustic parameters of bottom using FM acoustic signals is presented. The experiment was conducted at the stationary track between two ships. The source radiated linearly modulated signals within the band 25–95 Hz. Frequency spectra of received signals

are taken as input data for application of the Matched Field Tomography (MFT) method. For estimation of bottom parameters we compare the experimental and model interference patterns in frequency domain. A new element in our approach (for MFT-method) is modeling of the bottom using a two-components model: water plus mineral particles. The main parameter (used as one of the matching parameters), i.e., the characterizing sediment is porosity, sound speed and density of bottom can be expressed through porosity. Two mechanisms of bottom losses are considered: literally absorption-transformation of sound energy into heat (linearly depending on frequency), and volume scattering within bottom (depending on frequency as power function with the degree determined by dominating types of scatterers). As a result of MFT bottom parameters are estimated. It is shown that on the low-frequency (less than 40 Hz) bottom absorption dominates on the other side, for more the high-frequency loss in sediment is conditioned by volume scattering. [Work supported by RFBR, Grant 97-05-64878.]

TUESDAY MORNING, 2 NOVEMBER 1999

KNOX ROOM, 8:30 A.M. TO 12:05 P.M.

### Session 2aBB

## Biomedical Ultrasound/Bioresponse to Vibration and Signal Processing in Acoustics: Image and Signal Processing in Biomedical Ultrasound

Shira L. Broschat, Cochair

*School of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington 99164-2752*

Hua Lee, Cochair

*Department of Electrical Engineering, University of California, Santa Barbara, California 93106-2991*

### Invited Papers

8:30

**2aBB1. A two-dimensional array system for studies of ultrasonic imaging with aberration correction.** Robert C. Waag, Daniel B. Phillips, James C. Lacefield, Carsten G. Draeger, Feng Lin, and Makoto Tabei (Ultrasound Res. Lab., P.O. Box 648, Univ. of Rochester Med. Ctr., 601 Elmwood Ave., Rochester, NY 14642, waag@ece.rochester.edu)

A two-dimensional array system is described for pulse-echo studies of aberration correction. The transducer array is an 80×80 array with a center frequency of 3.0 MHz and a –6-dB bandwidth of 56%. At the center frequency, each element has a physical size of 1.04 wavelength and spacing of 1.2 wavelength. A multiplexer accesses any contiguous 128 elements for transmission and any contiguous 16 elements for simultaneous reception. Transmit electronics have independently programmable waveforms. Each receive channel includes a 20-MHz, 12-bit A/D converter, and a time varied gain programmable over 40 dB. Transmit and receive apertures up to the size of the array are formed synthetically. A method that iteratively predistorts transmit waveforms to produce a transmit focus compensated for aberration has been implemented. Point-spread functions have been measured for propagation through a water path and through a tissue-mimicking aberration path. Pulse-echo images have been formed through a water path, through a tissue-mimicking aberrator, and through the aberrator using aberration correction that consists of time-shift compensation in the transmit–receive aperture or backpropagation followed by time-shift compensation. The system is useful for pulse-echo measurements of aberration, development of adaptive focusing techniques, and formation of high-resolution ultrasonic images using aberration correction.

8:55

**2aBB2. Novel breast imaging techniques using transmission ultrasound.** Michael P Andre, Helmar S. Janee, Linda K. Olson (SDVA Dept. of Radiol. 114, Univ. of California, San Diego, San Diego, CA 92161, mandre@ucsd.edu), Constance D. Lehman (Univ. of Washington, Seattle, WA), and Barbara A. Fecht (Adv. Diagnostics, Inc., Richland, WA 99352)

Despite several limitations, ultrasound is an important adjunctive modality for detection and management of breast cancer, but it is not recommended for screening. We have explored novel approaches to breast imaging using transmission ultrasound to provide very large image fields, differentiation of breast masses, and improved detection of microcalcifications without the use of ionizing radiation or compression. One method uses a circular array and an iterative Born technique to reconstruct scatter properties in a series of coronal slices through the breast. The second method is based on acoustical holography in which an interference pattern is formed by combining a transmitted field with a reference beam. The signal is converted to the optical domain using a coherent laser detector. This latter method provides real-time images in a large fluoroscopy-like format but with pronounced soft tissue edge definition. A set

of large acoustic lenses provides zoom and focal plane control for magnified imaging with enhanced resolution and depth. Both systems have undergone preliminary investigation in large groups of patients with known abnormalities to assess their suitability for detection and characterization of breast disease and to provide accurate guidance for biopsy or tumor ablation. Results of laboratory and clinical findings will be presented.

9:20

**2aBB3. Recent progress on nonuniform fast Fourier transform algorithms and their applications.** Q. H. Liu, B. Tian, X. Xu, and Z. Q. Zhang (Dept. of Elec. Eng., Duke Univ., Durham, NC 27708-0291)

Recently, nonuniform fast Fourier transform (NUFFT) algorithms have received significant attention [Dutt and Rokhlin, *SIAM J. Sci. Stat. Comput.* **14**, 1368–1393 (1993)]. Unlike the regular fast Fourier transform (FFT) algorithms, the NUFFT algorithms allow the data to be sampled nonuniformly. The leading order of the number of arithmetic operations for these NUFFT algorithms is  $O(N \log_2 N)$ . Here, we review the recent progress of the NUFFT algorithms using the regular Fourier matrices and conjugate-gradient method for the forward and inverse NUFFT algorithms [Liu and Nguyen, *IEEE Microwave Guid. Wave Lett.* **8**, 18–20 (1998); Liu and Tang, *Electron. Lett.* **34**, 1913–1914 (1998)]. Because of their least-square errors, these NUFFT algorithms are about one order of magnitude more accurate than the previous algorithms. These NUFFT algorithms have been applied to develop the nonuniform fast Hankel transform (NUFHT) and nonuniform fast cosine transform (NUFCT) algorithms. Both NUFFT and NUFHT algorithms have been used to solve integral equations in computational electromagnetics and acoustics; The NUFCT has been used to solve time-dependent wave equations. Numerical examples will demonstrate the efficiency of the fast transform algorithms, and the applications in computational electromagnetics and computational acoustics.

9:45

**2aBB4. Noncontact ultrasonic imaging for the evaluation of thermal injury.** Joie P. Jones (Dept. of Radiological Sci., Univ. of California—Irvine, Irvine, CA 92697-5000)

Although conventional wisdom suggests that ultrasonic imaging of the body cannot be accomplished without direct contact (or at least via water coupling), we have shown that noncontact imaging through air is possible, certainly for superficial body regions, provided judicious choices of piezoelectric materials and matching layers are made. In preliminary experiments and clinical studies reported here, noncontact imaging is demonstrated for the evaluation of thermal injury (including the quantitative measurement of burn-depth), for the assessment of wound healing, and for the examination of assorted skin lesions. Specifically, in the case of thermal injury, reflections from the dermal/fat interface in human skin is clearly seen using a noncontact 5-MHz transducer. Such measurements are sufficient to determine burn-depth which, in turn, are sufficient to provide, for the first time, a quantitative and noninvasive method for burn evaluation and treatment specification. Evaluating over 500 burn sites in some 100 patients, noncontact ultrasound showed a much greater accuracy and sensitivity than standard clinical assessment. Our method is applicable to a conventional clinical environment as well as a battlefield situation and should prove particularly effective for large-scale medical triage.

10:10–10:20 Break

### Contributed Papers

10:20

**2aBB5. Focus-directed processing of acoustic holography images.** Ruming Yin, Shira L. Broschat (School of Elec. Eng. and Computer Sci., Washington State Univ., P.O. Box 642752, Pullman, WA 99164-2752), and Patrick J. Flynn (The Ohio State Univ., Columbus, OH 43210-1272)

Acoustic holography is a transmission mode imaging technique which was first proposed in the 1970s. As with optical holography, an image is obtained using coherent interference of the transmitted acoustic signal with a reference signal. The interference pattern is illuminated with a laser, and the resulting image is digitized. However, since image reconstruction is performed optically, acoustic holography introduces a focusing problem characteristic of optical systems. Ideally, an image is focused on tissues at a given depth along the optical axis—that is, in one planar slice of the object. In practice, the image is focused over a range of depths so that objects at different depths are blurred but still visible. In this paper, we consider several postprocessing algorithms developed to improve images obtained using an acoustic holography system. First, a focus measure technique is used to determine when the object of interest is best focused. Second, a technique called depth from focus is used to determine the depth of an object. Third, a technique is developed to increase the “in focus” interval, or focusing range. These techniques will be discussed and imaging results will be presented. [Work supported by the National Science Foundation and the Carl M. Hansen Foundation.]

10:35

**2aBB6. A new  $k$ -space method for simulation of ultrasonic propagation in tissue.** T. Douglas Mast (Appl. Res. Lab., Penn State Univ., University Park, PA 16802, mast@sabine.acs.psu.edu), D.-L. Donald Liu (Siemens Medical Systems, Issaquah, WA 98027), Laurent P. Souriau, Adrian I. Nachman, and Robert C. Waag (Univ. of Rochester, Rochester, NY 14642)

A new  $k$ -space method for large-scale computations of ultrasonic propagation is presented. In the new method, spatial derivatives from the second-order acoustic wave equation for inhomogeneous media are evaluated by Fourier transformation. Solutions are advanced in time using a  $k-t$  space Green's function. Computational results indicate that the new method shares advantages of both past  $k$ -space and pseudospectral methods. For scatterers with properties similar to soft tissue, the  $k$ -space method provides much higher accuracy and lower computational cost than a 2–4 finite-difference time domain method. The  $k$ -space method also allows high accuracy to be obtained for time steps much larger than those required by a leapfrog pseudospectral method. The low dispersion inherent to the  $k$ -space method is illustrated by large-scale quasi-one-dimensional computations, in which pulse waveforms incur negligible shape change for propagation distances as large as 1000 wavelengths. Example applications of the  $k$ -space method are demonstrated, including simulation of propagation through a large-scale tissue cross-sectional model and incorporation of a  $k$ -space solver into a nonlinear inverse scattering method employing eigenfunctions of the far-field scattering operator.

10:50

**2aBB7. Calibration of a two-dimensional array system for ultrasonic aberration correction.** James C. Laceyfield, Daniel B. Phillips, and Robert C. Waag (Dept. of Elec. and Computer Eng., Univ. of Rochester, Rochester, NY 14627-0126)

Two procedures for electronic compensation of array spatial and temporal nonidealities are compared. The motivations for array calibration are to account for signal variations due to differences in the impulse responses of the elements and to reduce beam degradation caused by nonideal element directivities. Both methods calculate inverse filters to equalize the signal at each element. One approach equalizes the array response to a planar reflector, while the other equalizes the response at a specified focal point. The point calibration method reduces the standard deviation of arrival time fluctuations in the measured wave front from a point reflector from 21 to 2 ns and reduces the standard deviation of energy level fluctuations from 2.5 to 1.7 dB. However, the point method also diverts more energy outside the main peak of the focused beam, which causes the -10-dB peripheral energy ratio to increase from 0.30 to 0.32. Point calibration of the receive aperture is nevertheless desirable for aberration correction using backpropagation followed by time-shift compensation because the fidelity of the correlation algorithm is dependent upon unbiased measurement of the echo wave front at the receive aperture.

11:05

**2aBB8. Comparison of high-frame rate and delay-and-sum imaging methods.** Jian-yu Lu and Anjun Liu (Ultrasound Lab., Dept. of Bioengineering, The Univ. of Toledo, Toledo, OH 43606, jilu@eng.utoledo.edu)

Recently, a high-frame rate imaging method has been developed with limited diffraction beams to construct either two-dimensional (2-D) or three-dimensional (3-D) images (up to 3750 frames or volumes/s for biological soft tissues at a depth of about 200 mm). In this talk, the new method is compared with the conventional delay-and-sum (dynamic focusing) method. Both computer simulation and experiment results show that the quality of images constructed with the two methods are virtually identical when the maximum Axicon angle of X waves in the high-frame rate method is approaching to 90 degrees. Theoretical analysis is carried out to confirm the results. This is significant because the high-frame rate method requires thousands of times less computations while achieving the same high-imaging quality as the conventional delay-and-sum method. [This work was supported in part by grant HL 60301 from the National Institutes of Health.]

11:20

**2aBB9. High-frequency dependence of the backscatter coefficient on selected bovine tissues from 10–30 MHz.** Subha Maruvada, Kirk K. Shung, and Shyh-Hau Wang (Grad. Prog. in Acoust. and The Bioengineering Prog., University Park, PA 16802)

Very high-frequency diagnostic ultrasonic imaging operates at frequencies of 20 MHz and higher. Thus it is critical to obtain data on ultrasonic attenuation and scattering in this frequency range. At high frequencies, it is not feasible to make scattering measurements with unfocused transducers due to their decreased sensitivity, therefore focused transducers are needed. Using the standard substitution method to calculate the backscatter coefficient, as is used with unfocused transducers, yields erroneous results for focused transducers. The assumption that the reflected echo from a perfect reflector in the far field can be calculated as though the transducer acted like a point source is not valid for focused transducers. A method is presented for focused transducers where the flat

reflector is substituted by a particulate reference medium whose backscatter coefficient is well known and documented, in this case, a red cell suspension. Results between focused and unfocused transducers match closely between 10 and 20 MHz. The backscatter coefficient for bovine tissues has been well documented between 1 and 10 MHz. These measurements have been extended to 30 MHz. The frequency dependence of backscatter on bovine tissues will be presented in the range 10–30 MHz and compared to previous results.

11:35

**2aBB10. Assessing arterial stenoses by tracking turbulence with Doppler ultrasound.** Megan M. Miller, Christy K. Holland (Dept. of Radiol., M.L. 0742, Univ. of Cincinnati, Cincinnati, OH 45219-2316, Christy.Holland@uc.edu), and Peter J. Disimile (Univ. of Cincinnati, Cincinnati, OH 45267-0700)

When a stenosis, or narrowing, causes a significant area reduction of a blood vessel, turbulence in the post-stenotic jet can be detected with Doppler ultrasound distal to the stenosis. A technique is investigated for assessing the severity of a stenosis, defined as the pressure drop across the lesion, by extracting the streamwise turbulence intensity (or the normalized square root of the velocity variance) from the Doppler ultrasound in an arterial flow model. The model consists of an optically and acoustically transparent polyurethane tube that mimics femoral artery compliance, a pump capable of continuous and pulsatile flow, 10- $\mu$ m glass spheres as an ultrasound and laser scatterer, and both blunt and rounded inlet stenoses. The flow field through three axisymmetric, Plexiglas stenoses with diameter reductions from 60% to 95% mH were investigated with an ATL HDI 3000 using the L7-4 linear array in Doppler mode (1.0-mm spatial pulse length) and a Dantec two-color 55X laser Doppler anemometer (0.7-mm major axis). To validate the ultrasound technique, correlation of Doppler ultrasound and laser Doppler anemometry flow measurements was examined. The correlation of the peak velocity, the maximum turbulence intensity, and the pressure drop across each stenosis was also investigated. This Doppler ultrasound technique could be sensitive to more subtle alterations in hemodynamics and therefore could aid early detection of atherosclerosis.

11:50

**2aBB11. Backward propagation algorithms for image reconstruction: Signal processing, algorithm architecture, and applications.** Hua Lee (Dept. of Elec. and Computer Eng., Univ. of California, Santa Barbara, CA 93106)

Traditionally, image reconstruction algorithms were developed, following the design and configurations of the data-acquisition systems. So, the algorithms are typically special-purpose and system-specific. Consequently, there is a lack of consistency in terms of algorithm architecture, organization, and performance. In this paper, we present a unified framework for algorithm design and development. This allows us to implement image reconstruction for various data-acquisition configurations including active or passive sensing, linear or circular receiving apertures, CW or wideband illumination, and monostatic or bistatic formats, based on a single theoretical framework in an organized manner. The computation schemes for both linear and circular apertures will be discussed in detail. The layered backward propagation technique, as the main processing modality, provides the flexibility for dynamic updating for changes of propagation parameters. In addition, we illustrate parallel processing and recognition as integrated components for the algorithm structure. The presentation includes the theoretical background on signal processing for image formation, and overview of algorithm architecture for various configurations, a discussion on computation complexity and commonality, and several applications.



## Session 2aEA

## Engineering Acoustics: Compatibility of Hearing Aids and Cellular Telephones: Standards Progress

Stephen C. Thompson, Chair

*Knowles Electronics, Inc., 1151 Maplewood Drive, Itasca, Illinois 60143*

Chair's Introduction—8:30

*Invited Papers*

8:35

**2aEA1. Suitcase lab for measuring digital cellphone interference in hearing aids.** Mead C. Killion (Etymotic Res., Inc., 61 Martin Ln., Elk Grove Village, IL 60007) and Harry Teder (Consultant to Hearing Industries Assoc., Excelsior, MN 55331)

A low-cost, "real-life" method for measuring the interference caused by cellular telephones in hearing aids is proposed. Data would be valid for specific phone and hearing aid models. Estimated equipment cost is about \$500. Real-ear recordings will demonstrate the audible interference.

9:05

**2aEA2. Real-ear measurement of hearing-aid interference from digital wireless telephones.** Harry Levitt (Ctr. for Res. in Speech & Hearing Sci., City Univ. of New York, 33 W. 42nd St., New York, NY 10036), Judy Harkins, Linda Kozma-Spytek (Gallaudet Univ., Washington, DC 20002), and Eddy Yeung (City Univ. of New York, New York, NY 10036)

Two studies were performed. The first measured hearing-aid interference under field conditions. Signal levels in the ear canal were measured using an in-the-canal probe system, the output of which was stored digitally in a laptop computer. Both bystander and user interference were measured as well as the user's ratings of intelligibility, annoyance, and usability (of the digital wireless telephone). The levels of user-interference were found to be unacceptably high for the large majority of hearing-aid wearers. In contrast, bystander interference was at a relatively low level under conditions typical of telephone use. The second experiment controlled the level of interference by means of the test mode in the digital wireless telephone. Speech-to-interference ratios were measured in the ear canal for various levels of intelligibility and usability, as rated by the subjects. Data were obtained for 37 subjects on each of three digital wireless technologies (PC1900, TDMA, CDMA). [Research supported by NIDRR.]

9:35

**2aEA3. Wireless hearing aid compatibility—handset testing requirements.** H. Stephen Berger (Siemens Information and Communication Products LLC, 2205 Grand Ave. Pkwy., Austin, TX 78728, stephen.berger@icp.siemens.com)

The Hearing Industry and Wireless Industry has been working together since 1996 to develop a standard to resolve this issue. This paper reports on the technical challenges and the testing requirements for wireless handsets. It concludes with a brief discussion of the need for consumer education and involvement in order to achieve the optimal solution to this issue. Hearing aids operate in two modes. In their primary mode a microphone is used to sense acoustic signals. The secondary mode uses an inductive coil, known as a T-Coil, to receive a magnetic signal which has been modulated by the audio signal. For the wireless handset, which may be used with a hearing aid, there are test requirements arising from both modes. The field strength of both the  $E$  and  $H$  field, in the area intended for use by a hearing aid must be measured and assured to be within specified parameters. For the T-Coil mode a magnetic signal, modulated by the audio, must be assured. However, in T-Coil mode there is also the issue of assuring that the signal has sufficient quality and does not suffer from magnetic interference from secondary sources.

10:05

**2aEA4. Hearing aid measurement and consumer counseling.** Thomas A. Victorian (Starkey Labs., Inc., 6700 Washington Ave. S., Minneapolis, MN 55344)

The relationship between digital cellular telephones and hearing aid devices requires specific measurement techniques to determine system compatibility between the two technologies. The ANSI C63.19 working group has developed such measurement techniques related to the actual usage of a digital phone while wearing a hearing aid—known as "near field" or "user" condition. The hearing aid rf near-field immunity test requirements involve hearing aid orientation and bias as well as rf near-field simulated test conditions. The testing is performed in both the audio and telephone-coil modes. The measurement figure of merit is defined as IRIL (input referred interference level) which normalizes the hearing aid acoustic gain. It can be shown that hearing aids will be required to withstand digital phone E-field emission levels greater than 50 dB V/m and H-field emission levels greater than  $-2$  dB A/m. By characterizing the simulated phone antenna properly this test enables hearing aid immunity testing within a controlled rf laboratory. The outcome of these measurements will be used to classify hearing aids for compatibility with categorized digital phones. This classification process will assist the hearing aid dispenser in patient counseling.

10:45

**2aEA5. Development of hearing-aid near-field EMI testing.** Marco Candiago (Unitron Industries, Ltd., 20 Beasley Dr., Kitchener, ON N2G 4X1, Canada)

With the introduction of digital cellular telephone technology, cellular telephone interference to hearing aids is now a noticeable problem for both hearing-impaired digital cellular phone users and bystanders. Previously, bystander hearing-aid immunity levels (far field) were defined by the IEC committee and are contained in the current IEC 118-13 hearing-aid electromagnetic immunity standard. With the realization that an up-scaled far-field test would be insufficient to provide an accurate correlation of digital cellular telephone usability, the ANSI C63.19 working group developed a near-field test procedure using calibrated dipole antennas. This paper discusses the development of near-field testing, broken down in the following topics: (1) digital cellular phone interference and spectral properties; (2) rf interference coupling into the hearing aid; (3) far-field testing and rationalization behind the use of 1-kHz 80% AM modulation to quantitatively determine hearing-aid immunity characteristics; (4) near-field simulation using dipoles and antenna calibration; (5) near-field repeatability testing; and (6) preliminary results of near-field hearing-aid interference measurements. The result of these developments has led directly to the formulation of the near-field test procedure described in Sec. 5 of the currently developing ANSI C63.19 standard.

11:15

**2aEA6. Validation of the C63.19 standard: Acoustic measurements and subjective evaluations.** Robert E. Schlegel, Hank Grant, and Tamy L. Fry (Wireless EMC Ctr., School of Industrial Eng., Univ. of Oklahoma, 100 E. Boyd, Ste. R-208, Norman, OK 73019, schlegel@ou.edu)

The ANSI C63.19 measurement and performance standard addresses the electromagnetic compatibility of hearing aids with wireless communication devices. Validation of the measurement procedures and performance criteria comprised a pilot study, objective validation, and subjective validation. Objective validation involved acoustic measurements of 80 combinations of hearing aids and wireless phones. The *E*-field and *H*-field immunity levels of each hearing aid and the *E*-field and *H*-field emissions of each wireless phone were measured according to the standard. These data were used to theoretically predict the overall input referenced interference level. Across a broad range of hearing aid immunity levels, the measured acoustic interference was accurately predicted from the individual C63.19 measurements. Subjective validation involved the testing of twenty hearing aid wearers with six phone technologies. Hearing aids offering a variety of immunity levels were custom manufactured for the participants and tested according to the C63.19 measurement procedures. While using each wireless phone, participants provided ratings of the speech intelligibility, usability, and annoyance caused by any rf interference. In addition to confirming the existence of compatible hearing aids and phones, the results of this study contributed to the setting of the immunity and emissions criteria for a usable hearing aid wireless phone system.

### *Contributed Paper*

11:45

**2aEA7. Clinical studies of digital wireless phone interference in hearing aids.** Robert E. Schlegel, Randa L. Shehab, Tamy L. Fry, and Hank Grant (Wireless EMC Ctr., School of Industrial Eng., Univ. of Oklahoma, 100 E. Boyd, Ste. R-208, Norman, OK 73019, schlegel@ou.edu)

Two clinical studies were conducted to evaluate the impact of electromagnetic interference from three digital wireless phone technologies on hearing aid wearers. Sixty-eight participants were tested with their own hearing aids to determine the interference detection threshold and annoyance, along with the impact on speech intelligibility. Bystander interference varied as a function of the hearing aid type and the phone technology.

The interference was significant only when the digital wireless phone was within 2 ft of the hearing aid user. Significant declines in speech intelligibility were observed as a function of the specific phones used. In the second study, interference signals for three phone technologies were generated at five sound pressure levels from 35 to 75 dB and mixed with speech at 65-dB SPL to test the speech intelligibility of 24 hearing-impaired participants. Scores for the TDMA-217 Hz phone signal at low speech-to-noise ratios (less than 10 dB) were significantly lower than those for CDMA and TDMA-50 Hz at the same level. Speech intelligibility scores at speech-to-noise ratios of 20 and 30 dB were similar to those for a "no noise" condition. The articulation index accurately represented the impact of the wireless phone interference on speech intelligibility.

## Session 2aED

## Education in Acoustics: Undergraduate Laboratory Experiments

Daniel A. Russell, Chair

*Science and Mathematics Department, Kettering University, 1700 West Third Avenue, Flint, Michigan 48504-4898*

Chair's Introduction—7:55

*Invited Papers*

8:00

**2aED1. Classical measurements of sound.** Elmer L. Hixson (Univ. of Texas, Austin, TX 78712)

From the 1500s the study of sound involved acoustics and vibrations. Early theoreticians tried to experimentally justify their analytical models. Many vibrating devices were available to generate acoustic waves but few means were available to quantify these sources and sound propagation. The human ear was the only means of amplitude measurement and spectral analysis. The development of electrical and electromechanical technology allowed great advancements in the measurement of sound. Bell's telephone transmitter/receiver and Flemming's valve brought great strides in quantifying sound. The capacitor microphone brought high accuracy. The advent of the oscilloscope and electrical filters provided analysis of complex sounds. This level of technology served the field well until the advent of the computer age.

8:30

**2aED2. Acoustics laboratory experiments for all.** Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809) and Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115)

For most students, understanding scientific principles requires hands-on experience, and that is the reason for having undergraduate laboratories. This is especially true of acoustics, in which the laboratory experience is important not only to beginning and advanced students in science and engineering but also to students taking a general education course. We will describe some favorite experiments on the introductory, intermediate, and advanced levels. Most of them are based on experiments in *Acoustics Laboratory Experiments*, a collection of 52 experiments on all levels. Some of them are appropriate for introductory courses in musical acoustics or speech and hearing, others are suitable for physics and engineering students in more advanced courses. Ways to incorporate acoustics experiments into courses not having a scheduled laboratory period will also be discussed.

9:00

**2aED3. A senior engineering design elective with industry and community-sponsored noise control projects.** Robert Celmer (Acoust. Prog. and Lab., College of Eng., Univ. of Hartford, 200 Bloomfield Ave., W. Hartford, CT 06117, celmer@mail.hartford.edu)

One of the transitions that engineering students must make as they enter the work force is the progression from theoretical concepts, or *book learning* to applied or *real world* applications. At the University of Hartford, engineering design courses serve as culminating experiences devised to bridge this passage. One such course, *Noise Control Design*, challenges the student to apply the past three years' conceptual base to a problem solving opportunity replete with actual scenarios encountered in industry or the community. Each year local firms and organizations approach the laboratory for assistance with a variety of sound or vibration problems. After an initial training period, students make use of the laboratory's dual channel FFT and real time analyzers, anechoic chamber, sound intensity and modal analysis software, acoustic modeling software, vibration tables and transducers, portable sound level meters and digital tape recorders. Using a consultant-client model, students work collaboratively in teams of two defining the problem, developing a method of approach, making appropriate measurements, devising alternate solutions, and ultimately delivering a written and oral presentation at the end of the semester. The paper discusses specific projects and some experiences students had with their first industrial assignment.

9:30–9:40 Break

*Contributed Papers*

9:40

**2aED4. Laboratory experiments as demos and projects in the underwater acoustics and sonar course.** Elizabeth L. Simmons and Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402, korman@nadn.navy.mil)

Underwater Acoustics and Sonar (SP411) is a 3-h course that is typically offered to Midshipmen in their senior year. General science majors take the course in the fall while the oceanography majors enroll in the

spring. A sprinkling of physics, electrical engineers, ocean engineers, and systems majors also populate the course (totaling ~110 students/yr). Since this course is offered without a lab, the "in-class" experience has been enhanced with the development (over many years) of our demo carts which surround the classroom. Although Friday is our major "demo day," demos are performed throughout the week. They motivate the students' "out-of-class" experimental projects. Demos include: (a) waves on slinkies; (b) Fourier analysis of tones in noise, homemade musical instruments; (c) harmonic synthesis; (d) receiver operating characteristics from pro-

cessed signals in noise; (e) two-element and loudspeaker beam patterns; (f) sound speed versus temperature in water; (g) target strength versus angle of a model sub; (h) Ref. coef. from an Al/water interface; (i) PC-IMAT (interactive multisensor analysis training) simulations of array steering, ray tracing, active sonar, propagation loss; and (j) FM detection and Doppler effects. Students get involved with the measurements, have fun, and their understanding of underwater sound is greatly enhanced.

9:55

**2aED5. The acoustic interferometer as a teaching laboratory experiment.** Thomas B. Gabrielson (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, tbg3@psu.edu)

Interferometry is a popular subject for teaching-laboratory experiments in optics. Furthermore, some important classes of acoustic and vibration sensors rely on optical interferometry. Consequently, the fundamentals of interferometry and the interpretation of the resulting signals are important concepts even outside the field of optics. Unfortunately, procurement, alignment, and maintenance of optical interferometers can be prohibitively expensive for a program that does not specialize in optics. For institutions lacking a substantial infrastructure in free-air or fiber optics, the acoustical interferometer is an attractive alternative. Most of the issues associated with optical systems can be incorporated into straightforward laboratory exercises. The components for a 40-kHz acoustical system are inexpensive and readily available. The alignment is greatly simplified by the much longer wavelength and the function of the components is more intuitive. Moreover, absolute displacement calibration can be done with conventional micrometer translation stages; the effects of different bias points can be investigated; and the characteristics of large-signal behavior in interferometers can be studied. For advanced experiments, a heterodyne interferometer can be constructed along with a simple demodulator. [Work supported by the Graduate Program in Acoustics at Penn State.]

10:10

**2aED6. Vortex-driven whistling tube undergraduate experiment.** Timothy W. Lancey and Tracy J. Haeggstrom (Dept. of Mech. Eng., California State Univ. Fullerton, Fullerton, CA 92834)

An experiment to investigate the tones produced by the generation and shedding of periodic vortices in a tube was designed, constructed, and implemented as a student project in a senior laboratory course. A 0.102-m-diam by 0.762-m-long lucite tube was installed in a subsonic wind tunnel; the longitudinal axis of the tube was located streamwise with the mean flow, laterally centered in the test section, 0.178 m above the floor. Two lucite rings, with minimal clearance at the tube inside diameter and each having inside diameters of 0.0762 m, were connected inside the tube. One of the rings was at the tube inlet plane, while the other was slightly downstream, at a location which could be varied. A microphone was attached to the tube near the exit plane and connected to a PC which incorporated a data acquisition system. Airflow through the wind tunnel was

varied until tones were heard, and the air velocities and frequencies of the tones were recorded. Results of the data analysis indicated that the frequency of vortex shedding, at the measured air velocity and ring separation, was in agreement with the natural frequency of the tube, for each of the audible tones.

10:25

**2aED7. Demonstration of Doppler effects and sound-speed measurement using an analog frequency-to-voltage (FM) detection circuit.** Angela K. Huegel, Michael S. Naff, and Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402, korman@nadn.navy.mil)

An audio frequency FM detector has been designed and built to measure very small Doppler shifts for classroom demonstrations. This detector uses a limiter, a series  $L-C$  phase-shift network, two sine-to-square wave circuits, a multiplier and low-pass filter. One can easily measure a deviation of 1/20 Hz in a 4000-Hz carrier. The apparatus consists of a stationary piezo-tweeter radiating a tone near a microphone that translates with periodic motion, using a motorized disk and linkage mechanism, which drives a translation stage. The output of the mic goes to the FM detector which outputs to a digital scope or meter for display. Calibration is extremely linear and is performed using the digital frequency readout on the generator. From a measure of the peak to peak Doppler shift  $\Delta f$  (from the periodic Doppler shift signal versus time), the source frequency  $f$ , the radius  $R$  of the disk, and the period  $T$ , one can predict the sound speed to be  $C = (f/\Delta f)(4\pi R/T)$ . Here, an approximation involving the linear velocity has been made. Results for  $C$  are typically within 1% of the expected value for air at room temperature.

10:40

**2aED8. New acoustics experiments on an old NeXT computer.** Daniel A. Russell (Sci. and Math Dept., Kettering Univ., Flint, MI 48504, drussell@kettering.edu)

The manufacture of NeXT computer hardware was discontinued in 1993. However, the operating system (now known as OPENSTEP and Macintosh OS-X) is still alive and kicking, and refurbished NeXT systems are still available at greatly reduced cost. Coupled with an A/D converter, the NeXT computer is a powerful, and relatively inexpensive, tool for an undergraduate acoustics laboratory. There is a fairly wide selection of free acoustics software available over the internet, including applications for digital sound recording, mixing, and editing; oscilloscopes; FFT analysis; spectrum analysis; MIDI sequencing; sound file conversion; and signal generation. In addition, there are some excellent new commercial software packages which allow the NeXT to act as a professional quality digital function generator and audiometer. This paper will summarize how NeXT computers are used at Kettering University for undergraduate acoustics laboratory experiments involving digital sound recording and manipulation, signal analysis, Fourier synthesis, principles of digital audio, noise, and human hearing.



## Session 2aMUa

## Musical Acoustics: Musical Instruments: Experiments and Modeling

Thomas D. Rossing, Chair

Physics Department, Northern Illinois University, DeKalb, Illinois 60115

## Invited Paper

8:30

**2aMUa1. Acoustics of Baltic psaltery.** Andres Peekna (Innovative Mech., Inc., 265 Coe Rd., Clarendon Hills, IL 60514-1029) and Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115)

The Baltic psaltery, a plucked-string instrument, is known in Finland as “kantele,” in Estonia as “kannel,” in Latvia as “kokle,” in Lithuania as “kankles,” and in northwestern Russia as the “wing-shaped gusli.” In its archaic form, it had a limited range, ~5–13 strings, usually tuned diatonically, though the lower strings were sometimes tuned as bourdons, several tones lower than the other strings. Results of experimental investigations on several instruments are presented. Methods used for calculating the lowest resonance of the air cavity are described. Investigations include determining the modes of vibration using electronic TV holography and recording sound spectra throughout the tonal range. Results indicate that the more successful instruments have many resonances close together in frequency, within their limited tonal range. This may account for why the archaic Baltic psalteries had several small soundholes rather than the single soundhole found on guitars and mandolins, which have much wider tonal ranges. Several instruments will be demonstrated.

## Contributed Papers

9:00

**2aMUa2. Asymmetrical properties of a vibrating wire and their effects.** Roger J. Hanson, H. Kent Macomber, and Andrew Morrison (Dept. of Phys., Univ. of Northern Iowa, Cedar Falls, IA 50614, hansonr@uni.edu)

Some effects of the asymmetries causing a splitting of the fundamental natural vibrating frequency of a wire have previously been reported [Hanson *et al.*, *J. Acoust. Soc. Am.* **103**, 2873 (1998)]. It has been demonstrated in this work on brass harpsichord wire that the splitting of the frequency is due to intrinsic properties of the wire itself and not of asymmetries in the end clamps. The two frequencies are associated with two definite orientations with respect to the wire. These two vibrational directions have been determined to be orthogonal within one degree experimental uncertainty. This orthogonality is in agreement with predictions of a simple model which assumes that, for small-amplitude free vibrations, the observed portion of the wire moves under the action of a linear anisotropic conservative restoring force. Measured splittings for several samples of harpsichord wire have ranged from 0.12 to 0.30 Hz for a frequency of about 70 Hz. The nonlinear effects of generation of motion perpendicular to the driving direction [Hanson *et al.*, *J. Acoust. Soc. Am.* **96**, 1549–1556 (1994)] and generation of higher harmonics are profoundly influenced by the orientation of the driving direction with respect to these vibrational orientations for low-driving forces. Related effects for plucked strings will be discussed.

9:15

**2aMUa3. Physical modeling of the piano: First results for a simple monochord.** N. Giordano (Dept. of Phys., Purdue Univ., West Lafayette, IN 47907)

Initial results are presented for a project aimed at constructing a physical model of the piano. Newton’s laws are used to calculate the motion of the hammer, strings, soundboard, and room air, and thereby compute the sound produced by the instrument from “first principles.” In this talk we focus on a prototype instrument consisting of a monochord with a small soundboard. Modeling results for this geometry are described, and com-

pared with experimental results for a laboratory version of the instrument. If possible, calculated tones will be presented. [Work supported by NSF Grant PHY-9722031.]

9:30

**2aMUa4. Sound production by a vibrating piano soundboard: Theory.** Minghui Jiang and N. Giordano (Dept. of Phys., Purdue Univ., West Lafayette, IN 47907)

We describe a theoretical study of the sound generated by a piano soundboard. The soundboard model, which is similar to that developed recently by our group, is driven harmonically at a point on one of the bridges, and the resulting vibrational motion obtained by numerical solution. This soundboard is situated in a numerical room, and the resulting room pressure is calculated with two different methods. In one approach the pressure throughout the room is obtained by solution of the appropriate coupled equations for the air pressure and velocity. In the other treatment every point on the soundboard is treated as a simple source, each of which contributes to the pressure at a specified listening location. The results of these two approaches are compared to each other, and to recent experiments by our group. [Work supported by NSF Grant PHY-9722031.]

9:45

**2aMUa5. Spectral characteristics and efficient critical-band-associated group synthesis of piano tones.** Hua Zheng and James W. Beauchamp (Dept. of Elec. and Computer Eng., Univ. of Illinois at Urbana-Champaign, 5237 Beckman Inst., 405 N. Mathews St., Urbana, IL 61801)

The physics of piano sound production has been studied extensively, while little has been done to explore its spectral characteristics for efficient and perceptually accurate digital synthesis. Piano tones with a large variety of explicit performance parameters were recorded from a MIDI-controlled acoustic piano, and spectral modeling methods were used for

the analysis. A number of features have been concluded: (1) relationship between harmonic structure and pitch for six levels of dynamic; (2) relationship between spectral envelope and dynamic for all pitches; (3) typical amplitude envelopes; (4) effects of the sustain and soft pedals; (5) inharmonicity characteristics; (6) noise characteristics. A multiple-wavetable synthesis algorithm implementing a group synthesis model was developed. Each harmonic group corresponds to one or more consecutive critical bands. Amplitude versus time envelopes of all harmonics within one group share a weighted-average shape with individual weight applied to each harmonic. With appropriate numbers of groups and characteristic noise, convincing results have been obtained for the entire range of pitches and dynamics. The synthesis algorithm has been implemented as a subroutine of an analysis/synthesis software package, as a Music 4C instrument, and as a real-time plug-in for a software sound server.

10:00

**2aMUa6. Acoustics of flute sound in laboratory experimental conditions.** Isabelle Cossette (Meakins-Christie Labs., McGill Univ., 3626 St-Urbain St., Montreal, QC H2X 2P2, Canada and NVC, Univ. of Sydney, NSW 2006, Australia) and C. William Thorpe (NVC, Univ. of Sydney, NSW 2006, Australia)

Experimental studies of musical performance are often performed in laboratories with sometimes invasive apparatus. A concern often raised by musicians is that recordings obtained in such laboratory conditions may not reflect what would occur in a real performance. This study aims to provide an assessment of the extent to which data acquired in an invasive experiment, involving the physiology of flute performance, represents what would occur in actual performance. Five professional flutists, playing on the same flute, performed four well-known pieces of the flute repertoire in invasive and noninvasive conditions. The invasive condition consisted of apparatus on the flute head joint to measure the jet velocity and lip aperture, and trans-nasal catheters to measure respiratory pressures. The noninvasive condition occurred before the apparatus had been attached, but with other conditions kept as consistent as possible. The acoustic output of the flute was recorded and characterized by measuring the fundamental frequency and relative magnitudes of the harmonics. The results did not show any significant change in the acoustic parameters overall, although several subjects exhibited some differences ( $<6$  dB) in harmonic amplitudes between the conditions. We therefore conclude that, in professional players at least, performance is not unduly compromised by experimental constraints.

10:15

**2aMUa7. A hybrid waveguide model of the transverse flute.** Mark A. Bartsch (Univ. of Dayton, Dept. of Elec. Eng., 300 College Park, Dayton, OH 45469, bartscma@flyernet.udayton.edu)

Recent years have seen substantial improvements in the modeling and synthesis of jet-reed instruments such as flutes and organ pipes. Developments in the modeling of woodwinds using so-called digital waveguides have allowed the efficient acoustical modeling of the main body and toneholes of the instrument. Further, a new and more complete model of the jet's behavior in the presence of the edge and the resonator has been formulated for recorderlike instruments [Verge *et al.*, J. Acoust. Soc. Am. **101**, 2925–2939 (1997)]. A new simulation model of the transverse flute is presented which combines the contributions of digital waveguide modeling and the new source model. This new model is defined almost entirely by physical parameters (such as dimensions of the instrument) rather than by the arbitrary adjustment parameters often employed. The model is evaluated by comparing the effects of certain parameters on the model's operation with their effects on the performance of an actual flute. The model is further evaluated for its tuning characteristics by comparing its frequencies of oscillation with the sounding frequencies of a simple flute with matched physical parameters. [Work supported by the University of Dayton Honors Program.]

10:30

**2aMUa8. Analysis of cymbal vibrations: Lyapunov spectrum and route to chaos.** Cyril Touzé and Antoine Chaigne (ENST, Dept. TSI, CNRS URA 820, 46 rue Barrault, 75634 Paris Cedex 13, France)

A new step in the analysis of cymbal vibrations is presented. The goal of this analysis is to gain a better understanding of the underlying mechanisms that govern the sound of these instruments. In a previous work, geometrical invariants (correlation dimension and percentage of false nearest neighbors) were extracted from the reconstruction of the cymbal vibration in phase space [Touzé *et al.*, Proceedings of ISMA'98, Leavenworth (1998)]. Here, the Lyapunov exponents of the system are computed in order to assess the presence of chaos and to quantify it. The method used consists of approximating the tangent flow of the trajectory in the reconstructed phase space. An algorithm which uses a higher-order polynomial approximation has been developed. This algorithm is validated on a number of well-known theoretical time series (Henon map, Duffing equation, etc.). The Lyapunov spectrum of the cymbal shows, in particular, the convergence of the greatest exponent, which is a strong indication for the presence of chaos. A closer analysis reveals a kind of Ruelle–Takens scenario for the transition from linear to chaotic motion: Power spectra and Lyapunov exponents show that mode-locking and quasi-periodicity are present in the vibrations just before the chaotic regime.

## Session 2aMUb

**Musical Acoustics and Committee on Archives and History: History of Musical Acoustics**

Uwe J. Hansen, Chair

*Physics Department, Indiana State University, Terre Haute, Indiana 47809*

Chair's Introduction—11:00

*Invited Paper*

11:05

**2aMUb1. Musical acoustics in the Twentieth Century.** Gabriel Weinreich (Randall Lab. of Phys., Univ. of Michigan, Ann Arbor, MI 48109-1120)

Of humankind's material creations that developed without benefit of logical understanding, musical instruments may well be the most complex and sophisticated. Although some of the greatest theoretical minds of the nineteenth century (such as Helmholtz and Rayleigh) applied notable efforts to the study of musical instruments, that period lacked the experimental tools to match the virtuosity of its theoreticians; so that it was not until the twentieth century, and more particularly, the period that overlaps the life of the Acoustical Society of America, that the field of Musical Acoustics was truly able to blossom forth. Specifically, ASA's first quarter century saw the development of sophisticated vacuum-tube technology (powerfully stimulated by World War II); the second produced solid-state electronics; and the third brought personal computers to every scientist's desk and made digital methods part of everyday life. But, unlike the situation in other fields, modern electronic methods not only made the physics of musical instruments more accessible, they themselves invaded their subject matter in the form of electronic music, giving rise to a totally new kind of symbiotic relationship. This talk will draw upon the annals of JASA to summarize that fascinating period of history.

## Session 2aPA

**Physical Acoustics: Earth and Air-Borne Sound**

Craig J. Hickey, Chair

*National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677**Contributed Papers*

8:45

**2aPA1. Simplified model of the seismic reverberation observed due to the high-resolution transient acoustic signal.** Yevgeniy Y. Dorfman, Peter A. Krumhansl, and Michael Goldsmith (BBN Technologies, Operating Unit of GTE, 70 Fawcett St., Cambridge, MA 02138)

Unexploded ordinance (UXO) presents a lethal hazard to individuals worldwide. To predict and/or simulate the performance of a seismic sonar for detecting UXO, one needs to develop a model of propagation and reverberation in the inhomogeneous seismic media. An underground seismic detection experiment was performed during fall 1998 as a part of the BBN Innovative Seismic System for Buried Unexploded Ordinance Detection and Classification program. A physics-based model of monostatic reverberation due to the low-directivity shear wave source pointed down was postulated. This model provides a way to (a) simulate the time series for different source-receiver geometries and source signals; and (b) to invert the field measurements for ground parameters. Exercising the model, it was found that the observed signal due to reverberation is a function of shear wave speed (usually known), exponential attenuation, and assumed reverberation strength (often not well-known). It was further found that under realistic assumptions for UXO detection scenario frequency-dependent attenuation must be accounted for in the modeling. A numerical code was developed based on the same physical assumptions in

order to account for nonmonostatic source-receiver configuration. It was found that collected data are matched well by the numerical model predictions. [Work supported by SERDP.]

9:00

**2aPA2. Acoustic landmine detection using laser scanning vibrometer.** James Sabatier and Ning Xiang (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Airborne acoustic waves that couple into the surface of the ground excite Biot compressional and shear waves. If a minelike target is present below the surface, the ground surface vibrational velocity will show distinct changes due to reflection and scattering of these waves. Taking advantage of a noncontact measurement technique, the surface vibrational velocity is detected with a laser Doppler vibrometer. Sound waves with a wavelength comparable to the object size are suitable for recognizing geometrical shapes of targets, while true wavelike acoustic scattering phenomena can be observed with a shorter acoustic wavelength. In the present contribution, both shape/size recognition and scattering phenomena of buried mines detected by the laser Doppler based acoustic detection will be visualized in terms of a number of 2-D and 3-D animations. [The work is supported by U. S. Army Communications-Electronics Command.]

9:15

**2aPA3. Notes on nonlinear elasticity of rocks.** Lev A. Ostrovsky (Univ. of Colorado/NOAA Environ. Technol. Lab., 325 Broadway, Boulder, CO 80303)

Recent experiments with samples of Earth materials such as sandstone have shown that they possess a strong nonlinearity, and that water saturation, even in small amounts, may strongly affect nonlinear properties of the material. Here, a theoretical model is considered which represents the rock as a conglomerate of "large" grains (of order 100  $\mu\text{m}$ ) with two types of contacts: hard ones which determine the basic elastic matrix, and soft contacts, due to much smaller contacting grains which can create strong nonlinearity. Also, the effect of capillary and Van der Waals forces in thin liquid layers between grains is considered. The theoretical estimates are compared with available experimental data.

9:30

**2aPA4. Resonance inversion for elastic moduli of anisotropic rocks.** Seiji Nakagawa, Kurt T. Nihei, and Larry R. Myer (Earth Sci. Div., E. O. Lawrence Berkeley Natl. Lab., 1 Cyclotron Rd., Berkeley, CA 94720, seiji@friction.lbl.gov)

In this research, acoustic resonance spectroscopy was applied to determine the dynamic elastic constants of isotropic glass and transversely isotropic rock cubes. This technique consists of resonating the specimen over a broad range of frequencies, measuring the resonance frequencies, and computing the elastic constants by nonlinear inversion of the measured resonance frequencies. Specimens were tested under unconfined, traction free conditions. Resulting surface vibrations were measured using a miniature accelerometer and their spectral characteristics were analyzed. The inversion was performed using a numerical algorithm based on the Rayleigh-Ritz method that minimized the difference between measured and computed resonance frequencies iteratively. Mode shapes of the anisotropic specimens were also measured using a laser Doppler vibrometer and compared with the prediction of the numerical model. Comparison between the elastic moduli of rock specimens determined by static loading tests, resonance inversion, and ultrasonic transmission tests showed good agreement between the ultrasonic and resonance results but the moduli determined from ultrasonic measurements were consistently higher than the resonance inversion. Such results may be due to the frequency-dependence of the wave velocity in microscopically heterogeneous rock and nonelastic (frictional) deformation of the rock specimen during the static loading tests.

9:45

**2aPA5. Acoustic signatures of a fracture during air injection.** Kurt T. Nihei, John E. Peterson, Jr., Larry R. Myer, and Ernest L. Majer (Earth Sci. Div., Lawrence Berkeley Natl. Lab., 1 Cyclotron Rd., M.S. 90-1116, Berkeley, CA 94720)

This study examines the acoustic signatures of transmitted, reflected, and guided waves during air injection into a single, natural, water-saturated fracture in limestone. The presence and location of the fracture were established in a series of geologic, hydrologic, and seismic studies (Queen and Rizer, 1990; Datta-Gupta *et al.*, 1994; Majer *et al.*, 1997) that ultimately led to its verification in core obtained from a slanted well. This work describes the results of a follow-up high frequency (1 to 10 kHz) crosswell survey that was designed to illuminate the fracture by air injection into the fracture. Zero-offset *P*-wave crosswell transmission and reflection measurements conducted during air injection showed a large decrease in the amplitude of the transmitted wave (approximately 10 times reduction at 3 kHz), and a smaller increase in the amplitude of the reflected wave (approximately 1.5 to 5 times increase at 3 kHz). Measurements of the *P*-wave and an interface wave propagating along the fracture also show a small increase in amplitude during air injection. Analyses of

these measurements using numerical boundary element and finite-difference simulations and the importance of including fracture stiffness heterogeneity arising from irregular distributions of air inside the fracture will be presented.

10:00

**2aPA6. Wind generated sound in standing corn.** David G. Browning (Dept. of Phys., East Hall, Univ. of Rhode Island, Kingston, RI 02881)

The "sound of corn growing" in farming folklore appears to be due to wind puffs, during otherwise calm conditions (especially at night), causing isolated audible popping sounds due to leaf striking leaf [D. G. Browning, *Am. J. Botany* **84**, 38(A) (1997)]. When the wind becomes steady, the number of events greatly increases, resulting in a rustling sound with a smoothed, broad spectrum between 1 and 5 kHz, and a peak at 2.5 kHz. For a windspeed of 20 mph the peak level is 20 dB above ambient. As the corn matures, the increased weight of the corn ears causes greater stalk sway at a given windspeed; also the leaves tend to become more brittle due to reduced moisture content. Both of these changes appear to alter the measured sound spectra, offering a means for evaluation.

10:15

**2aPA7. Seasonal variability in the atmosphere and its effect on infrasonic propagation.** David E. Norris and Robert Gibson (BBN Technologies, 1300 N. 17th St., Arlington, VA 22209)

Infrasonic waves can propagate thousands of kilometers in range and sample regions of the atmosphere from the ground up to and including the thermosphere. In this study, seasonal changes in the atmosphere and their effect on infrasonic propagation are characterized. The NASA/NRL empirically based models HWM-93 (for winds) and MSIS-90 (for temperature) are used. Three-dimensional ray traces are computed through the modeled atmosphere for several representative scenarios. Seasonal trends in both ray arrival times and ray azimuth bias are computed, and limited comparisons with data are made where possible. [Sponsored by Defense Threat Reduction Agency, Contract No. DSWA01-97-C-0160.]

10:30-10:45 Break

10:45

**2aPA8. Spectral broadening of sound scattered by atmospheric turbulence.** George H. Goedecke, Roy C. Wood (Dept. of Phys., New Mexico State Univ., P.O. Box 30001, Las Cruces, NM 88003-8001), Harry J. Auvermann (U.S. Army Res. Lab., Adelphi, MD 20783-1155), and Vladimir E. Ostashev (Environ. Technol. Lab., Boulder, CO 80303)

Scattering of a monochromatic sound wave by atmospheric turbulent eddies that are moving with the mean wind is described. The source and detector have wide radiation patterns and are at rest in a ground fixed frame. For eddies that make the dominant contribution to the detector signal, scattering angles change substantially with time, so the signal displays a time-dependent frequency which may include the full longitudinal Doppler width. A computer code is developed that calculates the time-dependent detector response and its Fourier spectrum due to one or many eddies, including a steady-state collection of eddies of many different scale lengths that models homogeneous and isotropic atmospheric turbulence. Several numerical results from this code are presented, including one for a simulation of a recent experiment. The predicted spectral characteristics are in very good agreement with the experimental ones. Some possible extensions of the model for describing anisotropic and intermit-



tent atmospheric turbulence are discussed. [Work supported in part by the U.S. Army Research Office under Contract Nos. DAAG55-98-1-0463 and DAAG55-97-1-0178, and an NRC-ETL Research Associateship.]

11:00

**2aPA9. Numerical solution of a second-moment parabolic equation for sound propagation in a random medium.** D. Keith Wilson (U. S. Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783) and Vladimir E. Ostashev (Environ. Technol. Lab., Boulder, CO 80303)

A parabolic equation for the second-moment of the acoustic pressure field in a refractive medium with random fluctuations is derived using the Markov approximation. The PE is then discretized and written as a matrix equation for the purpose of numerical solution. The matrix equation involves the product of an  $N_z^2 \times N_z^2$  tridiagonal matrix with “fringes” times the  $N_z^2 \times 1$  second-moment vector, where  $N_z$  is the number of points in the vertical grid. (The numerical formulation of the usual PE for the acoustic pressure involves an  $N_z \times N_z$  pure tridiagonal matrix.) Direct solution of this extremely large linear system is prohibitive. A pair of alternative strategies are therefore devised. One involves Cholesky factorization of the second-moment matrix. The other is based on iterative solution of the second-moment equation, with the coherent field used to start the iterations. Both of these methods are applied successfully to propagation in an upwardly refracting atmosphere, although solution of the second-moment PE is still very computationally intensive. [Work supported in part by the U.S. Army Research Office under Contract No. DAAG55-98-1-0463, and NRC-ETL Research Associateship.]

11:15

**2aPA10. Daytime sound levels and coherence as a function of height: Experiment and theory.** Xiao Di (Appl. Res. Lab., Pennsylvania State Univ., State College, PA 16804), Kenneth E. Gilbert (Univ. of Mississippi, University, MS 38677), and Richard D. Clark (Millersville Univ., Millersville, PA 17551-0302)

Daytime ground-to-ground acoustic propagation has been measured by a number of researchers over the years. Measurements of daytime sound levels as a function of height, however, have been made only in a limited number of near-ground experiments and apparently never at a considerable height. This paper reports measurements using a tethered balloon to measure the mean sound level and coherence from the ground up to 210 m at three separate ranges (300 m, 600 m, and 900 m) and at three frequencies (210 Hz, 380 Hz, and 600 Hz). A meteorological instrument package on the balloon measured the mean temperature, wind and humidity as a function of height. Fluctuation spectra were estimated using a stationary sonic anemometer at 5-m height which sampled the temperature and vector wind at 10 Hz. Acoustical inputs computed from the meteorological data were used with a three-dimensional parabolic equation model to predict the mean sound level and coherence as a function of height and range. The predictions and measurements agree reasonably well, both showing a

clear shadow-zone boundary that is a function of the meteorological conditions. Descending across the shadow boundary, the predicted and measured mean sound levels and coherence decrease dramatically.

11:30

**2aPA11. Propagation of a monochromatic sound wave in a turbulent atmosphere near the ground: Theory and laboratory experiment.** Vladimir E. Ostashev<sup>a)</sup> (Environ. Technol. Lab., 325 Broadway, Boulder, CO 80303), Julie Wasier, Philippe Blanc-Benon, and Danile Juvé (Ecole Centrale de Lyon, BP 163, 69131 Ecully, France)

The theory of sound propagation in a turbulent atmosphere near the impedance ground is presented. It is assumed that the monochromatic source and receiver are located above the ground, and there is no temperature or wind velocity stratification in the atmosphere. A formula for the mean square sound pressure at the receiver is obtained and studied analytically and numerically as a function of different parameters of a problem: the variance and integral length scale of atmospheric inhomogeneities, the horizontal distance between the source and receiver, the value of impedance of the ground, etc. Theoretical results obtained are also compared with data from laboratory experiment modeling sound propagation in an atmosphere with temperature fluctuations near the ground. The experiment was carried out at Ecole Centrale de Lyon. The agreement between theoretical and experimental results is generally good. [This material is based upon work supported in part by the U.S. Army Research Office under Contract No. DAAG55-98-1-0463, and NRC-ETL Research Associateship.] <sup>a)</sup> On leave from New Mexico State Univ.

11:45

**2aPA12. The effects of the wind velocity fluctuations on the sound backscattering cross section in the stratified moving atmosphere.** Vladimir E. Ostashev (Environ. Technol. Lab., 325 Broadway, Boulder, CO 80303) and D. Keith Wilson (U.S. Army Res. Lab., Adelphi, MD 20783-1197)

Recent studies have shown that, in the case of strong diffraction, the relative contributions from wind velocity and temperature fluctuations to many statistical moments of a sound field propagating in a turbulent atmosphere are determined by the ratio of the normalized variances of wind velocity and temperature fluctuations, multiplied by factor 4. On the other hand, for weak diffraction, these relative contributions are determined by the ratio of the normalized structure-function parameters of wind velocity and temperature fluctuations, multiplied by factor 22/3. In this paper, we calculate the vertical profiles of both ratios for different turbulent regimes in unstable atmospheric boundary layers. Furthermore, vertical profiles of the ratio for weak diffraction are used to study the effects of the wind velocity fluctuations on the sound backscattering cross section in a stratified moving atmosphere. This study is important for acoustic remote sensing of the atmosphere by monostatic sodars which measure a signal proportional to the backscattering cross section. [Work supported in part by the U.S. Army Research Office under Contract No. DAAG55-98-1-0463, and an NRC-ETL Research Associateship.]

## Session 2aPP

## Psychological and Physiological Acoustics: Detection and Resolution

Nandini Iyer, Chair

*Speech and Hearing Science, The Ohio State University, 110 Pressey Hall, 1070 Carmack Road, Columbus, Ohio 43210*

## Contributed Papers

9:00

**2aPP1. Hearing discrete elements of an auditory temporal sequence: Evidence for a central mechanism.** Dennis T. Ries, Jeffrey J. DiGiovanni, and Robert S. Schlauch (Dept. of Commun. Disord., Univ. of Minnesota, Minneapolis, MN 55455)

The ability to hear discrete elements of a temporal sequence was assessed. In one experiment, just-noticeable differences (jnds) in level between adjacent pulses of a 3- or 4-pulse temporal sequence that alternated in level was measured. The stimuli were pure tones (0.5-, 1.0- and 4.0-kHz tones), frozen noise, and white noise. All of the stimuli produced identical results. The jnd was about 5 dB for pulse rates less than 15 Hz and increased dramatically for higher rates. In a second experiment, listeners judged the initial direction change (high or low) of a frequency-modulated pure-tone sweep and a sinusoidal amplitude-modulated broadband noise. As in the first experiment, subjects performed poorly for rates higher than about 15 Hz. The similarity of the results for white noise, frozen noise, tones of different frequencies, AM noise, and FM sweeps is consistent with the notion that the factor-limiting performance is not in the auditory periphery. It is interesting to note that the approximately 15-Hz upper limit of performance on these tasks is consistent with the highest rates observed for phase locking in the auditory cortex. [Work supported by NIH-NIDCD R29 DC01542.]

9:15

**2aPP2. Effects of masker harmonicity on masked threshold.** William C. Treurniet and Darcy R. Boucher (Commun. Res. Ctr., P.O. Box 11490, Station "H," Ottawa, ON K2H 8S2, Canada, bill.treurniet@crc.ca)

The role of masker harmonicity in auditory masking is not yet fully understood. This paper compares the effect of harmonic and inharmonic maskers on the simultaneous masked thresholds of narrow bandwidth noise probes using a three-alternative, forced-choice method. The harmonic masker had an 88-Hz fundamental and 45 consecutive partials (random starting phases). The inharmonic masker differed in that the partial frequencies were adjusted to nearby prime numbers. Maskers were presented at several levels, and the probe at several frequencies. Thresholds were consistently lower for the harmonic masker (4–8 dB). The masking release persisted when masker and probe bandwidths were equal, so it is not due to comodulated auditory filter outputs. The difference between maskers disappeared when the harmonic partials were separated by more than 176 Hz, suggesting that the effect may be related to critical bandwidth. Further, the threshold of noise probes located between two partials was always lower for the harmonic masker. However, the threshold for a tonal probe at the frequency of an omitted partial was higher with the harmonic masker than with the inharmonic masker. The results are consistent with a harmonic template model that proposes inhibition at the template frequencies and disinhibition between those frequencies.

9:30

**2aPP3. Masking by harmonic complexes with different phase spectra in hearing-impaired listeners.** Jennifer J. Lentz and Marjorie R. Leek (Army Audiol. and Speech Ctr., Walter Reed Army Medical Ctr., Washington, DC 20307)

At the last meeting, we showed that in normal-hearing listeners, masking effectiveness of harmonic complexes depends on the maskers' phase spectra. Such masking differences are thought to partially reflect an interaction between the input waveform and phase characteristics of the normal auditory system. We have extended that work to explore masking by harmonic complexes in listeners with hearing impairment. The masker was a harmonic complex with equal-amplitude components, with phases selected according to scaled modifications of the Schroeder algorithm. Masking was measured for several different signal frequencies and levels. In contrast to normal-hearing listeners, most hearing-impaired listeners showed little change in masking for different phase selections. Some listeners did demonstrate masking differences, but the greatest reduction in masking occurred when the components had nearly equal phases (i.e., a highly modulated masker waveform). Apparently, in those listeners, the within-channel phase characteristic did not substantially alter the phase structure of the input stimulus, as is thought to occur in normal ears. In aggregate, these data suggest that sensorineural hearing loss is accompanied by deficits in temporal resolution, possibly involving changes in the phase characteristics of auditory channels. [Work supported by NIH DC00626.]

9:45

**2aPP4. Factors accounting for performance in a frequency-sample-discrimination task with distracters.** Donna L. Neff, Eric C. Odgaard, Walt Jesteadt, and Huanping Dai (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, neff@boystown.org)

This paper presents results of fitting several linear models, including the CoRE model of Lutfi [R. A. Lutfi and K. A. Doherty, *J. Acoust. Soc. Am.* **96**, 3443–3450 (1994)] to data from a study of the interaction of target and distracter tones in a sample-discrimination task. In this 2IFC task, ten listeners judged which of two pairs of target tones were drawn from the higher of two overlapping frequency distributions. After training with targets alone, two distracter tones were presented simultaneously with the target tones, with each distracter placed in a frequency region remote from the target region. The important variables were the frequency regions and degree of frequency variability of targets versus distracters. Main features of the data included a large range of individual differences, a dominance of the effects of lower-frequency context tones over the higher-frequency context tones in many conditions, and differences in performance in the baseline, no-context conditions. When stimulus variability, perceptual weights across frequency, and performance in the no-context conditions are included as factors in the analysis, both mean and individual performance can be described reasonably well. [Work supported by NIDCD.]

10:00

**2aPP5. Temporal factors in auditory peak detection of modulation of tones.** Julius L. Goldstein (Dept. of Elec. Eng., Washington Univ., Campus Box 1127, St. Louis, MO 63130) and Joseph L. Hall (Bell Labs., Lucent Technologies, Murray Hill, NJ 07974)

Auditory detection of tone modulation has been modeled as envelope-peak detection of the cochlear filterbank responses [J. L. Goldstein, J. Acoust. Soc. Am. **41**, 458–479 (1967)]. To study the ability of this model to quantify temporal smoothing in detection, new psychophysical experiments using a 4IFC paradigm were conducted, with the authors as subjects. AM or quasi-FM sinusoidally modulated tones were discriminated from the carrier tone. Only one should was modulated among the four pushed sounds in the paradigm. The first series of exploratory experiments (1994) focused on low modulation frequencies (4–16 Hz). A second series of systematic experiments (1996) with a wide range of modulation frequencies allowed estimation of detector smoothing at carrier frequencies of 0.25, 1, and 4 kHz. Despite differences between subjects in estimated spectral filtering, modulation thresholds for both subjects gave similar estimates of smoothing at each carrier frequency. At modulation frequencies above 20 Hz, the smoothing function showed similar dependence on carrier frequencies as found from the early experiments (1967). Below 20 Hz, the early experiments, which used a method of adjustment paradigm, reveal a hyperacuity. This suggests a second slower detection mechanism that exploits repeated presentations of test sounds. [Work supported by NSF Grant IBN-9728383.]

10:15–10:30 Break

10:30

**2aPP6. Statistical theory of peak detection for human hearing.** Julius L. Goldstein (Dept. of Elec. Eng., Washington Univ., Campus Box 1127, St. Louis, MO 63130)

Envelope peaks and peak factors of narrow-band sounds have been found in simulation studies to explain a wide range of experiments on sound discrimination [e.g., J. L. Goldstein, J. Acoust. Soc. Am. **99**, 2541(A) (1996)]. To extend understanding of peak detection, mathematical approximations were developed for the first and second moments of log (dB) envelope statistics of a tone centered in noise with duration  $T$  and bandwidth  $W$ . Noises considered are: uniform periodic noise (UPN) having uniform amplitudes and random phases, Gaussian periodic noise (GPN) having random amplitudes and phases, and true Gaussian noise (TGN). Key properties quantified include: (1) Three nearly independent causes of waveform fluctuation underlie peak variance, viz., phase noise, energy noise, and interaction of tone with common frequency noise (energy cross term). (2) The energy variances in peak detection are similar as for energy detection. (3) Peak factor variance is primarily phase noise, which, for noise alone, is  $\sim 1$  dB for  $T \times W > 2$ . (4) Peak and peak factor means differ by mean energy. (5) Mean peak growth for a small tone added to noise follows energy summation, while for a small noise added to a tone, the growth follows rms amplitude summation. [Work supported by NSF Grant IBN-9728383.]

10:45

**2aPP7. Detectability of Gaussian noise by humans and simulated observers: A stimulus-level correlation analysis.** Judi A. Lapsley Miller (Psychophys. Lab., Victoria Univ. of Wellington, New Zealand)

The correlation between ratings made by humans and simulated observers in detecting the same acoustic stimuli was used to estimate the detection (critical) bandwidth, and the form of rectification, temporal integration, and sampling strategy used by the human hearing system. By

using small- $WT$  Gaussian noise, bandwidth ( $W$ ) and duration ( $T$ ) were manipulated independently, resulting in 18 different combinations of bandwidth and duration for  $WT=1, 2, \text{ and } 4$ . The pattern of correlation, as properties of the simulated observer were systematically varied, indicated the best correlated observer. This tended to be the *full-linear* observer, consisting of a bandpass filter (which was generally wider than the bandwidth of the signal), full-wave rectifier, and a full (true) integrator. The full-linear observer was slightly, but consistently, better at describing human performance than the energy or envelope detector.

11:00

**2aPP8. Effect of BM compression on masking period patterns (MPPs).** Magdalena Wojtczak, Anna C. Schroder, Ying-Yee Kong, and David A. Nelson (Clinical Psychoacoust. Lab., Univ. of Minnesota, 516 Delaware St. S.E., Minneapolis, MN 55455)

MPPs were measured in listeners with normal and impaired hearing using tonal 4-Hz SAM maskers and short tonal probes with frequencies that were either identical to or higher than the carrier frequency of the masker. The probe frequencies were 500, 1200, and 3000 Hz for on-frequency masking, and 1200, 2400, and 6000 Hz for off-frequency masking. In normal-hearing listeners MPPs measured with off-frequency probes had valleys that were much longer and deeper than valleys observed with on-frequency probes. A similar result was observed in hearing-impaired listeners in the frequency region of mild hearing losses, where significant residual compression was presumably operating. However, in the frequency region with substantial hearing loss where compression is substantially reduced or absent, MPPs measured with the on- and off-frequency probes were very similar. A model consisting of peripheral filtering, compressive nonlinearity, and a sliding temporal window was used in an attempt to predict the data and to estimate the compression index. The results suggest that the similarity of on- and off-frequency temporal resolution in hearing-impaired listeners may be due in part to the lack of the compressive nonlinearity that is evident at the level of the basilar membrane in normal-hearing listeners. [Work supported by NIH-NIDCD DC00149.]

11:15

**2aPP9. Estimation of psychometric functions using a hybrid up-down maximum likelihood procedure.** He Yuan, Harry Levitt (Ctr. for Res. in Speech & Hearing Sci., City Univ. of New York, 33 W. 42nd St., New York, NY 10036), and James D. Miller (Central Inst. for the Deaf, St. Louis, MO 63110)

Rapid, efficient, and reliable estimation of psychometric functions is a much sought after goal in psychophysics. Adaptive techniques of the up-down type have many advantages in this regard, but they are not without their limitations. An alternative approach is the use of maximum likelihood estimation in placing observations. Whereas the maximum likelihood method is as efficient if not more efficient than any other method of estimation, this property only holds for very large samples. The use of maximum likelihood estimation in adaptive testing has significant limitations when the sample size is very small. In this study, a hybrid adaptive procedure was developed using an up-down technique initially and then converting to a maximum likelihood procedure once sufficient samples have been obtained for reliable estimation using the latter technique. The up-down technique that is used reduces the step size systematically employing a rule analogous to that used in analog-to-digital converters and is referred to as the AD procedure. Monte Carlo simulations of experiments using the hybrid adaptive technique showed promising results in comparison with other adaptive procedures. [Research supported by Grant 5P50DC00178 from NIDCD.]

2a TUE. AM

**Session 2aSA****Structural Acoustics and Vibration, Engineering Acoustics and Physical Acoustics:  
Acoustic Nondestructive Evaluation: New Directions and Techniques, Part I**

Joseph W. Dickey, Chair

*Center for Nondestructive Evaluation, Johns Hopkins University, Baltimore, Maryland 21218-2689***Chair's Introduction—9:25*****Invited Papers*****9:30****2aSA1. Advances in ultrasonic nondestructive evaluation.** Robert E. Green, Jr. (Ctr. for Nondestruct. Eval., Johns Hopkins Univ., Baltimore, MD 21218)

The fundamental principles of linear elastic-wave propagation in isotropic and anisotropic materials emphasizing the modifications to ultrasonic wave propagation caused by anisotropy will be reviewed. Among the topics covered will be wave-propagation modes, energy-flux vector, diffraction-beam spreading, attenuation, and nonlinear effects. Verification of these principles by optical probing of ultrasonic-wave fields in transparent solids will be presented. Recent advances and practical applications of noncontact ultrasound, including pulsed lasers, optical interferometers, electromagnetic acoustic transducers (EMATs), and air-coupled systems will be described. Application of x-ray diffraction topography to verify the mechanical vibration modes of quartz crystals will be illustrated. Finally, suggestions will be made about possible future advances in ultrasonic nondestructive evaluation of materials and structures.

**10:00****2aSA2. A nonlinear mesoscopic elastic class of materials.** Paul A. Johnson (Nonlinear Mesoscopic Elasticity, Los Alamos Seismic Res. Ctr., Los Alamos Natl. Lab., Los Alamos, NM 87545) and Robert Guyer (UMASS Amherst, Amherst, MA)

It is becoming clear that the elastic properties of rock are shared by numerous other materials (sand, soil, some ceramics, concrete, etc.). These materials have one or more of the following properties in common: strong nonlinearity, hysteresis in stress-strain relation, and discrete memory. Primarily, it is the material's compliance, the mesoscopic linkages between the rigid components, that give these materials their unusual elastic properties. It can be said that these materials have nonlinear mesoscopic elasticity and encompass a broad class of materials. Materials with nonlinear mesoscopic elasticity stand in contrast to liquids and crystalline solids whose elasticity is due to contributions of atomic level forces, i.e., materials with atomic elasticity. Atomic elastic materials are well described by the traditional (Landau) theory of elasticity; however, mesoscopic elastic materials are not. Mesoscopic materials are well described by the P-M (Preisach-Mayergoyz) model of nonlinear elasticity developed by Guyer and McCall. A sequence of experiments on numerous materials illustrate the evidence of nonlinear mesoscopic elastic behavior. In experimental analysis a surprising discovery was made: damaged atomic elastic materials behave as mesoscopic elastic materials. It is significant that the nonlinear mechanism(s) in mesoscopic elastic materials remains a mystery.

**10:30–10:45 Break****10:45****2aSA3. Noncontact ultrasonics for materials nondestructive evaluation.** Francesco Lanza di Scalea and Robert E. Green, Jr. (Ctr. for Nondestruct. Eval., Johns Hopkins Univ., Baltimore, MD 21218)

Recent advances in ultrasonic techniques for remote and noncontact nondestructive evaluation of structural materials will be presented. The principles and techniques of the laser generation and detection of ultrasonic waves in opaque solids will be reviewed. A laser-based beam-steering C-scan system will be described, along with a novel stabilization procedure for ultrasound detection by a confocal Fabry-Perot interferometer which appears to be more effective than previous approaches for high-sensitivity measurements in beam-steering scan applications. A hybrid system which is based on laser generation of narrow-band acoustic surface waves and their detection by gas (air)-coupled transduction will also be presented. This system uses fiberoptic light delivery which enhances its flexibility and remoteness from the test piece. The main issues governing the efficient coupling of high-power laser light into optical



fibers will be discussed. An improved theory for the laser generation of narrow-band surface acoustic waves by the spatial modulation technique will be outlined and the effect of the number of illumination sources on the bandwidth of the generated wave will be assessed experimentally and theoretically. Finally, the suitability of the hybrid system for the intelligent control of the tow-placement process of advanced thermoplastic composites will be demonstrated.

### Contributed Papers

11:15

**2aSA4. Computational methods for an inverse problem in acoustics.** Thomas DeLillo, Victor Isakov, Nicolas Valdivia, and Lianju Wang (Dept. of Mathematics and Statistics, Wichita State Univ., Wichita, KS 67260-0033 )

The problem of computing normal velocities on the boundary of a region from pressure measurements on an interior surface is considered. The pressure satisfies the Helmholtz equation and is represented by a single layer potential. Once the density function is found, the normal velocities can be easily computed by applying a second kind of integral operator. The problem of solving for the density function from the interior pressure measurements is ill-posed. Regularization methods using the singular value decomposition and iterative methods, such as the conjugate gradient method for the normal equations, are compared for pressure data with noise. [Work supported by NSF and Cessna Aircraft Company.]

11:30

**2aSA5. Remote ultrasonic classification of fluids in cylindrical containers by analyzing the response of circumferential guided waves.** Gregory Kaduchak and Dipen N. Sinha (Los Alamos Natl. Lab., Electron. and Electrochemical Mater. and Devices Group, MS D429, Los Alamos, NM 87545)

A novel technique for classifying fluids in sealed, metal containers at large stand-off distances has been developed. It utilizes a recently constructed air-coupled acoustic array made from inexpensive, commercial-off-the-shelf components to excite the resonance vibrations of fluid-filled

vessels. The sound field from the array is constructed by transmitting a high-frequency modulated carrier wave which utilizes the nonlinearity in the air medium to demodulate the carrier frequency along its propagation path in air. The array has a narrow beamwidth and an operating bandwidth of greater than 25 kHz. The vibrations are detected using a laser vibrometer in a monostatic configuration with the acoustic source. It is shown that the propagation characteristics of the  $ao$  Lamb wave are highly affected by different interior loading conditions on the interior wall of a cylindrical container. Classification of the interior fluid is obtained by analyzing the change of this response as a function of frequency. Experiments demonstrate that classification of the fluid-filler inside closed, steel vessels is possible with incident sound pressure levels of the demodulated wave as low as 80 dB at the container location. Preliminary experiments demonstrate that stand-off distances greater than 3 m are achievable for classification purposes.

11:45

**2aSA6. Acoustical free oscillation NDT method of cracks for turbomachinery.** Leonid M. Gelman, Olena V. Bezvesil'na, Sergey V. Gorpinich (Dept. of Nondestructive Testing, Natl. Tech. Univ. of Ukraine, 37, Peremogy pr., Kiev, 252056 Ukraine), and Aleksandr P. Stokov (GSKBD, Kharkov, Ukraine)

A new vibroacoustical nondestructive testing and evaluation method is proposed. The method uses phase spectral density harmonics of testing object free oscillations as features. New analytical expressions of phase spectral density of testing object free oscillations are received and considered.

TUESDAY MORNING, 2 NOVEMBER 1999 DELAWARE A & B ROOM, 9:00 A.M. TO 12:00 NOON

### Session 2aSC

### Speech Communication: Cross-Linguistic Aspects of Speech Production and Perception (Poster Session)

James E. Flege, Chair

*Department of Rehabilitation Sciences, University of Alabama, VH 503, Birmingham, Alabama 35294-0019*

### Contributed Papers

To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon. To allow for extended viewing time, all posters will be on display from 9:00 a.m. to 10:00 p.m.

**2aSC1. Searching for foreign accent.** Z. S. Bond, Verna Stockmal, and Danny R. Moates (Ohio Univ., Athens, OH 45701-2979)

Native speakers of English can identify non-native speakers with relatively little difficulty. Further, they claim to be able to identify the native language of non-native speakers, at least as suggested by such descriptive terms as French accent or Arabic accent, implying that non-native English carries properties which are characteristic of native languages. In four experiments, we investigated whether English speakers could match an unknown foreign language with a foreign accent. In the first two experiments, listeners heard a sample of accented English and were asked to

identify the native language of the speakers from a series of competitors. The target languages were Japanese and Latvian. Listener performance did not exceed chance. In the third experiment, listeners made "yes-no" responses to accented English matched with foreign languages, including the native language of the speaker. Although listeners thought some languages were more likely to be the source of the foreign accent than others, they did not identify the target language correctly. In the fourth experiment, listeners made ratings on a seven-point scale, judging the similarity of accented English and various foreign languages, with results very similar to those of experiment 3.

**2aSC2. Effects of age and L1 use on native Italian speakers' perception of English vowels.** James E. Flege (Dept. of Rehabilitation Sci., VH503, Univ. of Alabama at Birmingham, Birmingham, AL 35294, jeflege@uab.edu) and Ian R. A. McKay (Univ. of Ottawa, Ottawa, ON K1N 6N5, Canada)

This study examined the categorial discrimination of pairs of Canadian English vowels (/i/-/ɪ/, /ɪ/-/ε/, /ε/-/e/, /e/-/æ/, /eɪ/-/i/, /ɒ/-/ʌ/, /æ/-/ʌ/, /ɔ/-/ʌ/, /eɪ/-/i/) by groups of native Italian speakers who arrived in Canada at average ages of 7.6 and 19.5 years. Half of the 36 early and the 36 late bilinguals reported using Italian relatively seldom (*mean*=8%) or often (*mean*=48%). The subjects clicked "1," "2," or "3" to indicate the serial position of an odd item out, or "no" to indicate hearing three physically different instances of a single category. The discrimination (*A'*) scores were significantly higher for early rather than late bilinguals, and for bilinguals who used Italian seldom rather than often. The lack of a two-way interaction indicated that the L1 use effect was comparable for the early and late bilinguals. The "Early-High Use" group obtained lower scores than the native English comparison group for one constant (/ɒ/-/ʌ/) but the "Early-Low Use" group did not differ from the native English group for any contrast. The results suggest that both the state of development of the L1 phonetic system at the time L2 learning commences and language use influence the accuracy with which L2 vowels are perceived.

**2aSC3. The effect of vocalic F0 on the identification of Korean stops.** Mi-Ryoung Kim (Prog. in Linguist., Univ. of Michigan, Ann Arbor, MI 48109-1285, kmrg@umich.edu)

This study investigates the contribution of vocalic *F0* following the release of an initial stop to the identification of the three-way stop contrast (aspirated, lenis, and fortis) in Korean. Previous acoustic studies have shown an effect of initial stops on *F0* of the following vowel. Using synthetic stimuli, perceptual studies [Han (1996); Cho (1996)] have shown that listeners use *F0* in contrasting stops. However, the relative contribution of the consonantal and vocalic information remains unclear. To address this issue, a perceptual study was conducted using cross-spliced stimuli from three naturally produced syllables across all three places articulation in the /a/ context and for alveolars in the /i/ context. Stimuli were created by cross-splicing each consonantal portion (prior to voicing onset) with each vocalic portion (after voicing onset) to create all possible combinations. Twelve Korean listeners each responded to a total of 1444 tokens. The results showed that vocalic *F0* is more important than the consonantal onset to stop identification, suggesting that Korean uses tonal variation to differentiate consonant types.

**2aSC4. Using resynthesized speech in /r/-/l/ production and perception training.** Rieko Kubo, Reiko Akahane-Yamada, and Hideki Kawahara (ATR Human Information Processing Res. Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan)

To find effective speech stimuli for second-language (L2) training, we generated resynthesized speech using the STRAIGHT algorithm developed by Kawahara (1997). First, Japanese speakers were trained to produce English /r/ and /l/ in a reproduction task. Training proceeded by first using model words which were artificially lengthened by STRAIGHT, and then gradually introducing words at a natural speed. Spectrograms of the model speech and the learner's speech were given as feedback. Subjects improved significantly from pretest to post-test in their production ability as evaluated by human judges. Second, resynthesized speech was used in perception training. Previous perceptual training studies using synthetic speech have not always been successful. However, training based on resynthesized speech from STRAIGHT led to the same amount of improvement as training based on the original natural tokens, suggesting the potential for using STRAIGHT in L2 perception training. We then manipulated certain acoustic parameters of the stimuli, namely, lengthening and/or enhancement of the third formant. The training effect using these stimuli did not differ significantly from the effect found with the unmodified synthetic stimuli. These results suggest that resynthesized

speech from STRAIGHT is suitable for L2 training, but further effort is necessary to find more effective stimulus manipulations.

**2aSC5. Listeners' representations of within-word structure of native and nonnative words by Japanese speakers.** Takashi Otake (Dokkyo Univ., 1-1, Gakuen-cho, Soka, Saitama, 340-0042 Japan) and Kiyoko Yoneyama (Ohio State Univ., Columbus, OH 43210-1298)

This study further explored how Japanese represented within-word structure of native and nonnative spoken words in terms of two phonological units, morae and syllables, using a new task. The main question addressed in this study was to investigate which phonological unit was preferred to represent nonnative and native words. Three experiments were conducted with Japanese college students, using a tapping task. In experiment 1, 42 syllable Japanese materials which included a CVN syllable in the first syllable and fillers were presented to 20 Japanese college students to identify the number of chunks. The result showed that they preferred morae to syllables. In experiments 2 and 3, 42 syllable English and Spanish materials which included exactly the same syllable structure and fillers were presented to the same number of students, respectively. The results showed that the preference was around 50% each, respectively. These results obviously contradicted to the previous findings in which morae were preferred to nonnative words. The results in this study may suggest that the previous findings using an explicit segmentation task may have been influenced by orthography and that Japanese may be able to suppress morae to represent nonnative words. [Work supported by IFC and Ministry of Education.]

**2aSC6. The effect of duration on the perception of Cantonese level tones.** Patrick C. M. Wong and Randy L. Diehl (Dept. of Psych., Univ. of Texas, Austin, TX 78712)

An inverse relationship has been claimed to exist between the pitch of phonemic tones and the duration of the tone-bearing syllables [e.g., Blicher *et al.*, *J. Phonetics* **18**, 37-49 (1990)]. If this generalization is correct, syllable duration might be expected to serve as a perceptual cue to tone category. In this study, Cantonese-speaking listeners identified tones in sentence context. A naturally produced Cantonese sentence (/ha6 yat1 go3 zi3 hai6 si3/ "The next word is to try") was resynthesized in 25 versions such that the fundamental frequency (*F0*) contour of the context (the first five syllables) was shifted upwards, 1%, 3%, 6%, 9%, or 12% relative to the original, and the final, target syllable was resynthesized with durations that were 80%, 90%, 101%, 110%, or 120% of the original. We predicted that the higher context *F0*s and longer target syllable durations would yield more low-tone identification responses. The results of an analysis of variance showed a significant main effect of contextual *F0* in the expected direction, but no significant effect of target syllable duration. This last result calls into question the generality of the assumed perceptual relationship between duration and tone category. [Work supported by NIDCD.]

**2aSC7. Effect of linguistic background on accuracy of discrimination task and reaction time.** Aman Kumar (Prog. in Linguist., Univ. of Michigan, 1076 Frieze Bldg., Ann Arbor, MI 48109-1285)

This study examines how American English (AE) speakers categorize Hindi laryngeal and place of articulation contrasts. These results are interpreted with respect to the Perceptual Assimilation Model (PAM) [Best, "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (York Press, Inc., 1995), pp. 171-204]. Accuracy and reaction time data were obtained for discrimination and identification of Hindi bilabial and dental phonation contrasts and the retroflex-dental place contrast. Subjects included both AE speakers and a control Hindi speaker group. Discrimination accuracy averaged across the subjects for the phonation contrasts was nativelike while accuracy for the retroflex-dental contrast was significantly lower but above chance. Overall, the reaction time was inversely

proportional to discrimination accuracy. For some subjects discrimination performance for the retroflex-dental contrast was nativelike. However, the reaction time for those subjects was significantly higher than that of other speakers. In general, the reaction time was higher for the retroflex-dental contrast than for the other contrasts.

**2aSC8. Perception of English sentence stress by Yarmouk University English majors.** Fares Mitleb, Fawwaz Al-Haq, and Rasheed Al-Jarrah (Speech and Hearing Ctr., Yarmouk Univ., Irbid 21163, Jordan)

Perception of English sentence stress at discourse level was the focus of the present paper. Twenty English majors, freshmen and seniors at Yarmouk University, voluntarily took part in the study. In the scope of this paper, it was assumed that the subject would give diverse responses on the test material which comprised 13 English utterances, each of which was produced with the correct stress pattern by native speakers of American English. The subject's dependence on their native language perceptual pattern (i.e., Arabic in this case) made them fail to take advantage of the auditory correlates of stress such as pitch, duration, and loudness in the most appropriate way. The subjects' task was to mark out the constituent made most prominent by the native speaker in each utterance. Their responses were then classified as normal (if stress was correctly perceived), contrastive (if it was perceived on a different constituent), or undecided (if the utterance was perceived unstressed altogether or if two constituents or more were stressed simultaneously). It turned out that the subjects sometimes perceived stress on constituents that are hardly stressed such as pronouns; determiners and prepositions made it clear that the subjects' mastery of the phenomenon under discussion was inadequate.

**2aSC9. Intelligibility of Arabs' production of English grammatical words.** Fares Mitleb (Speech and Hearing Ctr., Yarmouk Univ., Irbid, Jordan)

Features of connected speech such as the use of weak forms of grammatical words in English are said to have an important role in the communication process between native and nonnative speakers. The present study is intended to describe how intelligible is the production of certain English grammatical words for British English listeners by Arabs. Six Arabs learning English and four British speakers produced 15 sentences with 17 grammatical words for the purpose of this study. Four native British listeners were asked to judge the stimuli they heard of certain grammatical words as either having a "strong" or "weak" form. They reported that Arabs' production of strong forms was almost the same as that of native subjects. However, the Arab subjects failed to produce a high percentage of intelligible weak forms. Findings of this impressionistic study were then examined against the well-established approaches to the problem of foreign-accented speech. Our results seem to support claims about the interference of segmental knowledge onto features of connected speech since the first is generally given too much emphasis in the teaching of the target language to the extent that the latter is most neglected. [Research supported by Yarmouk University.]

**2aSC10. Training Chinese speakers to perceive English /nI/. Anna M Schmidt, Amy Kaminski, and Hyunjoo Chung** (School of Speech Pathol. and Audiol., Kent State Univ. P.O. Box 5190, Kent, OH 44242, aschmidt@kent.edu)

Native speakers of some southern Chinese dialects have difficulty perceptually differentiating English /n I/. In this study, five native Chinese subjects participated in a training experiment designed to increase perception of the /n/-/I/ contrast using a fading technique [D. Jamieson and D. Morosan, *Percept. Psychophys.* **40**, 205–215 (1986)]. Stimuli consisted of initial /n/ and /I/ words produced by four native English speakers in which the initial consonants were increased by varying amounts. Subjects were

first trained on the longest exemplars, and progressed to the natural versions. Both discrimination and identification training sets were used along with immediate feedback. Pre- and post-training perception generalization and production data were obtained. Results and implications will be discussed.

**2aSC11. The discrimination of Mandarin Chinese alveolo-palatal fricative and affricate contrasts by American English and Mandarin Chinese speakers.** Feng-Ming Tsao, Huei-Mei Liu, and Patricia K. Kuhl (Dept. of Speech and Hearing Sci., Univ. of Washington, Seattle, WA 98195)

This study examined how language experience affects the perceptual sensitivity of adult speakers to nonnative phonetic contrasts. Three of Mandarin contrasts (two alveolo-palatal fricative versus affricate contrasts and one alveolo-palatal affricate aspirated versus unaspirated contrast) were computer synthesized. For each of these three contrasts, three different variations were contrasted by varying either amplitude rise time or fricative noise duration. Both English and Mandarin speakers participated in a speech discrimination task using these stimuli. When compared to the performance of Mandarin speakers, English speakers showed poorer performance for each stimulus pair in terms of accuracy and sensitivity measures ( $d'$ ). However, both English and Mandarin speakers demonstrated similar difficulty when tested with three stimulus pairs that were not typical exemplars of Mandarin phonetic categories. The results show the effect of language experience on the discrimination of phonetic contrasts. [Research supported by NIH grant to P. K. Kuhl.]

**2aSC12. Cross-language speech intelligibility in noise: The comparison on the aspect of language dominance.** Yasue Uchida (Dept. of Otorhinolaryngology, Nagoya Univ. School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, 466-8550 Japan, yasueu@nls.go.jp), David J. Lilly (Veteran Administration Natl. Ctr. for Rehabilitative Auditory Res., Portland, OR), and Mary B. Meikle (Oregon Hearing Res. Ctr., Portland, OR)

The purpose of this study is to obtain information about the influence of language characteristics on the results of speech intelligibility in two different languages: English and Japanese. This study investigates the speech intelligibility of both English and Japanese under quiet and noisy situation on 14 bilingual subjects aged 23 to 42 years with normal hearing. As test materials, the CID W-22 word lists which are meaningful monosyllables are used for English and the 57-S word lists which are nonsense monosyllables are used for Japanese. The subjects whose dominant language is English are 6. Results show that percentage-correct of speech intelligibility is higher in Japanese than in English at lower intensity and noisy situation. This advantage in Japanese is seen regardless of the language dominance. It is considered that the phonetic characteristics and the simple structure of Japanese can make it relatively resistant to the hard listening condition. [Work supported in part by Japan Foundation for Aging and Health.]

**2aSC13. Effect of syllable structure on syllable counting in English by Japanese.** Tsuneo Yamada, Donna Erickson (National Inst. of Multimedia Education, 2-12 Wakaba, Mihama, Chiba, 261-0014 Japan), and Keiichi Tajima (ATR Human Information Processing Res. Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan)

A previous study [Erickson, Akahane-Yamada, Tajima, and Matsumoto, *Proceedings of ICPhS* (1999)] showed that Japanese listeners have difficulty correctly counting the number of syllables in spoken English words, presumably because Japanese syllable structure is more restricted than English. A complex syllable like "stress" is not allowed in Japanese, and tends to get mapped onto many syllables, such as "su.to.re.su." To

test what factors contribute to Japanese listeners' perception of syllables, a set of nonsense words varying in number and voicing of initial and final consonants, vowel type, and number of syllables, was presented auditorily to Japanese and American listeners. The task was to count the syllables in the words. American speakers counted the syllables with nearly 100% accuracy; Japanese listeners performed at about 50% accuracy. The number of initial consonants seems to have a stronger negative effect on the syllable counting performance by Japanese listeners than the number of final consonants, or vowel type. These findings suggest that onsets may play a stronger role than codas in Japanese listeners' perception of the psychological weight of syllables, contrary to current notions of phonological weight which are sensitive to codas and ignore onsets. Implications for second-language speech perception/production learning will be discussed.

**2aSC14. A computational analysis of uniqueness points for Japanese auditory word recognition.** Kiyoko Yoneyama and Keith Johnson (Dept. of Linguist., Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, yoneyama@ling.ohio-state.edu)

This paper discusses uniqueness points for all words in a 65 000-word computerized lexicon in Japanese. The uniqueness point is the point in a left-to-right scan where the word is distinguished from all other words in the lexicon. This is an important concept in the Cohort Theory of auditory word recognition. Four different analyses were conducted based on two variables: the processing unit (segment or mora) and the phonetically transcribed representation (with or without pitch information). The probability of the uniqueness point being prior to the end of the word was highest when pitch information was included and/or when words were represented in segments rather than in moras. Still, in the best combination of variables (segments with pitch information) only 48% of the words reached a uniqueness point before the end of the word. This study predicts that if Japanese listeners use small processing units and attend to pitch, recognition performance will be maximized, but for most words this optimal processing strategy will still result in a Cohort size greater than one at word's end.

**2aSC15. Correlation between vowel production and perception for a given speaker.** Jean-Marie Hombert (Lab. Dynamique du Langage, CNRS and Université Lyon 2, Institut des Sci. de l'Homme, 14 avenue Berthelot, 69363 Lyon cedex 07, France, hombert@univ-lyon2.fr) and René Carre (ENST-CNRS, 75634 Paris cedex 13, France)

The goal of this paper is to investigate the correlation between production and perception of 10 French oral vowels at the individual level. Acoustic data ( $F1$ ,  $F2$  and  $F3$ ) from 10 French native speakers were collected. Measurements were obtained from the mid-point of the vowels using cepstral analysis. Each vowel was represented by 10 measurements. The same 10 speakers were subjected to perceptual experiments in which 53 synthetic vowel stimuli covering the entire vowel space were presented through headphones. The subject's task was to identify which French word from a list of 10 words (from the same list used in the production experiment) representing the 10 French oral vowels. Each synthetic stimulus was presented 10 times in randomized order. Subject responses were mapped in a  $F1$ - $F2$  space with groupings corresponding to 90% of identical labeling. These perceptual results were compared with the measurements obtained from the production data. Our results show that the target values in the perceptual domain are correlated for a given subject with this production data. This conclusion is extremely important for understanding how individual variations within the same Linguistic Community can be the source of sound change. [Work supported by CNRS-GDR Cognition and Linguistic Diversity.]

**2aSC16. Evidence for the independence of speech categories in perception and production.** Rachel Hemphill (Dept. of Linguist., Univ. of Chicago, 1010 E. 59th St., Chicago, IL 60637, rmhemphi@midway.uchicago.edu)

Models of speech processing, which attempt to account for speech input and output in phonological or categorical terms, generally rely on one single underlying representation for both perception and production. Previous research by Hemphill [J. Acoust. Soc. Am. **104**, 1757–1758 (1998)] has suggested that two separate representations are necessary to account for asymmetries in acquisition data. Sixteen native speakers of English were trained to categorically perceive a three way voicing contrast in Thai bilabial stop consonants. The results of perception and production tests following training indicate that subjects were unable to accurately produce the contrasts that they had learned from perceptual training, despite the fact that they could identify the perceptual stimuli correctly. Additional perception testing after a two-month interval found that subjects were still able to label the training stimuli accurately, but could not identify their own productions with greater accuracy than they identified the training stimuli or the productions of unfamiliar talkers. These results are explained in terms of separate perception and production representations in the mind and an hypothesis about the reorganization of perceptual categories is explored.

**2aSC17. Branching and phrasing in Seoul Korean.** Minkyung Lee (Dept. of Linguist., Indiana Univ., Bloomington, IN 47401)

This paper examines the relationship between the structural composition, the phrasing and duration and pitch range in Korean utterances. Previous linguistic models (Selkirk, 1986) predict that the syntactically left branching (LB) and right branching (RB) structures should correspond to different phrasings, which according to previous phonological models (Jun, 1993, 1995) should be indicated by the presence of rises in fundamental frequency. Speakers were instructed to produce syntactically ambiguous phrases in a frame which highlighted the ambiguity. The experiment indicates that LB and RB cases were collapsed together resulting in the same phrasing. Even though LB and RB structures usually exhibited the same phrasing, speakers did show differences in pitch range and duration which distinguished left from right branching structures. Initial words in RB cases were longer and had a relatively larger pitch range than RB cases. Such effects have also been found as a correlate of focus. Thus speakers tend to express structural properties of utterances by means of phonetic differences in relative peak  $F0$  height and duration, apart from differences in phrasing.

**2aSC18. Tonal alignment in Seoul Korean.** Byung-Jin Lim and Kenneth de Jong (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405)

Bruce found that a local  $F0$  peak is aligned very precisely in time with the segmental material in Swedish [Swedish word accents in sentence perspective (1977)]. Alternatively, it is possible that such  $F0$  peaks may function as edge markers, and hence not necessarily be aligned with any particular aspect of the word-internal structure. This study investigates how an initial high tone is aligned in time with the segmental material in Standard Korean. Recordings of two speakers of Seoul Korean producing three-syllable words with various syllable structure combinations in two prosodic conditions were digitized and analyzed. One speaker shows no apparent pattern of alignment with the segmental material; peaks are simply reached at a fixed duration from the beginning of the utterance regardless of the structure of the word. Another speaker, however, shows that initial high tones fall into two distinct groups, ones aligned with initial syllables and ones aligned with second syllables. This syllable association is statistically related to syllable composition. These results suggest that Korean tone alignment is currently in a state of fluctuation between durationally fixed edge tones and tones associated with internal syllables according to a stochastic rule.



**2aSC19. Articulatory effects of contrastive emphasis on the Accentual Phrase in French.** Helene Loevenbruck (Institut de la Commun. Parlee, INPG, Univ. Stendhal, UPRESA CNRS 5009, 46 av. F. Viallet, F38031 Grenoble, France)

Recent works (Beckman, 1996) show that prosody is itself a complex linguistic structure, and it is imperative to better describe its phonological and phonetic (acoustic and articulatory) characteristics. Articulatory studies of French prosody provide variable conclusions. The irregularities could come from the fact that prosodic structure is rarely considered and that different phenomena (“accents primaires,” “secondaires”) are examined together. Articulatory correlates of a prosodic entity, the Accentual Phrase (AP), are studied here, using a model of French prosody (Fougeron and Jun, 1998). The AP features an initial high tone Hi, also called “accent secondaire,” a final high tone H\* (“primaire”), and two low L tones preceding them. Sentences containing four-syllable words (APs), were recorded for two French speakers (one male, one female), using EMA. The position of the AP in the sentence varied, several speaking conditions were elicited. Displacement, peak velocity, and movement duration are analyzed for the tongue-middle vertical position. The results suggest that LHi could be related to hyper-articulation of the first or second syllable, and LH\* to even stronger hyper-articulation of the last syllable. With contrastive emphasis on the AP, the initial hyper-articulation can become as strong as, and even stronger than, the final one.

**2aSC20. Acoustics of Spanish vowels as spoken in two regions of Colombia.** Dolly Urueta Mazzilli, Ruth Bahr (Dept. of Commun. Sci. & Disord., 4202 E. Fowler Ave., BEH 255, Univ. of South Florida, Tampa, FL 33620, urueta-m@luna.cas.usf.edu), and Winifred Strange (City Univ. of New York, New York, NY)

It has long been believed that Spanish dialects do not differ with respect to vowels. However, acoustic differences between Spanish dialects have been reported anecdotally. A reasonable assumption then is that, like English, Spanish dialects would vary both within and across countries. In fact, dialect differences have been found within regions of Panama [M. C. McNair, “An acoustical analysis of Panamanian vowels,” unpublished Master’s Thesis (1996), University of South Florida]. Therefore, the following analysis was conducted to determine if acoustic differences exist within Colombian Spanish. Nine male monolingual Spanish speakers from two cities of Colombia (Barranquilla and Santa Fe Bogota) produced three tokens of each of the five Spanish vowels in bilabial, alveolar, and velar environments within the frame, “Escribe CVCV bien.” Spectral and temporal measurements were calculated for each vowel in the first syllable of the word. A repeated measures ANOVA revealed a significant four way interaction suggesting that differences between dialects were dependent upon specific contexts, vowels, and formants. While vowel differences between dialects were subtle, perceptual differences in prosody and consonant production emerged.

**2aSC21. Taiwanese final stops and following initial voiced stops and nasals.** Ho-hsien Pan (Dept. of Foreign Lang. and Lit., Natl. Chiao Tung Univ., 1001 Ta-hsueh Rd., Hsinchu, Taiwan 30050)

The places of articulation of unreleased Taiwanese final stops, e.g., /p,t,k/ and glottal stop, have been known to assimilate into that of initial voiceless stops in following a syllable, depending on the speed of articulation (Peng, 1997). Besides assimilation in place of articulation, nasalization can be observed across syllables between voiced stops and nasals (Pan, 1999). However, little is known about the assimilation in place of articulation, voicing, and nasalization between Taiwanese final stops, and following voiced stops and nasals across different prosodic boundaries. This study uses EPG, airflow, and acoustical data to investigate the coarticulation between Taiwanese final stops, and initial nasals, initial voiced stops, across syllable boundary, morpheme boundary, phrase boundary of narrow focus, and intonation boundary. Preliminary results showed that Taiwanese final stops are lost in certain contexts. There is assimilation in place of articulation between unreleased final stops and initial voiced stops

and nasals. No progressive nasalization is observed between final stops and initial nasals. Taiwanese final stops are easily coarticulated with following segments. Though the presence of final stops and short syllable duration are both vital cues to Taiwanese entering tones syllables, duration is a more invariant cue than the presence of final stops. [Project supported by NSC 87-2411-H-009-008.]

**2aSC22. An acoustic analysis of Cantonese rising tones.** Connie So (Dept. of Linguist., Simon Fraser Univ., 8888 University Dr., Burnaby, BC V5A 1S6, Canada)

The two Cantonese rising tones (the High Rising and the Low Rising) are traditionally described as Tone 35 and Tone 23. Recent studies, however, show that the High Rising tone has changed from Tone 35 to Tone 25. It starts at a pitch level similar to that of the Low Rising tone. The High Rising tone is also claimed to have a long and sharp rise from the midway of its fundamental frequency pattern. The present study attempts to look for these phenomena and to describe their characteristics in the two rising tones. Native Hong Kong Cantonese speakers were asked to read two sets of Cantonese rising-tone words formed of root-words /si/ and /fu/. Fundamental frequency patterns and duration of both rising tones in these two root-words are compared and discussed.

**2aSC23. Overlap at the interface: High vowel devoicing in Japanese.** J. Kevin Varden (Dept. of English, Meiji Gakuin Univ., Minato-ku, Tokyo, 108-8636 Japan)

High vowel devoicing (HVD) in Japanese has been characterized as both phonological [J. D. McCawley, *The Phonological Component of a Grammar of Japanese* (1968)], phonetic due to gestural overlap [S.-A. Jun and M. E. Beckman, LSA Paper (1993)], and as a result of overall vowel reduction [M. Kondo, Ph.D. thesis, University of Edinburgh (1997)]. Electromyographical and laryngeal spread data from one speaker [A. Tsuchida, Ph.D. dissertation, Cornell University (1997)] supports the existence of both phonological and phonetic devoicing in complementary environments, while data from ten speakers [J. K. Varden, Ph.D. dissertation, University of Washington (1998)] indicates the presence of both types of devoicing in overlapping environments. Results of Varden (1998) also indicate high-speaker variability in rate of devoicing, duration of voicing for voiced vowels, and context-dependency. A synthesis of these studies supports a characterization of HVD as the result of varying processes, each having the same acoustic end: the devoicing and overall reduction of high vowels in devoicing environments. The data in Varden (1998) also indicate a high incidence of devoicing of /u/ in the mora [tsu] before /n/ for some speakers. This indicates relaxation of the requirement of a following underlyingly voiceless segment.

**2aSC24. Analysis of voice onset time in Georgian.** Tamra M. Wysocki (Linguist. Dept., Univ. of Chicago, 1010 E. 59th St., Chicago, IL 60637, t-wysocki@uchicago.edu)

The Georgian stop system has a three-way opposition: voiced, voiceless aspirated, and voiceless ejective. This paper reports on results from an analysis of voice onset time (VOT) in word-initial stop consonants preceding /a/. Two native speakers of Georgian (one male, one female) were recorded reading two randomized token lists: one containing nonsense syllables (e.g., “ba”) and one containing actual Georgian words (e.g., “balaxi” “grass”). Tokens were repeated twice. VOT values were measured using spectrograms and waveforms. Results show that VOT clearly distinguishes two categories in Georgian: voiced and voiceless. Initial voiced stops had short positive VOTs, while the initial voiceless stops had relatively long positive VOT values. Within the voiceless category, however, mean VOT values for ejectives and aspirates overlapped. Articulatory and acoustic characteristics of ejectives and aspirates are discussed in relation to other possible cues for distinguishing between the two types of stops. In addition to stop categories, the results show a trend toward an-

other distinction: place of articulation. Labials had the lowest VOT values, followed by coronals, and then velars. The results from these analyses are compared to studies of voice onset time in other languages.

**2aSC25. A cross-gender examination of the breathy tone in Green Mong.** Jean E. Andruski (Audiol. and Speech-Lang. Pathol., Wayne State Univ., Detroit, MI 48202) and Martha Ratliff (Wayne State Univ., Detroit, MI 48202)

Tone is generally thought of as a pitch difference that carries information regarding word meaning, but tone often correlates with phonation type as well as pitch. This paper presents the acoustic portion of a cross-gender study on acoustic and perceptual correlates of the breathy tone in Green Mong. Three male and three female native speakers of Green Mong residing in the Detroit area were asked to produce a series of target words containing the breathy tone and comparison tones in sentence context. Target words were minimal sets in which only tone varied. Each speaker produced a sequence of minimal sets containing each possible voiceless stop and voiced fricative onset of Green Mong. Vowel identity also varied across minimal sets. Four acoustic correlates of tone (fundamental frequency or  $F_0$ ,  $F_0$  contour, duration, and breathiness) were measured in each target word. Degree of breathiness was measured by subtracting the amplitude of  $F_0$  from the amplitude of the second harmonic at the durational midpoint of the vowel. Each of these four acoustic correlates will be examined in the breathy tone and the comparison tones, and a cross-gender comparison of the results will be presented.

**2aSC26. Gender differences in fundamental frequency in focused words: A case from Japanese.** Kyoko Nagao (Dept. of Linguist., Indiana Univ., Bloomington, IN 47401, knagao@indiana.edu)

When a word is focused in some context in Japanese, its fundamental frequency ( $f_0$ ) range is usually increased [Pierrehumbert and Beckman (1988)]. It is also well known that Japanese females tend to employ higher frequencies when they speak in Japanese than Caucasian females do [Yamazawa and Hollien (1990)]. However, previous studies have not examined the interactions between these two factors, focus and gender. The present study was done to determine whether the same strategy for realizing focus is employed by female and male speakers, in terms of the  $f_0$  and temporal patterns. From the analysis of 252 utterances produced by six Japanese speakers, almost the same amount of increase in absolute  $f_0$  values was observed in the speech of both sexes. Therefore, when viewed in proportion to the speakers' normal  $f_0$  range, females showed smaller increases in the proportional  $f_0$  values on the focused words than males did. The results suggest that Japanese females have less room for focus expansion because they normally use higher frequencies within their pitch range. Since temporal focus effects do not occur in female speech, it is considered that the perception of intended words as emphasized will be more ambiguous in female than in male speech.

**2aSC27. Speaking fundamental frequency of young adult Arabic men during oral reading and spontaneous speaking in both Arabic and English languages.** Ali Abu-Al-Makarem and Linda Petrosino (Dept. of Commun. Disord., Bowling Green State Univ., 200 Health Ctr., Bowling Green, OH 43403)

Currently, there is a paucity of normative data on speech and voice characteristics of different linguistic and ethnic groups. In particular, there is no available published data on the speaking fundamental frequency (SFF) characteristics of the voice of the Arabic population. The purpose of this study was to obtain preliminary data on the SFF of a group of normal speaking, young adult Arabic males. Fifteen, native Arabic, adult men served as subjects and received the identical experimental treatment. Four speech samples were collected from each subject (Arabic reading, Arabic spontaneous speech, English reading, and English spontaneous speech). Results showed the mean and SD of SFF are ([146.9, 15.4], [145.8, 13.8],

[149.1, 12.6], and [145.5, 12.0]), respectively. No significant differences were found in the mean SFF between language and type of speech, nor between languages. A significant difference in the mean SFF was found between the type of speech ( $F_{1,14}=5.51$ ,  $p=0.03$ ). Reading was significantly higher than speech. Also, Arabic men in this study had higher SFF values than previously reported for young adult males of other ethnic groups [H. Hollien and B. Jackson, *J. Phonetics* 117–120 (1973); A. Hudson and A. Holbrook, *J. Speech Hear. Res.* 197–201 (1992)]. For Speech Communication Best Student Paper Award.

**2aSC28. Effect of computer-assisted training on production of English /r/ and /l/ by Japanese.** Reiko Akahane-Yamada, Erik McDermott, Takahiro Adachi, and Tomoko Takada (ATR Human Information Processing Res. Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan)

Japanese speakers were trained to produce American English (AE) /r/ and /l/ using a computer-assisted learning system which was developed to investigate how to provide useful and effective feedback to second-language learners regarding the goodness of their production in an automatic way. In experiment 1, two groups of Japanese speakers were trained to produce /r/ and /l/ in a reproduction task. In one group, spectrographic representations with formant-tracking results overlaid were used as feedback, and in another group, acoustic scores produced by an HMM-based speech recognition system were used as feedback. Learners in both groups significantly improved from pretest to post-test in their production ability as evaluated by AE judges. Experiment 2 investigated the order effect of production training and perception training. One group of trainees received production training before perception training, and a different group received training in the opposite order. Both groups improved in their production ability by an equal amount from pretest to post-test. However, the production-to-perception training group improved more in perception ability than did the perception-to-production training group, suggesting the relevance of training order in perception ability.

**2aSC29. Speaking rate, fluency, and accentedness in monolingual English and bilingual Czech, French, and Japanese speakers.** Cliff S. Burgess (Simon Fraser Univ., 8888 University Dr., Burnaby, BC V5A 1S6, Canada)

Speaking rates (SRs) and articulation rates (ARs) were calculated for monolingual English speakers and advanced ESL learners with Czech, French, and Japanese L1s; 64 speakers in total. Data were based on two reading tasks, a cartoon description, and a personal anecdote. Bilingual speakers did these tasks in both English and their respective L1s. Short excerpts were then taken from each English language sample (264 tokens plus 64 repeats) and played to 10 native English speakers who made scalar judgments on fluency and degree of accentedness. Judgments showed a wide range of correlation with SRs and ARs that varied among language group and task type. The influence of pause type, length, and frequency was also examined. The results suggest that cues to accent and fluency perception are multidetermined and that listeners weight these cues differentially. [Work supported by SSHRC.]

**2aSC30. Phonological representation in Japanese and American rap music: Further evidence for mora timing.** Atsuko Iwai (Dokkyo Univ., 1-1, Gakuen-cho, Soka, Saitama, 340-0042 Japan) and Takashi Otake (Dokkyo Univ., Saitama, 340-0042 Japan)

This paper investigated how words in Japanese and American rap music were represented in terms of two phonological units, morae and syllables. The main question addressed in this study was to find out which unit was preferred to represent words in Japanese and American rap music. One hundred CVN syllables in two or multi-syllable words were identified from 9 pieces of 2 Japanese musical bands and 27 pieces of 3 American musical bands. Then, the first author set these syllables to music

one by one, using a Japanese computer music software to determine how these syllables were assigned to notes. The analysis showed that 73% of the CVN syllables in Japanese rap music were assigned to a note which was two times longer than that of the following note and 27% were assigned to two separate equivalent notes, while 93% of the CVN syllables in American rap music were assigned to a single note which was equivalent to the following note. In other words, the computation of duration was considered to choose notes in the former, while it was not in the latter, suggesting that words in Japanese and American rap music were represented in terms of morae and syllables, respectively.

**2aSC31. Focus on Japanese-accented English interrogatives and declaratives.** Laura G. Knudsen (Dept. of Linguist., Indiana Univ., Memorial Hall 317, Bloomington, IN 47405, lwright@indiana.edu)

This paper presents the results of a study of focus realization in Japanese English intonation, offering a modified and expanded replication of an earlier study [M. Ueyama and S.-A. Jun, UCLA Phonet. Lab Wkg. Paper No. 94 (1996)]. Fourteen subjects and three controls read two sets of test interrogatives and declaratives, one with fixed utterance length but focus in varying locations, and one with fixed focus location but varying utterance length after focus. F<sub>0</sub> and absolute time measurements were taken at four or more points in each utterance, yielding measurements of focus contour, high- or low-plateau length, and utterance-final contour. Each test item was also analyzed qualitatively and coded phonologically using the ToBI transcription system. The current study includes twelve female and male EFL subjects in four levels, plus two more advanced ESL subjects in a fifth level. Results showed first of all that Japanese acquirers of English had large gaps in their inventory of intonational types. More advanced speakers did show evidence of having acquired aspects of the English intonational contours; however, not employing them as do native English speakers.

**2aSC32. The effect of the Lombard reflex on speech produced by Cantonese speakers of English.** Herman Chi Nin Li (Dept. of Linguist., Simon Fraser Univ., 8888 University Dr., Burnaby, BC V5A 1S6, Canada, clia@sfu.ca)

The Lombard reflex is a phenomenon according to which, in the presence of noise, people unconsciously raise their vocal levels in order to compensate for the level of the ambient noise. It is well documented that speech produced in noisy environments exhibits increases in duration, fundamental frequency, and amplitude. Most studies associated with the Lombard reflex have been carried out with native speakers, but research with second language speakers is much less common. The present study describes a pilot project that aims at studying the effect of cafeteria noise on the production of native English and Cantonese-accented speech. Participants read aloud to the experimenter a set of simple true/false English statements under both no-noise and noise conditions. In the latter, masking noise of 70 dB SPL was fed through headphones to the speaker to elicit

the Lombard speech. The characteristics of the sentence-length utterances produced under quiet and noise conditions by the two speaker groups, and by male and female speakers are discussed. [Work supported in part by the SSHRC.]

**2aSC33. Acquiring dynamic coarticulatory information.** Eunjin Oh (Dept. of Linguist., Stanford Univ., Stanford, CA 94305)

This paper investigates the acquisition of dynamic coarticulation information in French [du] by native English speakers. English [du] characteristically shows a concave-downward trajectory, while French [du] shows a concave-upward transition. As an index of coarticulation degree, an absolute understoot measurement [i.e., difference between the undershoot  $F_{2v}$  (the  $F_2$  frequency at the steady state of the vowel) and the estimated target value] is potentially problematic since  $F_2$  transitions with the same  $F_{2v}$  values would be treated as equivalent even when their dynamic natures are quite different [C. S. Crowther, UCLA Working Papers in Phonetics 88, 127–148 (1994)]. The learning effects of the dynamic coarticulatory information were quantified by calculating the value of the second derivative of the  $F_2$  transition curve and comparing the values of their English and French transition curve. The preliminary results showed (i) the dynamic coarticulatory information can be acquired by L2 learners; (ii) the advanced learners showed closer approximation to the native speakers in the value of the second derivative for  $F_2$  transition curve than the beginning learners. The results suggest that the L2 learners are sensitive to the detailed acoustic parameters and could progressively acquire the L2-specific dynamic patterns of coarticulation.

**2aSC34. Vowel epenthesis in productions of English consonant clusters by Japanese.** Keiichi Tajima and Rieko Kubo (ATR Human Information Processing Res. Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan)

Phonotactic constraints on syllable structure vary across languages. Japanese has a more restricted set of consonant clusters (e.g., ‘honda’) than English (e.g., ‘instruct’). This presents a problem for Japanese learners of English, who often avoid consonant clusters by inserting epenthetic vowels between consonants. This study examines how often Japanese speakers produce epenthetic vowels in English and the phonetic environments in which they commonly occur. A group of native Japanese speakers, and a control group of native English speakers, read English words embedded in a sentence, and isolated words spoken at normal and slow rates. Measurements were made from spectrograms of the duration of the target clusters, including any epenthetic vowels produced. Preliminary results suggest that Japanese speakers show more instances of epenthetic vowels than do English speakers, and that such vowels are more commonly observed in slow, careful speech than in normal-rate speech. Also, in the absence of vowel epenthesis, Japanese speakers’ productions of consonant clusters differ from those of native speakers in the durational characteristics of component segments. Thus, Japanese speakers’ difficulties in producing English consonant clusters are reflected not just through frequent vowel epenthesis, but also through differences in durational properties in the clusters.

**Session 2aSP****Signal Processing in Acoustics: Time-Frequency Applications in Acoustics I**

Patrick J. Loughlin, Chair

*Department of Electrical Engineering, University of Pittsburgh, 348 Benedum Hall, Pittsburgh, Pennsylvania 15261***Chair's Introduction—8:25*****Invited Papers*****8:30****2aSP1. Adapting time-frequency responses to speech and acoustic classification tasks.** Les E. Atlas (Dept. of Elec. Eng., Univ. of Washington, Box 352500, Seattle, WA 98195-2500, atlas@ee.washington.edu)

Much past effort, including our own, has gone into the design of high-resolution time-frequency representations. However, as found in most data-trained automatic classification applications, high resolution can significantly degrade performance. This behavior can be understood from the standpoint of complexity: The higher the resolution and complexity of the time-frequency representation, the larger the size of the training data needed for an accurate classifier. For most applications, where many regions in time and frequency are not salient to the classification task at hand, lowering resolution in these regions should actually improve classifier performance and generality. A new time-frequency approach, called "class-dependent kernels," which selectively smoothes in time and frequency, has been developed. This approach can be summarized by considering the points of a discrete auto-ambiguity function, which is the two-dimensional Fourier transform of a time-frequency representation, as a set of features, which can be ranked in terms of saliency to the particular classification task. Keeping and classifying based on only a small number of these ambiguity plane points corresponds to a flexible and data-adaptive smoothing of the corresponding time-frequency plane. This approach will be shown to be useful for a wide range of acoustic and vibration problems including monitoring acoustic emissions in titanium removal and as an adjunct to cepstra in speaker-independent phoneme recognition.

**8:50****2aSP2. Principal features for nonstationary signals from moments of the singular value decomposition of Cohen–Posch (positive time-frequency) distributions.** Dale Groutage (Naval Surface Warfare Ctr., Puget Sound Detachment, 530 Farragut Ave., Bremerton, WA 98314-5215)

This paper presents a new method for determining the principal features of a nonstationary time series process based on the singular value decomposition (SVD) of the Cohen–Posch positive time-frequency distribution. This new method uses density functions derived from the SVD singular vectors to generate moments that associate with the principal features of the nonstationary process. Since the SVD singular vectors are orthonormal, the vectors whose elements are composed of the squared-elements of the SVD vectors are discrete density functions. Moments generated from these density functions are the principal features of the nonstationary time series process. These descriptive features can be used in a variety of ways to gain information about the specific aspects of the time series process. This new technique has application to a broad area of fields and applications such as the medical field (noninvasive diagnostics and condition classification), the defense industry (classification of acoustics transient signals), the field of machinery diagnostics, the field of marine mammal acoustics, and the manufacturing industry. An example is presented which illustrates how the new technique can be used to classify acoustic transient signals from underwater vehicles.

**9:10****2aSP3. Cross-spectral methods with applications to speech processing.** Douglas J. Nelson (U.S. DoD, 9800 Savage Rd., Fort Meade, MD 20755)

Cross-spectral methods were first presented in the mid-1980's as a method for accurately estimating stable parameters such as modem baud rate and carrier frequencies. The stability of these signal parameters makes it possible to integrate large amounts of data to accurately estimate parameters even under degraded conditions. Since biological signals, such as speech, are not stationary, classical analysis methods, including normal cross-spectral methods, are poorly suited to the problem. Presented here are methods which take advantage of the structure of speech and the phase properties of the Fourier transform. They are based on the cross-spectral



methods of the 80's, but have the advantage that these newer methods provide good accuracy and resolution for nonstationary signals such as speech. In addition, they provide a simple method for taking advantage of signal structure, such as the harmonic properties of speech, which results from the quasi-periodic pitch excitation. Specific problems addressed are accurate pitch and formant estimation, and problems such as blind recovery of carrier frequencies for single side band AM radio transmissions.

**9:30–9:50 Break**

**9:50**

**2aSP4. Time-varying coherent AM–FM demodulation and denoising of acoustic signals.** Patrick J. Loughlin and Ferhat Cakrak (Dept. of Elec. Eng., Univ. of Pittsburgh, Pittsburgh, PA 15261)

Noise removal via linear time-invariant (LTI) filtering is most effective when the signal and noise spectra have minimal overlap in frequency. In particular, it can be difficult to extract, via LTI filtering, broadband signals from broadband noise, because often their spectra overlap. However, many broadband signals are locally narrow band (e.g., AM–FM signals with large FM and moderate to small AM), and this characteristic can be exploited to improve noise suppression for such signals. We present a method for extracting locally narrow band signals from broadband noise, based on an AM–FM decomposition of the signal and time-varying filtering. The center frequency and passband of a linear time-varying filter are determined from estimates of the instantaneous frequency and instantaneous bandwidth of the signal. Results on both synthetic signals and recorded whale sounds in ambient noise demonstrate a significant improvement in SNR compared to LTI-based filtering. [Supported by ONR Grant N00014-98-1-0680.]

**10:10**

**2aSP5. Reconstruction of Formula 1 engine instantaneous speed by acoustic emission analysis.** Giorgio Rizzoni (Dept. of Mech. Eng. and Ctr. for Automotive Res., The Ohio State Univ., 206 W. 18th Ave., Columbus, OH 43210-1107, rizzoni.1@osu.edu)

This talk presents some results of a method aimed at extracting instantaneous engine speed information from acoustic emission measurements obtained from Formula 1 (F1) vehicles during a race, using joint time-frequency analysis methods. The analysis method used in this work is applied to acoustic emission data recorded by the microphone of the in-car cameras mounted on F1 vehicles. The data analyzed were acquired during the 1998 Grand Prix of San Marino (Imola), and pertain to the performance of the Ferrari and McLaren–Mercedes vehicles. The analysis presented in the paper is based on data acquired in three different sections of the Imola circuit: the starting straightaway and two curves, to highlight the capabilities of the method. The result of the analysis demonstrates that it is possible to estimate a number of useful variables from sound measurements using joint time-frequency analysis methods. These estimates are related to engine performance (e.g., engine speed and its acceleration, top engine speed), to engine architecture (e.g., gear ratios), to driving strategy (e.g., shifting strategy), and to vehicle performance (tire adhesion, aerodynamic behavior). The analysis includes validation against engine speed data obtained via telemetry measurements.

**10:30**

**2aSP6. Applications of time-frequency analysis in musical acoustics.** Gregory H. Wakefield, Maureen Melody, Rowena Guevara, and William Pielemeier (Dept. of Elec. Eng. and Computer Sci., Univ. of Michigan, Ann Arbor, MI 48109-2122, ghw@eecs.umich.edu)

The modal distribution (MD) is a member of Cohen's class of time-frequency distributions that is designed specifically for signals that are well-modeled as the sum of time-varying partials, such as those generated by many musical instruments [W. Pielemeier and G. H. Wakefield, *J. Acoust. Soc. Am.* **99**, 2382–2396 (1996)]. When combined with local operators on the time-frequency surface, MD analysis provides substantial improvement over techniques based on the spectrogram with respect to simultaneously resolving variations in amplitude and frequency. Furthermore, the analysis degrades gracefully as the acoustic signal varies from the specified model and, in several cases, can be extended to handle a broader class of signals. Examples drawn from violin vibrato, singing, and the piano are used to illustrate MD analysis and its extensions. Issues of system identification, time-varying signal models, and musical synthesis are also discussed on the basis of these examples. [Research supported by grants from the Ford Motor Company, the Office of Naval Research, the National Science Foundation, and the Office of the Vice President for Research at the University of Michigan.]

## Session 2aUW

## Underwater Acoustics: Scattering

John Oeschger, Cochair

*Coastal Systems, Station Code R21, 6703 West Highway 98, Panama City, Florida 32407-7001*

Christopher Feuillade, Cochair

*Naval Research Laboratory, Stennis Space Center, Mississippi 39529*

## Contributed Papers

8:15

**2aUW1. Acoustic band gaps and localization in water with air-cylinders.** Emile Hoskinson (Dept. of Phys., Univ. of Berkeley, CA) and Zhen Ye (Natl. Central Univ., Chungli, Taiwan, ROC)

Multiple scattering of waves may lead to many peculiar phenomena such as complete band gaps in periodic structures and wave localization in disordered media. Within the gaps excitations are evanescent, when localized waves remain confined in space until dissipated. Here we show that waves are not always localized in 2-D, but when localized there is a coincidence between the gaps in periodic structures and the ranges of localized states in the case of acoustic propagation through water containing an array of air-cylinders. We report that inside the gap or localization regime, an interesting collective behavior of the cylinders appears. Exact numerical calculations reveal the unexpected result that localization is relatively independent of the precise location or organization of the scatterers.

8:30

**2aUW2. Validity of the sonar equation and Babinet's principle for object scattering in a shallow water waveguide.** Nicholas C. Makris, Purnima Ratila, and Yisan Lai (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

It has recently been shown that the sonar equation can provide a reasonable approximation to nonforward scattering from a noncompact ( $ka > 1$ ) sphere submerged in an ocean waveguide. The omnidirectional scattering characteristics of the sphere in nonforward directions enables the sphere's far-field plane wave scattering function to be factored in the single-scatter waveguide solution just as it is in free space. The highly directional nature of the sphere's scattering in the forward direction, however, prevents this factorization and leads to significant departures from sonar equation predictions in the forward direction. By Babinet's principle, the forward scattered field from a noncompact sphere subject to an incident plane wave in free space will be nearly the same as that of a disk of the same projected area. Accordingly, it is shown that the sonar equation can significantly overestimate the field scattered from an upright plate or disk submerged in an ocean waveguide. These flat objects yield some of the most directional scattering possible from a finite body. It is also shown that in an ocean waveguide the upright disk and an equivalently located sphere with the same great circle area can have nearly identical forward scattered fields, just as in free space.

8:45

**2aUW3. Nonlinear scattering of crossed ultrasonic beams in the presence of turbulence: Multiple scattering effects.** Rebecca A. Manry and Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402, korman@nadm.navy.mil)

The nonlinear scattering of two finite-amplitude mutually perpendicular crossed ultrasonic beams—interacting in the presence of turbulence—generates a scattered sum frequency component that radiates outside the

interaction region. In the absence of turbulence, virtually no scattering, at the sum frequency, is observed [M. S. Korman and R. T. Beyer, *J. Acoust. Soc. Am.* **84**, 339–349 (1988); **85**, 611–620 (1989)]. Here, two primary cw ( $f_1 = 2.05$  MHz and  $f_2 = 1.95$  MHz) beams are generated by 2.54-cm-diam circular plane array piston transducer units ( $T_1$  and  $T_2$ ). A 4-MHz receiving unit ( $R$ ) is of similar construction. All beam axes form a common plane and overlap region with the axis of a submerged circular water jet, which generates the turbulence. With  $R$  fixed,  $T_1$  and  $T_2$  rotate on radius arms—always keeping the beams perpendicular. Symmetry suggests an angle  $\theta_*$ , where  $\theta_* = 0^\circ$  defines forward scattering. This geometry allows a nonlinear forward scattering intensity component to exist without concern for a coherent component. Here, multiple scattering effects are needed to predict their results. Supplemental scattering experiments will be presented in an attempt to identify a transition from single (Born approximation) to multiple scattering. [Work supported by Naval Academy Research Council.]

9:00

**2aUW4. An indirect method for computing troublesome coefficients in the boundary integral equation method.** Ronald T. Kessel (Defence Res. Establishment Atlantic, 9 Grove St., P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

The boundary integral equation (BIE) method is ideal for modeling scattering of waves from arbitrarily shaped objects embedded in unbounded domains. Its singular or “self”-coefficients require special numerical treatment because the kernel of the integral equation exhibits the singularity of the Green's function for a fundamental point source. Regularization for direct numerical integration is particularly complicated for elastic waves in layered media where the Green's function is a tensor whose elements cannot be written in closed form. Here the singular coefficients are computed indirectly, by inferring their values from a family of related propagation scenarios whose solutions are known in advance by virtue of Huygens' principle. The method can be applied quite generally to compute other potentially troublesome coefficients, which may occur in thin platelike objects, for example, when one boundary element lies close and parallel to another on the opposite face, or in horizontally stratified media when one element lies directly above another, making the Green's function poorly convergent. The method is verified here for scattering from elastic spheres (low to moderate  $ka$ ), and is demonstrated for scattering from an ice keel in floating sea ice, and for a sphere half-buried in seafloor sediments.

9:15

**2aUW5. Scattering enhancements for penetrable tilted circular cylinders in water: The computed evolution away from the meridional plane.** Philip L. Marston and Florian J. Blonigen (Phys. Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

Significant contributions to the high-frequency backscattering by penetrable bluntly truncated solid circular cylinders include meridional leaky Rayleigh waves [K. Gipson, Ph.D. thesis, Washington State University (1998)] and the caustic merging transition [F. J. Blonigen and P. L. Mar-

ston, J. Acoust. Soc. Am. **102**, 3088 (1997)]. The latter is important for plastic objects (where Rayleigh waves become subsonic relative to water) and is associated with merging transmitted and internally reflected rainbow rays. One way to explore the dependence of these processes on tilt is to compute the exact scattering by infinite circular cylinders away from the meridional plane. The meridional Rayleigh-wave feature for infinite metallic cylinders is a dip in the total scattering and a peak in the background-subtracted scattering [P. L. Marston, J. Acoust. Soc. Am. **102**, 358–369 (1997)]. We find it evolves in a way bounded by the locus of those rays reflected with their local angle of incidence matching the Rayleigh-wave coupling angle. In the caustic merging case, the evolution of the rainbow enhancement is similar to optical observations [Mount *et al.*, Appl. Opt. **37**, 1534–1539 (1998)]. [Supported by the Office of Naval Research.]

9:30

**2aUW6. Backscattering enhancement mechanisms for thin tilted circular plates in water.** Brian T. Hefner and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

A thin tilted circular glass plate in water was illuminated with high-frequency sound. When the angle of incidence along a diameter of the plate corresponds to the extensional and flexural wave coupling angles, there is an enhancement in the backscattered response. At each of these angles, the largest response arises from a leaky wave which travels along the diameter of the tilted plate and reflects from the plate edge. A similar enhancement was studied previously for the scattering of sound from the flat end of a tilted solid cylinder and, in that case, the reflection of the leaky Rayleigh wave was modeled using methods in ray theory [K. Gipson and P. L. Marston, J. Acoust. Soc. Am. **104**, 1754 (1998)]. For the circular plate, in addition to the scattering associated with this central ray, further enhancement was found near the extensional wave coupling angle associated with rays traveling along off-center paths. The mechanism for this response is believed to depend upon multiple reflections around the rim due to the geometry of the scatterer and may have a significant effect on the overall response of the scatterer. [This work is supported by the Office of Naval Research.]

9:45

**2aUW7. Anderson (1950) revisited.** C. Feuillade (Naval Res. Lab., Stennis Space Center, MS 39529-5004) and C. S. Clay (Univ. of Wisconsin-Madison, Madison, WI 53706)

The Anderson fluid sphere scattering model [J. Acoust. Soc. Am. **22**, 426–431 (1950)] has been reexamined to clarify three issues which have been the source of misunderstanding among underwater acousticians. First, the accuracy of the Morse large-range approximation for the spherical Hankel functions is investigated. It is shown that the minimum range for use of the approximation is strongly mode-number dependent, and should be carefully evaluated in short-range and/or high-frequency applications. Second, the precise characterization of the forward scatter region is studied. When the scattered field and the incident plane wave are combined, it is shown that little advantage is obtained in detection and localization applications by using forward scattering, rather than backscattering. Third, the translational response, or “rebound,” of the sphere under the action of the incident field is examined. By demonstrating that Anderson’s theory is a limiting case of Faran’s scattering model [J. Acoust. Soc. Am. **23**, 405–418 (1951)] for an elastic sphere, which contains the rebound response, it is shown that the response is completely explainable within Anderson’s theory, and is consistent with a description which uses a normal mode expansion around a fixed origin. [Work supported by ONR.]

10:00

**2aUW8. Surface backscattering statistics for observations conducted near Kiel, Germany.** Timothy H. Ruppel, Christopher Feuillade, and Stephen J. Stanic (Naval Res. Lab., Stennis Space Center, MS 39529)

In May 1993, an experiment was performed off the coast of Kiel, Germany to study the temporal variability of backscattering from the ocean boundaries. In the sea-surface backscattering component of the experiment, two dual-mode acoustic transducers were used to ensonify the surface with 1 or 3 ms pings at 20–90 kHz and incident angles of 65.1°–80.9° from normal. The transducers also recorded the backscattered sound. Analysis of the overall returns shows that they vary significantly from ping to ping, such that the distribution of backscattered energy appears nearly Gaussian. However, close comparison of all the signals from a given run indicates the presence of strong salient patterns in some cases, which suggest scattering from individual features (presumed to be wave crests) which move through the beam pattern from ping to ping. The statistics of the backscattered signal from these features, and their relation to the overall backscattered energy statistics, is the topic of this paper. [This work was funded by the Office of Naval Research.]

10:15–10:30 Break

10:30

**2aUW9. Formulation of the interaction of guided waves with buried structures in the shallow ocean.** M. F. Werby (NRL Code 7181, Stennis Space Center, MS 39529)

In this work the interaction of a guided wave with a buried object is formulated based on a unified propagation-object scatter model. The model makes use of a range-dependent normal mode description of propagation and a  $t$ -matrix approach to scattering. In a sediment, certain modes may propagate into the bottom, while others never penetrate and the remaining modes attenuate quickly into the bottom. These facts are used to derive a mathematically plausible method that also has the advantage of yielding intuitive information. Here ocean refraction, boundary interaction at the water-bottom interface, and absorption in the sediment are taken into account. Since this formulation makes use of general  $t$  matrices there is little limitation on the type of targets that may be considered at the bottom. A result from this formulation is the expected angular spectrum and their weights, which are useful for the experimentalist.

10:45

**2aUW10. High-frequency acoustic scattering from medium variability.** John Oeschger (Coastal Systems Station, Code R21, Panama City, FL 32407)

Laboratory results of high-frequency acoustic scattering measurements from thermal and saline driven plumes are reported. These experiments used broadband signals in a multi-static configuration. Far-field weak scattering theory describes the acoustic scatter from the medium variability. Expressions relating the acoustic scatter to the scattering field are given in terms of deviations from ambient of temperature and salinity or alternatively, sound speed and density. The angular dependence of the coefficient for the salinity difference has a stronger angular dependence than does the coefficient for the temperature difference. By using the common Bragg wave number comparisons [J. Oeschger, Proc. of 16th ICA and 135th ASA **2**, 1335–1336 (1998)], multi-static measurements can be used to determine the source of scattering, either salinity or thermal in nature, without resorting to *a priori* information. Each angular dependent source-term normalization factor is applied in turn to the data and the one that yields the best agreement predicts the scattering source correctly. Results are also presented for medium mixing where the relative density fluctuations dominate the scattering mechanism. Again the common Bragg wave number comparisons using the multi-static data are able to determine correctly the source of scattering as predicted by supporting environmental measurements.

11:00

**2aUW11. Parabolic equation techniques for diffraction and scattering problems.** Michael J. Mills and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

The parabolic equation method has proven to be useful for solving single-scattering problems [M. D. Collins and R. B. Evans, *J. Acoust. Soc. Am.* **91**, 1357–1368 (1992)]. Parabolic equation techniques for more difficult scattering problems will be presented. An approach involving parabolic approximations for range derivatives and an iteration scheme for the reflected and transmitted fields has been developed for solving problems such as the Sommerfeld diffraction problem and scattering from baffled membranes [Kriegsmann *et al.*, *J. Acoust. Soc. Am.* **75**, 685–694 (1984)]. This approach has been generalized to include multiple scattering events, with applications such as scattering from compact objects. The parabolic equation techniques will be described and illustrated with examples. [Work supported by ONR.]

11:15

**2aUW12. Bistatic scattering: A new way to improve sonar detection capabilities?** Francoise Schmitt and Franck Daout (Ecole navale, batiment des laboratoires, BP 600, Lanveoc Poulmic, 29 240 Brest naval, France)

In the field of seabed sonar imagery, it is necessary to establish local scattering models to improve the performances of the detection or recognition algorithms. In this paper, we present the Probability Density Function (PDF) of the acoustic intensity scattered by a natural profile as a function of the bistatic angle. To do this we have developed a 1-D bistatic scattering model called NEWS (Numerical Estimation for Waves Scattering) that incorporates physical phenomena like multiple reflections, shadow and the reflection coefficient of the profile. Moreover, NEWS takes into account acquisition parameters like sensors characteristics and their positions in relation to the center of the illuminated area. Gaussian spectra for the profile height fluctuation are considered. Five hundred pro-

files are generated. For each profile NEWS's algorithm gives the angular distribution of the scattered field in amplitude and in phase for all geometries and as a function of incident and scattered wave. The acoustic intensity is then treated as a random variable, and histograms are established. The PDF of the scattered intensity is compared to the K, Weibull and lognormal distributions and we examine the statistical informations providing by bistatic sonar.

11:30

**2aUW13. Building a bottom scatter database from measured bottom scatter data in the eastern Mediterranean Sea.** Peter Neumann and Gregory Muncill (Planning Systems, Inc., 7923 Jones Branch Dr., McLean, VA 22102-3304, pneumann@plansys.com)

The current U.S. Navy Standard bottom scatter database for low frequencies (up to 5 kHz) is Lambert's Law with a mu coefficient of  $-27$  dB for the world. In recent years, work to characterize scattering from both rough interfaces and the sub-bottom volume has greatly increased the understanding of acoustic interaction with the ocean bottom. However, the analysis of measured bottom scatter data requires an approach that includes the effects of the measurement geometry, higher-order multipaths, and contributions from the water-sediment interface, sub-bottom volume, and the sediment-basement interface. The SCARAB (scattering and reverberation from the bottom) model along with an automated inversion approach (simulated annealing with a downhill simplex) is designed to invert measured bottom scatter data for seven bottom scatter parameters. These seven parameters characterize the water-sediment interface, sub-bottom volume, and sediment-basement interface using a spectral representation for each. The SCARAB model is being used to analyze bottom scatter data measured by NAVOCEANO in the Eastern Mediterranean Sea and produce a database for prediction of bottom scattering strength for active sonar systems. The database is to be submitted to the OAML-SRB in late 1999. [This work is being sponsored by SPAWAR (PMW-185) under ONR management.]

TUESDAY AFTERNOON, 2 NOVEMBER 1999

KNOX ROOM, 3:00 TO 5:50 P.M.

### Session 2pAA

## Architectural Acoustics and Musical Acoustics: Integration of Synthesized Musical Instruments with Acoustic Environments

Richard H. Campbell, Cochair

*Bang-Campbell Associates, Box 47, Woods Hole, Massachusetts 02543*

Anthony J. Hoover, Cochair

*Cavanaugh Tocci Associates, Inc., 327F Boston Post Road, Sudbury, Massachusetts 01776*

### Invited Papers

3:00

**2pAA1. Driving the virtual orchestra with a conductor gesture interpreter.** Frederick W. Bianchi (Music Div., Humanities Dept., Worcester Polytechnic Inst., 100 Institute Rd., Worcester, MA 01609, bianchi@wpi.edu) and Richard H. Campbell (Worcester Polytechnic Inst., Worcester, MA 01609)

The interactive multichannel computer music system known as the "virtual orchestra" (VO) has been used several times in professional opera and theater as an alternative to a live pit orchestra. In each case, the VO has been operated by a "tapper" who follows the conductor and operates the computer which controls the VO. Recent research in noncontact, video-based conductor gesture interpretation has shown interesting possibilities in tempo and limited musical dynamics management when the gesture interpreter is directly connected to the VO controller. Two different approaches to conductor gesture detection will be discussed and the virtual orchestra will be demonstrated.



3:30

**2pAA2. Incorporating gestures into the musical control stream with the conductor's jacket.** Teresa Marrin Nakra (MIT Media Lab, 20 Ames St., Rm. E15-368A, Cambridge, MA 02139)

Much contemporary music is created in a studio, put to tape, and performed through speakers to a passive audience. In the process, some of the excitement and spontaneity of live performances is lost. There is arguably a growing problem that our contemporary musical forms have become so complicated that they cannot readily be performed in live concerts. One solution to this problem is to build real-time sensor-based systems for artists to interactively and spontaneously perform with. However, as has been shown by numerous others, these systems are either too simple in their functionality or require such constrained gestures that they force the performer to conform to unnatural or nonintuitive gestures. This paper presents an approach that adopts a unique synthesis-by-analysis method, whereby the signals created by numerous professional musicians are gathered under various rehearsal and performance conditions, studied, interpreted, and used to inform the choices in a real-time performance system. The final results, which will be presented in this paper, demonstrate a variety of attempts to generate more natural, expressive, and intuitive mappings between gesture and sound. We will discuss three different examples, including one taken from a conducting-style gestural metaphor, one from tai chi, and one from dance.

4:00

**2pAA3. Intelligent computer accompaniment systems.** Roger B. Dannenberg (School of Computer Sci., Carnegie Mellon Univ., Pittsburgh, PA 15213, rbd@cs.cmu.edu)

Because real-time computer music systems are automated, portable, and affordable, they are finding increasing application in live performance. Unfortunately, most computer music systems offer little more than a tape recorder in terms of their abilities to interact with live musicians. Computer accompaniment is designed to go beyond this "tape recorder" model of human-computer interaction. Computer accompaniment is a process in which a computer "listens" to a live musician, follows along in a score, and synchronizes an accompaniment score with the live player. Computer systems can reliably accompany a live musician in spite of tempo changes and wrong notes. Many variations of computer accompaniment have been developed. The original work assumed only one live musician playing a monophonic wind instrument such as a trumpet or flute. One variation is to follow the polyphonic performance of a keyboardist by sensing the motion of keys. In ensemble accompaniment, the task is to accompany multiple players. In addition to playing wrong notes, individual players might drop out or become lost, so there are interesting new issues to be dealt with. This work addresses practical problems of integrating computer-based performers with live performers.

4:30

**2pAA4. Local performance recording and reproduction: Application to a string quartet.** William M. Hartmann and Zachary A. Constan (Dept. of Phys. and Astron., Michigan State Univ., East Lansing, MI 48824)

A goal of traditional multichannel sound recording is to transport the listener into the environment of a musical performance. Ideally, the sound field at the ears of the listener resembles the sound field of a live performance, including ambiance cues to the acoustical character of the performance space. The goal of the local performance recording and reproduction technique (LPR/R) is the reverse. It attempts to transport the music into the environment of the listener. In this technique both the recording and the reproduction steps are different. The LPR/R technique requires at least one separate recording channel for each musical instrument. Recorded channels contain no crosstalk among instruments and no reflected sound from the room surfaces. On reproduction, the channels from different instruments are similarly kept separate; they are never electrically mixed. Instead, the channel(s) for each individual instrument are reproduced by a dedicated array of loudspeaker drivers, attempting to simulate the radiation pattern of the instrument. The LPR/R technique is particularly suited to small ensembles. Listening tests show that it is especially effective when the listener moves. An eight-channel application of the technique to a Mozart string quartet is described and demonstrated.

5:00

**2pAA5. Evaluation of multitrack recording practices in rock and roll.** Alexander U. Case (Fermata, 117 Atlantic Ave., North Hampton, NH 03862, alex@fermataco.com)

Pop and rock music reaches its listeners using loudspeakers as the interface between the recording medium and the acoustic environment. The art of creating pop music recordings relies on a terrific amount of audio equipment to not just capture but also enhance the musical instruments recorded. In the never-ending search for a more compelling loudspeaker playback experience, the equipment of the recording studio is often used in unexpected and creative ways. The application of signal processing to multitrack pop music occurs using approaches that are accessible, but not necessarily intuitive to the acoustician. In this presentation the snare drum is studied to provide a view of some modern multitrack music recording and mixing techniques. The sound of the drum is analyzed both before and after some typical signal processing in an attempt to quantify the motivation behind these recording practices. Microphone selection and placement priorities are discussed. The application of equalization, compression, and ambiance effects is analyzed. The perceptual significance of some processes is demonstrated through the playback of audio examples, both before and after signal processing. A live multitrack mixdown in which discreet elements of the mix can be isolated for evaluation is presented.

2p TUE. PM

5:30

**2pAA6. A probabilistic method for tracking a vocalist.** Lorin Grubb (Andersen Consulting, 1909 Andover Dr., Dover, PA 17315, lorin.grubb@ac.com)

When a musician gives a recital or concert, the music performed generally includes accompaniment. To render a good performance, the soloist and the accompanist must know the musical score and must follow the other musician's performance. Both performing and rehearsing are limited by constraints on the time and money available for bringing musicians together. Computer systems that automatically provide musical accompaniment offer an inexpensive, readily available alternative. Computer accompaniment requires software that can listen to live performers and fol-

low along in a musical score. An implemented system automatically accompanies a singer given a musical score. The focus is a method for robustly detecting the score position and tempo (performance rate) of the singer in real time. Robust score following requires combining information obtained both from analyzing a complex signal (the singer's performance) and from processing symbolic notation (the score). The singer's score position is characterized statistically using a model that combines the available musical information to produce a probabilistic position estimate. By making careful assumptions and estimating statistics from a set of actual vocal performances, a reasonable approximation of this model can be implemented in software and executed in real time during a performance.

TUESDAY AFTERNOON, 2 NOVEMBER 1999

MARION ROOM, 12:55 TO 5:00 P.M.

### Session 2pAB

## Animal Bioacoustics: Detecting and Identifying Animals Using Acoustics I

David A. Helweg, Chair

*SPAWARSYSCEN-San Diego, Code D351, 49620 Beluga Road, San Diego, California 92152*

Chair's Introduction—12:55

### Invited Papers

1:00

**2pAB1. Detection and species identification of baleen whale calls.** David A. Helweg (Code D351, SPAWARSYSCEN San Diego, 49620 Beluga Rd., San Diego, CA 92152)

Baleen whales live over extensive home range and time scales. Study of how these animals use their vocalizations for communication requires massive data sampling over long periods. This paper describes a system for automating the sampling and analysis of baleen whale calls. Many species produce stable, homogeneous call structures which lend themselves to automated species identification. We have benchmarked a series of bioacoustical call identification algorithms against a set of blue and fin whale calls while systematically manipulating the signal-to-noise ratio. Blue (*Balaenoptera musculus*) and fin (*B. physalus*) whale calls are very stereotypical. Blue whale "A" and "B" calls have fundamental frequencies of approximately 17 Hz, narrow bandwidth, well-defined harmonic structure, and typical duration of 15–25 s. Fin whale "pulses" have fundamental frequencies of approximately 17 Hz, but are broadband in nature and short (approximately 1-s) duration. The results demonstrated a typical tradeoff of speed versus accuracy. The best algorithm was inserted into an underwater sound recording system and its signal-detection theoretic performance was quantified. Results will be discussed with respect to technological, ecological, and conservation aspects of baleen whale bioacoustics. [Project CS-1082 of the Strategic Environmental Research and Development Program.]

1:20

**2pAB2. Acoustic feature extraction for characterizing and classifying animal sounds.** Kurt Fristrup (Cornell Lab. of Ornithology, 159 Sapsucker Woods Rd., Ithaca, NY 14850)

Standardized measures for animal sounds were investigated as the basis for characterizing and classifying animal sounds. These measures were computed from a noise-compensated spectrogram. The performance of this system was evaluated in terms of the stability of measurements under varying noise conditions, and the ability to correctly identify sounds to species or individual in a variety of trials. Results from analyses of bird and marine mammal sounds will be presented. General considerations regarding feature and classifier selection will be discussed.

1:40

**2pAB3. Using vocal identifiers of individuality in research.** Robert Gisiner (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217-5660, gisiner@onr.navy.mil)

Many animals use vocal cues for individual recognition to mediate interactions between parents and offspring, between mates or between territorial neighbors. These cues or other features of vocalizations with no apparent social role can also be used by researchers to identify individuals or estimate the gender, age, size or other features of the vocalizing individual. In order to establish the reliability of the vocal cue the researcher must first have access to a pool of individuals of known identity, or gender, size, etc. Simple

statistical measures such as the within and between measures of variance from an ANOVA can provide a quantitative measure of the potential utility of a vocalization as an identifier. Sorting of individuals can often be accomplished with a discriminant function analysis using relatively few features (three to five may be all that are needed), but more often a simple neural network classifier can achieve comparable classification results without preselecting acoustic features for sampling. Field application of a trained classifier can greatly aid research requiring rapid identification of individuals when visual identification is difficult or impossible.

### Contributed Papers

2:00

**2pAB4. Preliminary results on the acoustic characterization of the Northern Right whale.** Francine Desharnais and Mark G. Hazen (Def. Res. Est. Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, desharnais@drea.dnd.ca)

The Northern Right whale (*Eubalaena glacialis*) is the most endangered large whale species. Collisions with ships is the cause of a large percentage of documented deaths. An experiment is planned for the summer of 1999 to record vocalizations of the Northern Right Whale. The recordings will be made in the Bay of Fundy (Southeast of Grand Manan Island) and Roseway Basin (South of Nova Scotia). These areas were chosen because of the large occurrence of whale sightings in the summer months. Preliminary analysis of acoustic signals from the Right Whale will be shown. The statistical analysis of these signals should eventually allow the development of processing algorithms to recognize whale calls. Identification of unique features in the whale calls may allow localization and tracking of the Northern Right whales with a field of sonobuoys.

2:15

**2pAB5. The dependence of target strength of the northern right whale (*Eubalaena glacialis*) on the acoustic properties of blubber.** James H. Miller, Thomas Weber, Angela Tuttle (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882, miller@uri.edu), and David C. Potter (Northeast Fisheries Sci. Ctr., NMFS, Woods Hole, MA 02543)

Whale detection and tracking using active sonar is a subject of recent study [J. H. Miller, D. C. Potter, T. Weber, and J. Felix, *J. Acoust. Soc. Am.* **105**, 992 (1999)]. In that previous work, the measured target strength of a northern right whale (*Eubalaena glacialis*) was found to be less than that measured for a humpback whale (*Megaptera novaeangliae*) of the same size and aspect (about 0 and 5 dB, respectively). The difference was explained by modeling the thicker blubber of the right whale as a lossy layer in a plane wave reflection coefficient. However, values for the acoustic properties of the blubber were estimated from tissue properties of other mammals [R. C. Chivers and R. J. Parry, *J. Acoust. Soc. Am.* **63** (1978)]. In April, 1999, a 60-ton northern right whale named Staccato was found dead near Cape Cod. Blubber samples were acquired by the National Marine Fisheries Service and acoustic properties determined using a multi-sensor core logger in URI Marine Geomechanics Laboratory. Sound speed and density data have been used to refine the plane wave reflection coefficient model and corroborate the earlier work. [Work supported by NMFS.]

2:30

**2pAB6. Visual and acoustic surveys of whales: A Monte Carlo model.** Kathleen J. Vigness (Marine Acoust., Inc., 901 N. Stuart St., Ste. 708, Arlington, VA 22203)

There has been much interest lately in decreasing the high variability of cetacean abundance estimates by coupling passive acoustics with visual surveys. To estimate the advantages and disadvantages of a dual-mode survey, a Monte Carlo model was created. The model can reflect the species and the methodology of a particular survey. The whale population is defined by the parameters of group size and encounter rate, swimming speed and direction, surfacing rate and pattern, and source level and rate of vocalization. The survey process is constrained by the additional parameters of vessel speed, sighting probability, hydrophone array characteris-

tics, and sound propagation. With results from the model, researchers can compare the abundance estimates and variability from a visual survey only, a passive acoustic survey only, or a dual-mode survey. They can also determine which parameters are important for a species, region, and methodology, and design their surveys accordingly. Sample results from a hypothetical minke whale population demonstrate that adding passive acoustics significantly increased the number of groups detected and decreased the time the groups were in the survey area before being detected.

2:45

**2pAB7. Detection and censusing of blue whale vocalizations along the central California coast using a former SOSUS array.** Ching-Sang Chiu, Christopher W. Miller, Therese C. Moore, and Curtis A. Collins (Dept. of Oceanogr., Naval Postgrad. School, Monterey, CA 93943)

To investigate the feasibility of automating the detection and censusing of blue whale vocalizations over a large coastal region using the Naval Postgraduate School (NPS) Ocean Acoustic Observatory (OAO), a four-day experiment was conducted along Central California in the summer of 1997. During the experiment, array data were archived continuously at the NPS OAO, a former SOSUS array. In addition to shore-based acoustic monitoring, an aircraft was assigned to locate blue whales in the Monterey Bay National Marine Sanctuary, and a research vessel, manned with observers and instrumented with a towed hydrophone array, was used to confirm locations of the blue whales and classify the vocalized near-field signals. The shipboard measurements were required to provide a means to separate the source signal characteristics from their multipath signatures. The towed array data were deconvoluted, source level and characteristics were estimated, and call-to-call variability was studied. Based on robust signal parameters, two simple autodetection correlation filters for the "A" and "B" blue whale calls were constructed, respectively. By applying the autodetection filters to the OAO data, the number of blue whale calls for the summer of 1998 was tallied. [Research supported by SERDP/ONR.]

3:00–3:15 Break

3:15

**2pAB8. Geographic variations of the Hawai'ian spinner dolphin (*Stenella longirostris*) whistle repertoire: Do they exist?** Carmen Bazua-Duran and Whitlow W. L. Au (MMRP, Univ. of Hawai'i, P.O. Box 1106, Kailua, HI 96734, bazua@soest.hawaii.edu)

In the present work, groups of spinner dolphins (*Stenella longirostris*) off four of the main Hawai'ian islands (Hawai'i, Lana'i, Maui, and O'ahu) have been studied. The whistle repertoire of dolphin groups from each island will be compared to search for geographic variations. In addition, behavioral, ecological, and acoustical variables have been quantified in order to correlate the dolphin whistle repertoire with specific behavioral and ecological factors. Geographic variations in the whistle repertoire of coastal bottlenose dolphins (*Tursiops truncatus*) from the Gulf of Mexico have been found. The differences were not attributable to different behavioral states, but were related to group structure (group size and contacts established between groups) and coastal habitat. These findings on the whistle repertoire of coastal bottlenose dolphins support the hypothesis that delphinid whistles serve to coordinate group behavior. Preliminary results on group structure of spinner dolphin groups off Hawai'i suggest that geographic variations should exist. Gulf of Mexico bottlenose and Hawai'ian spinner dolphin societies share several characteristics. Thus, it

2p TUE. PM

is expected that the Hawai'ian spinner dolphin whistle repertoire will show variations as does the bottlenose dolphin whistle repertoire. [This work was supported by a Leonida Memorial Scholarship and Seed Money Grant from the University of Hawai'i.]

3:30

**2pAB9. Source level estimation of dolphin communication calls as a potential method for species identification.** Kristin Kaschner (Marine Mammal Res. Unit, Univ. of British Columbia, Hut B-3, 6248 Biol. Sciences Rd., Vancouver, BC, Canada), A. David Goodson, and Paul R. Connelly (Loughborough Univ., LE11 3TU, UK)

Research into the acoustic structure of communication calls of different odontocete species has attempted to identify distinctive features that might assist in discrimination of individuals, pods or species. Most studies have concentrated on frequency modulation characteristics, such as the general shape or contour of the call. Although the source level (SPL *re*: 1  $\mu$ Pa at 1 m) of echolocation signals has been studied and appears to reflect some body size dependency, signal intensities of "whistles" have not been investigated closely. In this study maximum source levels of communication calls were estimated in open sea conditions, using an acoustic multi-path ranging technique. Preliminary results suggest species-specific differences in maximum call intensities that may provide a useful additional cue to acoustically distinguish species in the wild. Data used for spectral and statistical analysis were recorded using a single hydrophone attached to a pelagic trawl fishing in continental shelf waters in the Bay of Biscay. Maximum SPLs of social calls were extracted and distances between hydrophone and vocalizing animals were obtained exploiting information inherent in recorded multi-path signals. Concurrent visual observation provided species identification. Source levels of calls were calculated based on a calibrated reference signal. Methodology and limitations of this technique are discussed.

3:45

**2pAB10. Vocalizations in two captive born Florida manatees (*Trichechus manatus latirostris*).** Katherine J. Frisch (Western Illinois Univ., Macomb, IL 61455-1390)

Florida manatees (*Trichechus manatus latirostris*) are an endangered species with a population of less than 3000 individuals. They are known to produce sounds. Some attempts have been made to characterize their vocal repertoire. However, little has been documented about intermanatee differences in these vocalizations. This pilot study was undertaken to examine the acoustic differences in vocalizations of two captive born manatees at Mote Marine Laboratory in Sarasota, FL. The manatees were recorded at random intervals while engaged in typical activities. The recordings were analyzed along a number of dimensions. Despite a great deal of overall variability, there was a clear bimodal distribution of frequency characteristics, particularly near the end of the vocalization. This suggests that manatee vocalizations may have an individually determined natural frequency range. These acoustic cues could provide additional means of identifying and distinguishing individuals from each other in the wild. Such a finding has important applications, for example, in the effort to preserve the species as an aid in the cataloging of individual animals.

4:00

**2pAB11. Vocal discrimination of two species of manatees (*Trichechus inunguis*) and (*T. manatus manatus*) in Brazil.** R. S. de Sousa Lima (PGECEMVS/Depto. Zoologia/Univ. Federal de Minas Gerais, Belo Horizonte, Brazil), A. P. Paglia (UFMG, Belo Horizonte, Brazil), and G. A. B. Fonseca (UFMG/CABS/Conservation Intl.)

Localizations from 15 West Indian manatees (*T.m. manatus*) and 14 Amazonian manatees (*T. inunguis*) were recorded and selected for spectral analysis. Several physical parameters of these vocalizations were used to describe and compare the vocal repertoires of these two species. Five of those variables were used in a principal components analysis (PCA) in order to verify if the vocal parameters could be grouped by species. Both approaches (repertoire comparison and PCA) discriminated these two spe-

cies. The main differences were in the limits of the fundamental frequency (range) of each species, 1.07 to 8 kHz for (*T. inunguis*) and 1.07 to 4.98 kHz for (*T.m. manatus*), and the number of notes per vocalization. Only the Amazonian manatee presented more than one note (1 to 4) per vocalization. The first axis of the PCA explained almost 70% of the data variation. The most important variables in this axis were related to the fundamental frequency. Therefore, the characteristics of the fundamental frequency are responsible for the discrimination of the vocal patterns of (*T. inunguis*) and (*T.m. manatus*). Such patterns are specie-specific and could be used as a taxonomic feature. [Work supported by FBPN, MacArthur Foundation, Conservation International, USFWS, and CNPq.]

4:15

**2pAB12. A two-stage process for automatic recognition of harbor seal vocalizations.** David K. Mellinger<sup>a)</sup> (MBARI, 7700 Sandholdt Rd., Moss Landing, CA 95039, dave@mbari.org)

Male harbor seals (*Phoca vitulina*) were continuously recorded for a year from an array of hydrophones in shallow water off the coast of central California. A two-stage automatic recognition system was used to extract sounds of interest. The first stage, fast but crude, processed the entire sound archive. It operated by: (1) making a spectrogram; (2) normalizing the spectrogram in several ways to remove some background noises and interfering sounds; and (3) detecting sounds in the 100–1000 Hz range with a minimum duration of 1 s. The second stage, slow but accurate, operated on the sounds extracted by the first stage and classified them as being either harbor seal roars or not. Classification was done by measuring a variety of acoustic characteristics—duration, frequency span, amplitude variation, etc.—in several frequency bands, and applying statistical pattern recognition techniques to the resulting feature vectors. Training data consisted of 1011 roar examples and 850 nonroar sounds. Recognition accuracy greater than 95% was achieved, with the principal errors occurring because of close resemblance between seal roars and breaking-waves. These results show that acoustic monitoring combined with automatic recognition can be a viable method for continuous monitoring of populations of wild animals. <sup>a)</sup>Currently at PMEL, 2115 S. E. OSU Dr., Newport, OR 97365.

4:30

**2pAB13. Passive detection and localization of marine mammals in the Ligurian Sea.** Angela D'Amico, Joseph E. Bondaryk, and Nicola Portunato (SACLANT Undersea Res. Ctr., Viale S. Bartolomeo 400, 19138 La Spezia, Italy, damico@saclantc.nato.int)

SACLANTCEN sponsored an environmental acoustic measurement in the Ligurian Sea in August 1999. The objectives of this cruise were to detect, localize and track marine mammals using passive acoustic techniques. During this cruise, the R.V. ALLIANCE passively operated several advanced sonar arrays near Corsica. Acoustic techniques were used to localize and track several species of cetaceans in position and depth over time from their vocalizations only. The acoustic results are compared to ground truth provided by visual observations and data from several radio tags with time-depth recorders. Statistics on whale vocalizations and behavior are tabulated. At the time of this abstract submission, it is anticipated that data on Fin, Sperm and other cetaceans will be available.

4:45

**2pAB14. Passive acoustical survey of finless porpoises in the Yangtze River, China.** Tomonari Akamatsu (National Res. Inst. Fisheries Eng., Ebudai, Hasaki, Kashima, Ibaraki, 314-0421 Japan, akamatsu@nrife.affrc.go.jp), Ding Wang, and Kexiong Wang (The Chinese Acad. of Sci., Wuhan 430072, PROC)

Finless porpoises (*Neophocaena phocaenoides*) are distributed in Asian waters. The narrow-band and high-frequency pulse sonar signals produced by this species are distinctive from background noises. Underwater sound monitoring by a hydrophone (BK8103) along board sides of a



research vessel concurrent with visual observation were conducted in the Yangtze River from Wuhan to Poyang Lake in 1998. The peak-to-peak detection threshold level was 133 dB *re*: 1  $\mu$ Pa, which could be roughly converted to the theoretical detection range of 316 m, hypothesizing the porpoise directing to the hydrophone and 168-dB source level of the sonar signal under a shallow water sound propagation (combination of spherical and cylindrical spreading). In a total of 1064-km cruise, 717 finless por-

poises were observed. The acoustical monitoring system could detect the sonar signals from finless porpoises, found within 200 m from the research vessel. The sonar signals could also be detected at night and under windy weather conditions. Basically, the acoustical observation system was operated automatically and free from the individual difference of observers. The high-frequency acoustical monitoring seems to be an effective method for the survey of small cetaceans which produce sonar signals.

TUESDAY AFTERNOON, 2 NOVEMBER 1999

KNOX ROOM, 1:00 TO 2:30 P.M.

## Session 2pBB

### Biomedical Ultrasound/Bioresponse to Vibration: Role of Vibration in Medicine

Diane Dalecki, Chair

*Department of Electrical and Computer Engineering, University of Rochester, Hopeman Building, Rochester, New York 14627*

#### Contributed Papers

1:00

**2pBB1. Lung response to low-frequency underwater sound.** Diane Dalecki, Carol H. Raeman, Sally Z. Child, Stephen A. McAleavey, and Edwin L. Carstensen (Dept. of Elec. and Computer Eng. and the Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627)

The response of the lung to exposure to low-frequency underwater sound was investigated for frequencies of 100–500 Hz. Several different experimental and theoretical approaches demonstrated that the lung responds to low-frequency underwater sound as a resonant structure. Measurements of acoustic scattering near murine and rat lung indicated that the response of the lung to low-frequency acoustic fields can be described well by the theory of linear oscillations of a bubble in water. The measured mean resonance frequency was 325 Hz for adult mice (30 g), 420 Hz for young mice (10 g), and 175 Hz for rats (320 g). Noninvasive measurements of the displacement amplitude of the lung using an ultrasonic pulse-echo ranging technique confirmed that the lung oscillates in response to exposure near the resonance frequency. At the resonance frequency the response of the lung is maximized and the thresholds for damage to the lung and surrounding tissues (such as liver) were lowest. For exposure of adult mice at the resonance frequency, the threshold for lung hemorrhage and for liver hemorrhage was  $\sim 184$  dB *re*: 1  $\mu$ Pa (i.e., 1.6 kPa).

1:15

**2pBB2. Measurements of lung vibration in response to low-frequency sound.** Stephen A. McAleavey, Carol H. Raeman, Kevin J. Parker, and Diane Dalecki (Dept. of Elec. and Computer Eng., Univ. of Rochester, Rochester, NY 14627)

The amplitude of vibration of lung in response to low-frequency (100–500 Hz) sound was measured in mice and rats *in vivo*. Low-frequency sound fields were generated in an inertial water column driven at its base with an electromagnetic transducer. Estimates of vibration amplitude of the lung were obtained by calculating the variations in the round-trip delay of pulsed wideband ultrasound bursts. The relative time delays were calculated by determining the maximum of the cross-correlation of the first echo with subsequent echoes. The ultrasound bursts were emitted at a rate of 1–5 kHz, allowing ten displacement estimates to be calculated per cycle of lung oscillation. The mean resonance frequencies obtained through measurements of vibration amplitude were 330 Hz for mice and 189 Hz for rats. The maximum observed displacements were on the order of 0.1 mm. The resonance observed through measurements of displacement amplitude was equivalent to that obtained through acoustic scattering measurements and consistent with observations of lung hemorrhage.

1:30

**2pBB3. The effect of age on vibrotactile adaptation.** Alan K. Goble (Dept. of Psych., Bennett College, 900 E. Washington St., Greensboro, NC 27401-3239, goble@bennett1.bennett.edu)

The aging process is known to affect the processing of somatosensory information. For example, a number of recent studies have shown that detection thresholds for vibrotactile stimuli are elevated at low and high frequencies in elderly individuals, as compared to comparable measurements in young adults. The present study is the first to directly measure the effect of aging on vibrotactile adaptation. Detection thresholds for 25-Hz sinusoidal vibrations presented to the thenar eminence via a 5-mm-diameter contactor were measured in a small group of senior citizens ( $n = 4$ ) before and after exposure to a 25-Hz 20-dB SL adapting stimulus and compared to comparable measurements previously obtained on younger adults ( $n = 4$ ). As expected, baseline thresholds were significantly higher in the older observers by some 12 dB ( $p < 0.0000$ ). Furthermore, older individuals exhibited significantly less adaptation than their younger counterparts (71 vs 14.4 dB,  $p < 0.0000$ ). These results provide evidence of age-related changes in both peripheral and central nervous system function, and suggest directions for future research. [This research was supported by NIH Grant No. 2 S14 GM44780-08.]

1:45

**2pBB4. Excitation and response of surface waves on isotropic and nonisotropic viscoelastic half-spaces with application to medical diagnosis.** Thomas J. Royston (Dept. of Mech. Eng., Univ. of Illinois at Chicago, Chicago, IL 60607), Hussein A. Mansy, and Richard H. Sandler (Rush Medical College, Chicago, IL 60607)

An analytical solution is proposed for the problem of surface wave generation on a viscoelastic half-space by a finite rigid circular disk located on the surface and oscillating normal to it. The solution is an incremental advancement of the work reported in two articles published in the 1950s in the Proceedings of the Royal Society. While the application focus of those articles was seismology, the application of interest here is medical diagnostics. Consequently, the solution is verified experimentally using a viscoelastic phantom with material properties that approximate biological soft tissue. Also, the effect of an inclusion on surface wave behavior within the otherwise isotropic medium is investigated. Measurement of wave motion on the skin surface caused by internal biological functions or external stimuli has been studied by a few researchers for rapid, noninvasive diagnosis of a variety of medical conditions. Conditions considered

include those associated with cardio-vascular dynamics, oedema and other skin ailments, hardened tissue regions, e.g., tumors, and extraluminal air in the abdomen, a.k.a. pneumoperitoneum. It is hoped that the developments reported here will advance these techniques and also provide insight into related diagnostic techniques, such as sonoelastic imaging and other vibro-acoustic methodologies. [Work supported by the Whitaker Foundation.]

2:00

**2pBB5. Dynamic response characteristics of upper respiratory system.** Ahmed Al-Jumaily (Diagnostics and Control Res. Ctr., Auckland Inst. of Technol., Auckland, NZ) and Ammar Al-Saffar (Univ. of Sci. and Technol., Irbid, Jordan)

A theoretical model is presented to study the general dynamic characteristics of the upper human respiratory system for a healthy as well as an unhealthy lung. The model is formulated by using perturbation technique on the fluid momentum and continuity equations taking into consideration the structural dynamic features of the air passage walls. A recursion formula is generated to compute the input impedance along the principal path. Also the distributed pressure response along a single pathway in the breathing system is investigated for different frequencies. The results indicate that for a healthy lung the magnitude of the impedance at the throat decreases to a minimum value at a frequency of 379 Hz, then increases again, and continues to fluctuate between maximum and minimum values. This prediction agrees very well with available experimental data and disputes earlier theoretical observations of double minimum values. However, it is indicated that a lung with some abnormality at a Horsfield order of 24 and above has its first minimum at a frequency of 238 Hz. A detailed

explanation of the dynamic characteristics in terms of the impedance ratio and the pressure distribution is presented in the paper.

2:15

**2pBB6. Electrical bioimpedance monitoring of cardiovascular effects of noise.** Goran Belojevic (Inst. of Public Health of Serbia "Dr Milan Jovanovic Batut," Dr Subotica St. No. 5, 11000 Belgrade, Serbia, FR Yugoslavia), Vesna Stojanov (Inst. for Cardiovascular Diseases, Belgrade, Serbia, FR Yugoslavia), Branko Jakovljevic, and Jelena Ivanovic (Belgrade Univ. School of Medicine, Belgrade, Serbia, FR Yugoslavia)

Cardiovascular effects of recorded traffic noise ( $L_{eq}=89$  dBA) were monitored with an "AVL 2001" electrical bioimpedance apparatus on 12 medical students (4 male and 8 female), and compared to quiet conditions ( $L_{eq}=40$  dBA) before and 10 min after exposure. There were significant changes of the following cardiovascular parameters in quiet-noise-quiet conditions (Mean  $\pm$  SD): cardiac work/kgm/m<sup>2</sup>/(7.0  $\pm$  2.0; 6.4  $\pm$  1.5; 7.1  $\pm$  2.1; respectively;  $P<0.01$ , Friedman ANOVA test); global flow [L/min/m<sup>2</sup>] (5.8  $\pm$  2.0; 5.4  $\pm$  1.5; 6.0  $\pm$  2.0; respectively;  $P<0.01$ ), pump output/beat [ml/m<sup>2</sup>] (75  $\pm$  26; 69  $\pm$  18; 78  $\pm$  21, respectively;  $P<0.05$ ) and peripheral vascular resistance (afterload) [ $\Omega$  m<sup>2</sup>] (1347  $\pm$  476; 1387  $\pm$  458; 1273  $\pm$  430; respectively;  $P<0.01$ ). No significant effects of noise were observed concerning pulse frequency, systolic and diastolic arterial blood pressure, preload, pump efficiency, thoracic fluids, and contractility. These results indicate that intense noise has a strong vasoconstrictive effect, which in longer terms may lead to disturbances in the regulation of arterial blood pressure and in coronary circulation.

TUESDAY AFTERNOON, 2 NOVEMBER 1999

HARRISON ROOM, 1:00 TO 4:00 P.M.

### Session 2pEA

## Engineering Acoustics, Physical Acoustics and Structural Acoustics and Vibration: Acoustic Nondestructive Evaluation: New Directions and Techniques, Part II

P. K. Raju, Chair

*Mechanical Engineering Department, Auburn University, Ross 201, Auburn, Alabama 36849-5341*

### Invited Papers

1:00

**2pEA1. Impulse response characterization of composite materials and structures for design and manufacturing.** Ronald F. Gibson (Mech. Eng. Dept., Wayne State Univ., Detroit, MI 48202, gibson@eng.wayne.edu)

Impulse response testing may be used for fast and efficient characterization of both the elastic and viscoelastic properties as well as the structural integrity of composite materials and structures. Impulsive excitation is used to induce vibrations in the structure and modal frequencies and damping factors are extracted from the response using frequency domain and/or time domain techniques. The use of measured modal frequencies in the solution to the inverse problem for composite beams or plates yields the elastic constants for the composite. A variation on this method yields the spatial distribution of fibers in the composite. For certain configurations, the damping has been found to be a good indicator of interlaminar fracture toughness of a composite. Adhesively bonded composite structures having various flaws and defects have also been tested using this method, and both frequencies and damping factors have been found to vary from those of the "good" structures. PC-based virtual instrumentation software has been developed for data acquisition and analysis, and the method has also been used to test full scale composite automotive vehicle components as well as small laboratory samples. The long range goal of the research is the development of rapid, inexpensive screening tests for composite components.

1:30

**2pEA2. Elastodynamic response of ordered materials.** Vikram K. Kinra (Aerosp. Eng., Texas A&M Univ., College Station, TX 77843-3141)

The interaction of a normally incident plane longitudinal wave with an ordered or disordered single layer of spherical inclusions embedded in a polyester matrix was measured. Inclusions were arranged in a periodic (square or hexagonal) and random arrays with area fraction ranging from 0.07 to 0.65. Measurements were carried out at wavelengths that are large, equal, or small compared to the two characteristic lengths of the composites, namely, the particle radius and the interparticle distance. The transmission and reflection

spectra for periodic layers are characterized by several resonances, the frequencies of which are close to the cut-off frequencies of the appropriate shear lattice modes. The excitation of resonances is accompanied by the propagation of mode-converted shear waves, which are propagated along certain symmetry directions within the plane of the particles. These waves were detected by the use of a shear wave transducer. At the first lattice resonance frequency, the reflection coefficient of a layer drops down to near-zero value. The vibrations of an individual particle have been measured by a laser interferometer. Finally, a simple model of the elastodynamic event was constructed. In spite of its simplicity, the model captures the essential features of the experimental data.

### Contributed Papers

2:00

**2pEA3. Differential ultrasonic stress-strain measurements.** Sissay Hailu (Dept. of Elec. Eng. and Computer Sci., CWRU, Cleveland, OH 44106), Gary R. Halford (NASA, John H. Glenn Res. Ctr., Cleveland, OH 44135), Dov Hazony, and Gerhard Welsch (CWRU, Cleveland, OH 44106)

Fatigue tests are often encumbered by lack of the specimen's exact length due to mechanical grip effects and by environmental noise factors. Such issues may be mitigated when differential measurements are undertaken between consecutive data states. Our principle-monitoring tool is an ultrasonic pulse-echo process where the primary and secondary echoes along the principle specimen's axis provide both length and gage length, respectively [D. Hazony, *Circuit Systems Signal Process.* **14**(4), 525–538 (1995)]. The process is also useful for crack detection [I. Mostafa *et al.*, *Int. J. Fracture* **85**, 99–109 (1997)]. Compared with absolute measurements, differential stress-strain monitoring is shown to be highly reproducible and more sensitive providing changes in Young's modulus and Poisson's ratio as well as a focus for the detection of an emerging crack. [Work supported by NASA.]

2:15

**2pEA4. Acoustic detection of pressure in sealed drums.** R. Daniel Costley (Miltec, Inc., Natl. Ctr. for Physical Acoust., University, MS 38677, [dcostley@mil-tec.com](mailto:dcostley@mil-tec.com)) and Mark Henderson (Mississippi State Univ., Starkville, MS 39762)

At many waste sites, transuranic (TRU), low-level, and mixed wastes are stored in 55-gallon drums. Many of these drums contain hazardous, organic wastes as well. Radiolysis or other physical or chemical processes may result in gaseous emissions inside these drums. When this happens the pressure within the drum will increase, sometimes to unacceptable levels. In more drastic cases, these emissions may produce flammable or explosive atmospheres (e.g., hydrogen from radiolysis). Currently regulatory procedures requires that each drum be individually opened and inspected for the presence of hazardous organic waste. This situation will be dangerous for workers. A noninvasive technique has been developed which will detect an increase in pressure over ambient levels and alert workers of potential danger and greatly increase safety. When a drum lid is tapped, it vibrates at specific frequencies. It turns out that the natural frequencies of vibration of the drum lid increase as the pressure inside the drum increases. Thus the pressure within the drum can be determined by measuring the frequency at which the drum lid vibrates. Experimental results and plans to incorporate this into a simple handheld device will be discussed. [Work supported by DOE.]

2:30–2:45 Break

2:45

**2pEA5. Acoustic sensors and NDE techniques for process control in drug, cosmetic and food manufacturing industries.** Hasson M. Tavossi (Dept. of Eng. Sci. and Mech., The Penn State Univ., 227 Hammond Bldg., University Park, PA 16802-1401)

New techniques of acoustic sensors for process control in manufacturing industries are presented. Two practical applications of acoustic sensors are considered. The first case involves remote determination of bulk-temperature of a paste heated in an agitated mixer, when other means for temperature measurement are not tolerated. During chemical processing the precise value of bulk-temperature of the paste must be known for

quality control of the final product. To determine bulk-temperature, thermal expansion of a rod, placed in the mixer, is measured by acoustic sensors, in the pulse-echo and the through-transmission modes. The results show that in the range of 60 °F–210 °F a temperature change of 1 °F can be measured. The second case relates to the nondestructive measurement of the solid content of the packaged food products. Accurate value of the solid-weight-ratio is required for correct thermal processing. Nondestructive methods with acoustic sensors in the through-transmission mode are utilized to measure solid-weight-ratio of a binary mixture of water and a food product. Using Wood's equation for porous media, a precision of less than 5% in the value of solid-weight-ratio is obtained by this new technique.

3:00

**2pEA6. Profiling a surface using its echoes.** Gareth Block and John G. Harris (216 Talbot Lab., 104 S. Wright St., Urbana, IL 61801)

The electromechanical reciprocity relation is used to construct a theoretical model of the imaging of a sinusoidal fluid-solid interface using a cylindrically focused acoustic beam. The electromechanical reciprocity relation is used to connect the change in the voltage measured at the electrical terminal of the transducer to the perturbation in the mechanical wave field caused by a change in the profile of the interface. It does so by mixing an unperturbed reference wave field with one containing the perturbed wave field. We use a regular boundary perturbation expansion to obtain an approximate boundary condition that fits directly into the reciprocity relation. This perturbation assumes both the amplitude and the slope of the profile are small. We limit our discussion to imaging a sinusoidal interface with the understanding that resolution of this profile is central to understanding that of more general profiles. We find that depending on the choice of parameter values the measured profile is not simply a replication of the original sinusoidal profile. Time permitting, we briefly indicate preliminary work on a related model of the imaging of the mechanical properties of a thin solid film using a confocal arrangement of point-focused transducers. [Work supported by NSF.]

3:15

**2pEA7. Detection of cracks in plates using guided waves.** Christoph Eisenhardt, Laurence Jacobs (School of Civil and Environ. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0355), and Jianmin Qu (Georgia Inst. of Technol., Atlanta, GA 30332-0405)

This research combines laser ultrasonic techniques with the two-dimensional Fourier transform (2-D-FFT) to investigate the effect of cracks on the dispersion of Lamb waves propagating in thin aluminum plates. The high fidelity and broad bandwidth of these optical techniques are critical elements to the success of this work. The experimental procedure consists of measuring a series of equally spaced, transient waves in aluminum plates containing notches; a crack is simulated with a 1-mm-thick saw-cut notch. The frequency spectrum (dispersion curves) for each plate is obtained by operating on these transient waveforms with the 2-D-FFT; this procedure extracts steady-state behavior from a series of transient waveforms. This study quantifies the effect of notch depth (two notch depths are examined: one-fourth and one-half of plate thickness) on the dispersion curves of three different plate thicknesses (nominal thicknesses of 1, 1.5, and 3 mm). These dispersion curves show that a notch reduces the transmitted energy by an amount that is directly proportional to a notch's depth. In addition, scattering by a notch causes definitive reductions in energy (evident in all modes) at certain frequency-wave-number combinations, thus providing experimental evidence of the relationship between crack size and the scattered Lamb wave field.

2p TUE. PM

**2pEA8. Detection of fatigue crack initiation and growth in steel specimens from Rayleigh scattering of 5-MHz Rayleigh waves.** Daniel A. Cook and Yves H. Berthelot (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Detecting the formation of surface microcracks remains an important challenge for condition-based maintenance (CBM) and remaining life prediction algorithms. A relatively new method [see M. Resch and D. Nelson, ASTM STP **1149**, 169–196 (1992)] consists of monitoring the reflection of pulsed surface waves off the crack. The method is extended to monitor the Rayleigh forward and backscattering (low-ka regime) with pulsed 5-MHz surface waves propagating on steel specimens under tensile cyclic loading. The cross-correlation of the received signals measured as a function of the number of fatigue cycles can be used to detect the presence and the growth of small cracks. Experimental results will be presented and discussed. [Work supported by the Office of Naval Research, N00014-95-1-0539, MURI Center for Integrated Diagnostics.]

**2pEA9. Vibroacoustical nondestructive evaluation of fatigue crack size.** Nadejda I. Bouraou (Dept. of Orientation and Navigation Systems, Natl. Technol. Univ. of Ukraine, 37, Peremogy Pr., Kiev, 252056 Ukraine, nadye@burau.inec.kiev.ua) and Alexander N. Tyapchenko (Natl. Technol. Univ. of Ukraine, Kiev, 252056 Ukraine)

For the vibroacoustical nondestructive evaluation of the fatigue crack's relative size the feature set, components of which are ratios of the testing object free or forced oscillations spectral amplitudes of higher and basic harmonics, is used. The features being used are functions of the relative crack size and do not depend on initial excitation intensity. The evaluation of the above-mentioned fault is produced by using the generalized likelihood ratio method. The maximum likelihood equations for using the feature set are solved. New expressions of the relative crack size estimation are obtained. The estimation of each feature is represented as the estimation set component. On the basis of the estimation set the statistics (mean, standard deviation) are computed.

TUESDAY AFTERNOON, 2 NOVEMBER 1999

GRANT–HARDING ROOM, 12:55 TO 3:00 P.M.

### Session 2pED

## Education in Acoustics: Integrating Acoustics into the Curriculum of Other Disciplines

Daniel R. Raichel, Chair

*Department of Mechanical Engineering, City College of New York, 140 Street and Convent Avenue, New York, New York 10031*

Chair's Introduction—12:55

### Invited Papers

1:00

**2pED1. One approach to architectural acoustics in education.** J. Christopher Jaffe<sup>a)</sup> (Jaffe Holden Scarborough Acoustics, Inc., 114A Washington St., Norwalk, CT 06854)

In the fall of 1997, Dean Alan Balfour of the School of Architecture at Rensselaer Polytechnic Institute asked me to introduce an undergraduate course entitled "Sonics in Architecture." I had previously taught a number of two-credit survey courses at Julliard, RPI, and City College of New York. My limited exposure as an adjunct professor did not prepare me for the difficulty of developing and integrating a 14-credit certificate course in what was in reality a vertical studio (a studio with students from different classes and disciplines). This paper discusses the curriculum I developed, the strengths and weaknesses of my initial plan, examples of student work and my revised curriculum for next year. In addition I will share my concerns regarding the teaching methods currently prevalent in many schools of architecture today, and how building science professionals might assist in addressing these issues. <sup>a)</sup>Distinguished Visiting Professor at the Rensselaer Polytechnic Institute.

1:30

**2pED2. Integrating acoustics into mechanical engineering education.** M. G. Prasad (Dept. of Mech. Eng., Stevens Inst. of Technol., Hoboken, NJ 07030)

Acoustics is generally considered as an area of specialization. However, the significance of acoustics can be seen in several disciplines including science, engineering, and arts, etc. Thus acoustics as knowledge that deals with production, propagation, transmission, and reception of sound plays an important role in many different fields. Acoustics can be integrated into mechanical engineering education through core as well as elective courses. As core material, acoustics nicely fits into a course in dynamical systems. The topics such as room acoustics and Helmholtz resonator are effective examples of first- and second-order systems, respectively. Acoustics along with vibrations can be part of core course material in design and machine dynamics. Acoustics with mechanical engineering applications such as noise control using source-path-receiver concepts, active control, noise control materials, product design, acoustical measurements, etc. can be included as elective courses. It is important to integrate acoustics in engineering curriculum because it gives the students an experience of the system effects and their impact on design. The integration of acoustics through both core and elective courses into mechanical engineering will be presented.



2:00

**2pED3. A course in musical acoustics for nonengineering majors at the University of Hartford.** Robert Celmer (Acoust. Prog. and Lab., College of Eng., Univ. of Hartford, 200 Bloomfield Ave., W. Hartford, CT 06117, celmer@mail.hartford.edu)

The University of Hartford's College of Engineering provides a service course entitled "Introduction to Musical and Architectural Acoustics" for nonengineering students majoring in such degree programs as music (at the Hartt School), technology fields, or the liberal arts. The material covers: acoustics terminology and concepts; a study of the science of each musical instrument family; sound radiation patterns of instruments and the implications for sound recording technology; the auditory system and hearing conservation; and an introduction to architectural acoustics related to such applications as band room and auditorium design. This presentation will describe some of the multi-media techniques for in-class presentation of the material as a means of better disseminating complex information using a sensory-rich environment. The extent to which this course fits into the assessment goals and general education science requirements of these nonengineering majors will also be discussed.

*Contributed Papers*

2:30

**2pED4. Development of an acoustics curriculum for architectural engineers.** Ralph T. Muehleisen (Dept. of Civil, Environ., and Architectural Eng., Univ. of Colorado, Boulder, CO 80309)

The University of Colorado is currently developing a series of acoustic classes for undergraduate Architectural Engineers and graduate students in the Building Systems Program of the Department of Civil, Environmental, and Architectural Engineering. The two classes currently undergoing development emphasize noise control in buildings and the design of acoustic venues. The goal of the classes is to teach the fundamentals of acoustics as needed by Architectural Engineers in order to design quieter buildings and

interact with acoustic consultants. In the classes, the students are given both theoretical and practical problems which involve both analysis and design. In addition, students get hands on experience by taking measurements of real rooms using standard equipment.

2:45

**2pED5. Should every physics student study acoustics?** Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Should every physics student learn about sound? Many of my colleagues say "no;" I say "yes." My arguments for restoring acoustics to the physics curriculum will be presented.

TUESDAY AFTERNOON, 2 NOVEMBER 1999

MCKINLEY ROOM, 1:00 TO 3:45 P.M.

**Session 2pMU**

**Musical Acoustics: African Musical Instruments and Traditions**

David Avorgbedor, Cochair

*School of Music, The Ohio State University, Columbus, Ohio 43210*

James M. Pyne, Cochair

*School of Music, The Ohio State University, Columbus, Ohio 43210*

*Invited Papers*

1:00

**2pMU1. Voiced noise: The "heterogeneous sound ideal" as preferred acoustic environment in selective sub-Saharan African instruments and ensembles.** Daniel Avorgbedor and James Pyne (School of Music, Ohio State Univ., 110 Weigel Hall, Columbus, OH 43210-1170)

This paper draws on the African-American composer Olly Wilson's notion of a "heterogeneous sound ideal" in which the author discusses the common preference for certain sound combinations and timbres in African and African-American musical traditions. The purpose of this presentation is to demonstrate, in detail, specific aspects of this concept by drawing on a variety of instrumental and ensemble traditions from sub-Saharan Africa with focus on the manipulation and intentional modulation of timbres that characterize much of the musical traditions of sub-Saharan Africa and which are often overlooked by analysts. The paper argues that this common emphasis on "unusual" timbres constitutes one of the major parameters distinguishing sub-Saharan African musical traditions from those of European art music traditions. A "heterogeneous sound ideal" suggests new approaches to the identification, measurement, evaluation, and appreciation of acoustic phenomena in the contexts of sub-Saharan African musical traditions. Additional demonstrations involving spectral-timbral projections of individual and ensemble instruments will be presented.

1:30-1:45

**Performance Demonstration**

2p TUE. PM

1:45

**2pMU2. “Greenotation:” A system for representing African drum sounds and techniques.** Doris Green (Pan African Performing Arts Preservation Assoc., Inc., 700 Southern Pkwy., Uniondale, NY 11553)

Within the past few decades technology has made it possible not only to write African music on paper like western notation, but also to retrieve and perform it from the printed page. In this paper I will present an overview of and demonstrate a system for writing African music I designed and tested over the past 20 years. Greenotation (after my name, Doris Green) was created for percussion instruments of Africa because Western musical notation could not notate the nuances and actions found in percussion music. Greenotation can notate the music of bells, rattles, drums, talking drums, sticks, stamping tubes, xylophones, hand clapping, and water drums; it is also able to represent dance movements that are integral to many sub-Saharan African drum and dance traditions.

2:15

**2pMU3. African influences on jazz: A lecture—demonstration.** Ted W. McDaniel (Div. of Jazz, School of Music, 110 Weigel Hall, Ohio State Univ., Columbus, OH 43210-1170)

This presentation focuses on selective aspects of sound patterns and performance traditions from sub-Saharan African societies that are retained in various ways in African–American jazz and performance styles. The talk will examine specific timbres and their related practices of repetition, collective improvisation, the acoustic phenomenon of pitch-bending, and texture. Recorded examples as well as live performances will highlight these elements and establish their relationships to the sub-Saharan African examples. A group of jazz musicians drawn from the OSU Jazz Division will perform specific items that further illustrate the concepts and practices outlined.

2:45–3:00

**Performance Demonstration**

3:00

**2pMU4. The influence of African musical traditions on gospel music.** Jan McCrary and Raymond Wise (School of Music, Ohio State Univ., 110 Weigel Hall, Columbus, OH 43210-1170, mcclary.5@osu.edu)

The paper examines historical and cultural influences of African music traditions on African–American gospel and spiritual music. Gospel music is one of several African–American genres of sacred music developed during the 1920’s and 30’s when musicians combined elements of blues and jazz music with church hymns and spirituals. The paper examines scholarly investigations of the impact of African music traditions on the development of gospel and spiritual music in America. In addition, the paper will explore unique African and American cultural influences on specific vocal-style characteristics. The session will feature The Ohio State University African American Music Chorale, which will perform representative short examples of African–American gospel and spirituals. The chorale also will present a short concert of complete works immediately following the session.

3:30–3:45

**Performance Demonstration**

TUESDAY AFTERNOON, 2 NOVEMBER 1999

GARFIELD ROOM, 1:30 TO 5:00 P.M.

**Session 2pNS**

**Noise: Active Noise Control, Sound Quality and Noise Control Analysis**

Glenn E. Warnaka, Chair

*Future Technologies, LLC, 1612 South Allen Street, State College, Pennsylvania 16801*

**Contributed Papers**

1:30

**2pNS1. An inverse structure for active noise control or combined active noise control/transaural sound reproduction.** Stephan Quednau and Martin Bouchard (School of Information Technol. and Eng., Univ. of Ottawa, 161 Louis Pasteur, Ottawa, ON K1N 6N5, Canada)

In this presentation, the use of inverse models of the acoustic plants is investigated for active noise control (ANC) systems. It is shown that the use of the inverse models combined with a predictor can: (1) greatly improve the convergence speed of stochastic gradient descent algorithms such as the multichannel filtered-X LMS or the modified filtered-X LMS and (2) significantly reduce the computational load of the resulting ANC

system. Moreover, for systems combining active noise control and transaural sound reproduction (TSR), the reduction of the computational load will be even greater, since TSR systems implicitly need the use of the inverse models of the acoustic plants, and the proposed ANC inverse structure already performs this inverse filtering. For broadband feedforward ANC controllers, the drawback of using the typically noncausal inverse models of the acoustic plants is to increase the delay required between the reference signal(s) of the controller and the disturbance signal(s) to be reduced. In the case of feedback controllers, this additional delay would likely limit the application of the proposed structure to the control of periodic disturbances only.

1:45

**2pNS2. Multichannel RLS algorithms and FTF algorithms for active noise control and sound reproduction systems.** Martin Bouchard (School of Information Technol. and Eng., Univ. of Ottawa, 161 Louis Pasteur, Ottawa, ON K1N 6N5, Canada)

In the fields of active noise control (ANC) and transaural sound reproduction (TSR), multichannel FIR adaptive filters are extensively used. For the learning of such FIR adaptive filters, recursive-least-squares (RLS) algorithms are known to typically produce a faster convergence speed than stochastic gradient descent techniques, such as the basic least-mean-squares (LMS) algorithm or even the fast convergence Newton-LMS, gradient-adaptive-lattice (GAL) LMS and discrete-cosine-transform (DCT) LMS algorithms. In this presentation, multichannel RLS algorithms and multichannel fast-transversal-filter (FTF) algorithms are introduced, with the structures of some stochastic gradient descent algorithms used in ANC: the filtered-x LMS, the adjoint-LMS and the modified filtered-x LMS. The new algorithms can be used in ANC systems or for the deconvolution of sounds in TSR systems. Also, heuristic techniques are introduced, to compensate for the potential ill-conditioning of the correlation matrix in ANC or TSR systems, and for the potential numerical instability of the multichannel FTF-based algorithms. Simulations of ANC and TSR systems will compare the performance of the different multichannel LMS, RLS, and FTF-based algorithms.

2:00

**2pNS3. Reduction of time-domain aliasing in adaptive overlap-add algorithms.** Uwe Rass and Gerhard H. Steeger (Georg-Simon-Ohm Univ. for Appl. Sci., FB NF, P.O. Box 210320, D-90121 Nuremberg, Germany, Gerhard.Steeger@fh-nuernberg.de)

The adaptive overlap-add algorithm (OLA) is an attractive means for acoustical signal processing (e.g., filtering, compression), since it is computationally effective and provides magnitude and phase information. It performs well when the input data blocks, the DFT, and the filter impulse response are of suitably chosen lengths. If the filter frequency response is altered on-line by an adaptation rule formulated in the frequency-domain, the resulting impulse response will, in general, be too long. This results in serious distortions, due to time-domain aliasing during the inverse DFT. Limiting the filter length by a filter design procedure needs much computation time, which is inhibitive for many acoustical applications. Time-domain aliasing distortions are particularly disturbing when musical signals are processed, due to their nonharmonic nature. An algorithm is proposed which convolves the filter frequency response with a short window sequence, solely in the frequency domain. Thereby, the aliasing components can be reduced by a predictable amount. The computational burden is low (approximately 12% add-on to the basic OLA algorithm) and, in addition, can be traded with the attenuation of the aliasing components. Discrete prolate spheroidal sequences have proven to be the optimal window type for this purpose.

2:15

**2pNS4. A method for evaluating automotive steering column noise.** Allan K. Kennedy (Delphi Automotive Systems, 3900 Holland Rd., Saginaw, MI 48601) and William E. Niehoff (Automotive Eng. Management Services, Inc.)

Automotive component suppliers are often called on to supply components which meet very vague requirements such as no audible noise from component. As odd as this might seem, this type of specification usually has a number associated with it. This number could be the measured A-frequency weighted sound pressure level, the total loudness, some averaged band sound pressure level, or even a vibration response level. The irony comes from the fact that most of these components do not generate noises on their own but only when subjected to intense vibration. To deal with this, component suppliers must somehow construct a measurement system which will capture these squeaks and rattles as the component is

vibrated. This paper describes a measurement system and method for evaluating the noises generated by automobile steering columns. Descriptions are given of the various vibration inputs used, the equipment required, and the parameters measured.

2:30

**2pNS5. Modifications of a handheld vacuum cleaner for noise control.** Gerald C. Lauchle and Timothy A. Brungart (Appl. Res. Lab. and Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804, GCL1@psu.edu)

Sound power measurements are described for a commercially manufactured, handheld vacuum cleaner. An objective assessment of the measured results is performed to identify the aeroacoustic sources of noise, suggested routes to noise control are then implemented, and a re-evaluation of the modified units is conducted and compared to the baseline results. The blade rate tone of the vacuum working fan is found to be the most annoying source, and it is reduced by up to 8 dB through a modification of the shroud that surrounds it. Unevenly spaced fan blades are also implemented as a noise control measure. This results in a comparable 8-dB reduction of the blade rate tone, but it introduces several new, side band tones to the sound power spectral data; the overall sound power increases slightly. Regardless, a subjective jury survey indicated that the modified unit having the unequally spaced fan blades is preferred 3:1 to the unit with the modified shroud only. [Work supported by Royal Appli-ance Manufacturing Co.]

2:45

**2pNS6. Measurement of acoustic properties by the two-cavity method.** Minor Nice (Owens Corning Testing Systems, 2790 Columbus Rd., Rte. 16, Granville, OH 43023-1200)

In this paper, results of the measurement of the infinite thickness properties (characteristic impedance, propagation constant) of acoustic materials by the two-cavity method will be presented. Results for some standard fiberglass materials will be shown along with comparison to some well-known correlations (Delaney and Bazley, Beranek). Results for materials other than fiberglass will also be presented. The impact of various measurement parameters (cavity spacing, averaging, impedance tube size, etc.) will be discussed relative to the measurement technique.

3:00–3:15 Break

3:15

**2pNS7. Vibration analysis and design of wiper motors.** Manuel Recuero, Juan Sancho, and Antonio Minguez (Departamento de I+D de Acustica, INSIA - UPM, Carretera de Valencia, Km. 7, 28031, Madrid, Spain, mrecuero@insia.upm.es)

This work has two objectives: (1) to characterize the manner of vibration of each part of a wiper motor; and (2) to find the sources of vibration so the noise emitted by the wiper motor will be known. Different measurement systems to analyze the vibrations of the wiper motor have been used. Results have been compared and the better system has been found. The results of the measurements have also been used to make some design recommendations for a new more silent wiper motor. This work started with four wiper motors. Two of them were characterized by the manufacturer of the motors as noisy and the other two as acoustically OK.

3:30

**2pNS8. Holographic sound-field imaging as a diagnostic tool.** Karl B. Washburn and Richard D. Godfrey (Owens Corning Testing Systems, 2790 Columbus Rd., Rte. 16, Granville, OH 43023-1200, karl.washburn@owenscorning.com)

Approaching its third decade, sound-field imaging using near-field acoustical holography has matured as an analysis technique. It has been established as the tool of choice for source motion mapping in structural acoustics and for source identification in noise control. The strength of near-field holography lies in its complete description of the sound field,

from source to far field. This opens a door to using sound-field imaging as an acoustical diagnostic tool. Several current trends in applying holography, including moving sources, reverberant spaces, and lower-cost systems, are surveyed. New diagnostic applications of sound-field reconstruction in audio-source imaging, transmission path analysis, and boundary material characterization are described.

3:45

**2pNS9. Disk brake squeal analysis using the ABLE algorithm.** G. Lou, T. W. Wu (Dept. of Mech. Eng., Univ. of Kentucky, Lexington, KY 40506), and Z. Bai (Univ. of Kentucky, Lexington, KY 40506)

Disk brake squeal noise is due to unstable friction-induced vibration. In order to study the stability of the disk system, the finite element method (FEM) is used to model a typical assembly that includes one rotor and two pads. Coulomb's friction law is applied at the contact interface between each pad and the rotor. Due to continuity of displacement in the normal direction, some degrees of freedom are condensed out. The FEM matrices of the dynamic system then become unsymmetric, which will yield complex eigenvalues. Any complex eigenvalue that has a positive real part will indicate an unstable mode. In real-world applications, the FEM model could include thousands of degrees of freedom. Fortunately, the resulting FEM matrices are sparse and a dynamic allocation scheme can be used to store only the nonzeros. Then a recently developed iteration method called ABLE (Adaptive Block Lanczos Method) is used to search only the complex eigenvalues with a positive real part in a certain user-specified frequency range. This algorithm has been shown to be very efficient and effective in predicting the unstable vibration modes.

4:00

**2pNS10. New geometric sound absorbers.** Glenn E. Warnaka (Future Technologies, L.L.C., 1612 South Allen St., State College, PA 16801)

This paper presents new concepts for simplified geometric sound absorbers that have a number of advantages over previous geometric sound absorbers and conventional sound absorbing materials. The new absorbers can be made of nearly any substance. Hence, the material for the absorbers can be selected on the bases of material costs, construction costs, and environmental considerations. They may be painted and cleaned without affecting their sound absorbing qualities. They also can be made very rugged for use in highway noise barriers and other outdoor applications. With the new geometric sound absorbers, it is possible to design absorbers that fulfill nearly any frequency range required. An infinite variety of designs is possible. In addition, the amount of sound absorption can be changed even after the sound absorbers have been installed. This may be of interest in multi-purpose rooms where the sound absorption can be altered for different functional uses of the room. The geometric sound absorbers may also be formed into flat treatments. Since the frequency range is a matter of design, high acoustic absorption can be maintained to frequencies well below 100 Hz, depending on design requirements, in treatments that are 6–16 mm thick.

4:15

**2pNS11. Analysis of impact noise of nail ejection in wire nail machines.** Jack Ding, Ahmed Al-Jumaily (Diagnostics and Control Res. Ctr., Auckland Inst. of Technol., Auckland, NZ), and Doug Wilson (Unitech Inst. of Technol., Auckland, NZ)

A wire nail production machine can be classified as one of the punch press machines. By nature, the production of wire nails involves the development of a series of force impulses, which in turn generate various sound pressure levels with different frequencies. Typically, the machine

operations generating impact noise include pressing, cutting, ejecting, wire gripping and feeding. The punch pressing is the movement of punch header to press forming the nail head. It produces a large force impulse, which in turn generates a high level of impact noise and consequently high levels of ancillary impact noise such as the impact noise by the backlashes. It has been found that this type of impact noise is not necessarily the most annoying one. The impact noise generated by nail ejection, however, has been identified as one of the annoying noise components. The ejector used in a nail machine for executing the movement of nail ejection produces ringing noise at high frequencies when it is struck by the punch header in each cycle of the machine operations. This paper presents the analyses of and experiments on the ejection force impulse and the characteristics of the impact noise generated by nail ejection.

4:30

**2pNS12. Acoustical analysis of infant/toddler rooms in daycare centers.** Tom Frank (Dept. of Commun. Disord., Penn State Univ., 5-A Moore Bldg., University Park, PA 16802) and Matthew V. Golden (Penn State Univ., University Park, PA 16802)

The purpose of this ongoing research is to determine ambient noise levels, signal-to-noise ratios (SNRs), and RTs in daycare centers. To date, 248 measures of ambient noise levels have been obtained in occupied infant/toddler rooms of 11 daycare centers and 109 measures in 7 of the 11 daycares unoccupied. The mean occupied levels were 57 dBA and 66 dBC and the mean unoccupied levels were 38 dBA and 56 dBC. A 1/3 OB analysis revealed that the occupied levels were relatively constant (46–50 dB) from 63 to 2000 Hz while the unoccupied levels decreased as frequency increased especially after 125 Hz. Using several different speech spectrums compared with the occupied 1/3 OB levels, estimated SNRs ranged from -3 to 12 dB. Dosimeter measurements ( $N=932$ ) revealed that the occupied daycares had an average peak sound level of 103 dB, high-threshold level of 62.5 dB, time-weighted average of 55.9 dB, and noise dose of 1.3%. Additional noise measurements and RTs are currently being collected. Overall, the unoccupied dBA levels and the estimated SNRs were less than recommended in the ASHA 1995 guidelines for classroom acoustics. [Work supported by PHS/NIH (1-R01-HD31540-01A2) Otitis Media, Behavior and Attention in Daycare.]

4:45

**2pNS13. Prediction of cutoff noise in centrifugal fans.** Zhichi Zhu, Song Li, and Dongtao Huang (Dept. of Eng. Mech., Tsinghua Univ., Beijing 100084, PROC)

The most important noise in centrifugal fans is aerodynamic noise, in which cutoff noise is the most main component. Cutoff noise is mainly caused by the fluctuating pressure on the cutoff in the volute impacted by the nonuniform flow from the exit of the rotating impeller. A numerical method is presented for prediction of cutoff noise in centrifugal fans. The contents of the method are mainly the following. (1) The fluctuating pressure serving as the source of cutoff noise is given out by using the existing computation software of 3D viscous flow in centrifugal fans and the time-frozen hypothesis. (2) The sound field in the volute of a centrifugal fan and the sound power levels of cutoff noise are computed by using the forth-order MacCormack scheme and making some important numerical measures. The sound field in the volute and the sound power levels of cutoff noise for three practical volutes with different cutoff clearance have been computed and the acoustic pressure distribution in space and change with time are discussed. Finally, the sound power levels of cutoff noise have been measured. The errors between the predicted and the tested sound power levels of cutoff noise are less than 3 dB. [Work supported by NSF.]



## Session 2pPA

## Physical Acoustics: Theoretical Physical Acoustics

R. Glynn Holt, Chair

Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street,  
Boston, Massachusetts 02215

## Contributed Papers

1:30

**2pPA1. A priori assessment of the effect of small scales on sound radiation from a subsonic axisymmetric jet.** Wei Zhao, Steven H. Frankel, and Luc Mongeau (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907)

The effect of small scales on sound generated from a subsonic axisymmetric jet was investigated by filtering near-field data from a direct numerical simulation to determine the accuracy of the large eddy simulation technique for jet sound predictions. Lighthill's acoustic analogy was employed to predict the far-field sound. The direct numerical simulation results were in agreement with published results [Mitchell *et al.*, *J. Fluid Mech.* **383**, 113–142 (1999)]. A new approach to handle the large spatial extent of the Lighthill source term yielded predictions of the far field in good agreement with the simulation results at low frequencies even for shallow angles from the jet axis. It was found that the small scales have little effect on the low-frequency sound, which is dominant in this case. The levels computed from both the filtered and unfiltered computed sound pressure field were in good agreement with those predicted by direct computations. As expected, filtering removed the small scale fluctuations in the near field, thereby reducing the magnitude of the source term for the high-frequency sound. [Work supported by NIH DCO 3577-02, RO1 grant from NIDCD.]

1:45

**2pPA2. Hot jet noise computed using large eddy simulation, computation of compressible free jet turbulence.** David B. Schein (Northrop Grumman Corp. for UCLA/MAE Dept., 9HI 1/GK, 8900 E. Washington Blvd., Pico Rivera, CA 90660) and William Meecham (Univ. of California Los Angeles, Los Angeles, CA 90095)

A computational fluid dynamics model for free, heated jet flow and resultant far-field sound has been developed which uses large-eddy simulation (LES) and Lighthill's acoustic analogy. A deductive, subgrid scale model (based on a Taylor series expansion of the weighting function in the Favre expansion) is used for the turbulent simulation. The simulation has been tested using published experimental mean flow field and rms fluctuation data [W. R. Quinn and J. Militzer, *Phys. Fluids* **31** (1988)] for a turbulent, free, square jet (known to be the same as a round jet of the same area). The ultimate objective is to address large Reynolds number, high subsonic (compressible) flow with realistic geometries, more representative of aircraft engine exhausts than can be considered using direct numerical simulation (DNS). In the simulation, Gaussian random velocity fields are introduced at the jet exit to excite the turbulence. The far-field sound and directivity is computed using the time-derivative form of Lighthill's source-integral result (formulated in terms of quadruple sources from the simulated flow field), which is integrated in time and contains the fluctuations set up by the time-varying stress tensor. Simulation for a WR 19-4 turbofan engine exhaust ( $Re=2 \times 10^6$  based on exit velocity and diameter) was performed, and propagated jet noise results compared with experimental acoustics data.

2:00

**2pPA3. Calculation of second harmonic beam patterns for ultrasonic sources of arbitrary shape.** Brian Landsberger (Caterpillar, Inc., P.O. Box 1875, Peoria, IL 61656-1875)

Ultrasonic transducers with nonaxisymmetric radiating surfaces are common in both industrial and medical applications. This presentation describes a numerical technique for calculating the second harmonic field produced in the field radiated by a transducer of arbitrary shape. The angular spectrum of the source is calculated with a two-dimensional Fourier transform, a weighted integral over the source spectrum is evaluated, and the inverse transform of the result yields the second harmonic pressure in any plane parallel to the source plane. Arbitrary absorption and dispersion are taken into account. The method is similar to that used by Alais and Hennion [*Acustica* **43**, 1–11 (1979)] to evaluate the field of a parametric array. In the present theory, however, we do not make use of the parabolic approximation. Beam steering and focusing are introduced by appropriate phasing of the source function. Calculations of radiation patterns produced by rectangular and other sources are presented.

2:15

**2pPA4. Nonlinear modeling of focused acoustic fields.** B. Edward McDonald (Saclant Undersea Res. Ctr., 19138 La Spezia, Italy) and William A. Kuperman (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

A series of simulations was performed to address the following question: If a circular piston projector were capable of producing finite amplitude beam pulses of arbitrary amplitude and bandwidth, what parameters would lead to the greatest concentration of energy within the smallest focal volume? The simulations were performed using the NPE time domain nonlinear acoustics model [B. E. McDonald and W. A. Kuperman, *J. Acoust. Soc. Am.* **81**, 1497 (1988)] adapted to azimuthal symmetry about the beam axis [G.-P. J. Too and J. H. Ginsberg, *J. Acoust. Soc. Am.* **91**, 59 (1992)]. For a given amplitude of the focusing wave, simulations were performed over a range of values for the parameter of nonlinearity and the effective piston aperture at the focus. Results show a strong maximum in focal gain as a function of the parameter of nonlinearity for each aperture considered. The maximum can be interpreted as competition between increasing amplitude at the piston and nonlinear defocusing as the wave propagates. Results may be rescaled to imply an optimum amplitude for a given parameter of nonlinearity. [Work supported by Saclantcen.]

2:30

**2pPA5. Effects of small-amplitude fluctuations on thermoviscous shock wave structure.** David G. Crighton and Pablo L. Rendon (Dept. of Appl. Mathematics and Theoretical Phys., Univ. of Cambridge, Silver St., Cambridge CB2 9EW, UK)

We study the low-diffusivity limit of the plane Burgers equation when small-amplitude fluctuations are introduced behind the shock region. Experiments in hydraulic jump propagation suggest that these downstream fluctuations can produce large-scale effects in the shock region. A self-consistent model is proposed in which the fluctuations themselves largely determine propagation over the mean field, and are thus permitted to am-

ply sufficiently so as to broaden the nonuniform part of the mean flow. By means of asymptotic matching we find an expression for the power amplification of any given spectral component in a mean field where a particular gradient is sought in the shock region. Shock thickening and profile distortion are predicted for certain values of fluctuation amplitude and frequency, and these results are confirmed numerically using a pseudo-spectral method.

2:45

**2pPA6. Numerical solution of a statistical version of the Burgers equation.** Penelope Menounou (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712)

The Burgers equation (BE) is transformed into an unclosed set of linear equations that describe the evolution of the joint moments of the sound signal. The set of the equations is represented by a recursion equation termed Statistical Burgers Equation (SBE). Unlike for the BE, the time signal at the source is not required for the solution of the SBE and the power spectral density of the signal (or of a stochastic process) can be given instead [Menounou and Blackstock, *J. Acoust. Soc. Am.* **99**, 2539(A) (1996)]. The SBE is solved numerically by appropriately selecting only a finite number of equations from the infinite set, provided that the joint moments of the signal (or the stochastic process) are known at the source. The properties of the joint moments are presented for two source conditions: (i) a sinusoidal signal; and (ii) a Gaussian stationary and ergodic stochastic process. The finite difference scheme employed for the prediction of the joint moments' evolution is presented and stability criteria for the algorithm are derived. Finally, the sequence of computing the equations within the set is investigated and its effect on the stability of the numerical algorithm is demonstrated. [Work supported by the F. V. Hunt Postdoctoral Fellowship.]

3:00–3:15 Break

3:15

**2pPA7. Mode counts for rooms and waveguides.** Christopher L. Morfey, Matthew C. M. Wright, and Seong-Ho Yoon (Inst. of Sound and Vib. Res., Univ. of Southampton, Southampton SO16 1BJ, UK)

The density of modes in a room (3-D) or waveguide cross-section (2-D) is known to be approximated at high frequencies by expressions involving room volume and surface area, or (in 2-D) the area and perimeter of the waveguide cross-section. Deviations of the actual mode count from the smoothed approximation are studied numerically and analytically for simple shapes: rectangular (2-D and 3-D) and annular (2-D). Periodic clustering of eigenvalues, associated with cyclic rays [R. Balian and C. Bloch, *Ann. Phys.* **69**, 76–160 (1972)] is demonstrated at high frequencies. Expressions are given for estimating the standard deviation of 2-D and 3-D mode counts from their smoothed values, in a finite frequency band, when one dimension is much smaller than the others.

3:30

**2pPA8. Band-gap engineering in a two-dimensional periodic system of fluids.** M. S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla 72570, Mexico, manvir@sirio.ifuap.buap.mx) and B. Djafari-Rouhani (Univ. of Sci. & Technol., 59655 Villeneuve D'Ascq, Cedex, France)

This work emphasizes that periodic binary systems can give rise to genuine acoustic band gaps (or stop bands) within which sound and vibrations remain forbidden. Extensive band structures for two-dimensional (2-D) periodic arrays of air cylinders in water background are computed. Complete, multiple, huge stop bands are found for both square and hexagonal lattices. The lowest stop bands are largest for a range of filling fraction  $10\% \leq f \leq 55\%$ , with a gap/midgap ratio of 1.8. The most inter-

esting finding of the present investigation is that the low-frequency, flat passbands for a perfectly periodic system correspond to the discrete modes of a *single* airy cylinder. This almost exact correspondence is attributed to the low-filling fraction and the huge density contrast in air and water. It is stressed that such a simple inhomogeneous system as made up of air and water exhibits the largest stop bands ever reported for 2-D or 3-D elastic as well as dielectric (photonic crystals) composites. [Work partially supported by CONACyT Grant No. 28110E.]

3:45

**2pPA9. Dispersion relations of surface waves in semi-infinite periodic scattering arrays.** D. Caballero and J. Sánchez-Dehesa (Dept. Theoretical Condensed Matter., Autonomous Univ., 28049 Madrid, Spain)

Lord Rayleigh was the first in proposing the existence of surface waves in elastic systems at the end of the 19th century. Since then, people have observed elastic surface waves in very different contexts: from the huge wavelengthed seismic movements to the ultrasonic waves traveling along electronic devices. The aim of this work is to study the propagation of elastic surface waves through structures with periodically distributed inhomogeneities with cylinder symmetry. The theoretical study is based on variational tools developed by the authors. The periodicity led us to characterize the system by a band structure. In principle, those results can be used to design devices to forbid the transmission or to enhance and drive the propagation of elastic surface waves of selected frequencies. [Work supported by CICYT of Spain.]

4:00

**2pPA10. Experimental determination of acoustic bands in two-dimensional sonic band gap crystals.** J. Sánchez-Dehesa, D. Caballero (Dept. of Theoretical Condensed Matter, Autonomous Univ., 28049 Madrid, Spain, jsdehesa@uamca3.fmc.uam.es), R. Martínez-Sala, C. Rubio, J. V. Sánchez-Pérez, F. Meseguer, and J. Llinares (Unidad Asociada CSIC-UPV, 28049 Madrid, Spain)

The dispersion relation of acoustic bands in sonic band gap crystals built up with periodic arrays of rigid cylinders in air have been characterized both experimentally and theoretically. A technique based on the analysis of the phase-shift experienced by the sound is used to construct the experimental acoustic dispersion relation. Measurements have been performed in square, triangular, and honeycomb lattices for different filling fractions. The experimental setup also allows to detect the so-called "deaf bands" [J. V. Sánchez-Pérez *et al.*, *Phys. Rev. Lett.* **80**, 3080–3083 (1998)]. A variational method is employed to calculate the corresponding acoustic bands. The good agreement between theory and experiment supports the experimental technique here described as an important characterization tool for these systems. [Work supported by CyCIT of Spain and Generalitat of Valencia.]

4:15

**2pPA11. Band gap engineering in three dimensional system of air bubbles in water.** M. S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla 72570, Mexico, manvir@sirio.ifuap.buap.mx), B. Djafari-Rouhani, L. Dobrzynski (Dept. of Phys., Univ. of Science & Technology, Lille-I, 59655 Villeneuve D'Ascq, Cedex, France)

This work reports systematic and extensive evidence for the existence of complete, multiple, huge stop bands in the band structures for cubic arrays of air bubbles in water. All three important structures: face-centered cubic (fcc), body-centered cubic (bcc), and simple-cubic (sc) arrangements are investigated using the Fourier-series expansion (of position dependent density and elastic constant) method. It is noteworthy that this formulation does not require matching of the messy boundary conditions. The lowest stop bands are largest for a volume fraction  $f \leq 10\%$ , with a gap/ midgap ratio of 1.8, for all three geometries. It is found that the low-frequency, flat

passbands for the perfectly periodic systems correspond to the discrete modes of a *single* bubble. This is an artifact of the low filling fraction and huge density contrast in air and water. It is stressed that such a simple inhomogeneous system as made up of air bubbles in water gives rise to the

largest stop bands ever reported for elastic/acoustic as well as dielectric composites—save the similar 2D composites discussed in the preceding work. [This work was partially supported by CONACyT grant No. 28110E.]

4:30–4:40 Break

### Contributed Poster Papers

Papers 2pPA12 and 2pPA13 will be presented in poster format. Authors will be at their posters from 4:40 p.m. to 5:10 p.m.

**2pPA12. Stop bands of tri-dimensional sonic band gap crystals.** P. Ribaute, D. Caballero, and J. Sánchez-Dehesa (Dept. of Theoretical Condensed Matter, Autonomous Univ., 28049 Madrid, Spain)

Acoustic and ultrasonic waves are powerful probes for many macroscopic systems ranging from the oceanic environment to devices under mechanical stress. From the technological point of view, tools must be developed in order to understand the propagation of acoustic waves through inhomogeneous media. In this work, a variational method is developed to study the acoustic response of three-dimensional systems with periodically distributed scatterers. The dispersion relations of the corresponding acoustic bands are obtained. They determine the stop bands. Also, it is shown that symmetry plays an important role in the full problem of transmitting acoustic waves through a periodic structure. If the symmetry of the external wave does not match with the band at its corresponding frequency, the acoustic energy cannot be transmitted. The existence of

these bands has been previously reported in two-dimensional sonic crystals [J. V. Sánchez-Pérez *et al.*, Phys. Rev. Lett. **80**, 3080–3083 (1998)]. [Work supported by CICYT of Spain.]

**2pPA13. A novel time-domain method to study the propagation of acoustic waves across composite media.** D. Bosquetti, J. Sánchez-Dehesa, and D. Caballero (Dept. of Theoretical Condensed Matter, Autonomous Univ., 28049 Madrid, Spain, jsdehesa@uamca3.fmc.uam.es)

It has been previously shown that the symmetric-split-operator technique is a useful method to study evolution problems in classical mechanics. In this work the technique is applied to obtain a finite-difference scheme which allows us to analyze the propagation of acoustic waves in time domain. This scheme is developed for a general three-dimensional composite system. Numerical results are presented here for some one-dimensional structures. [Work supported by UAM and CICYT of Spain.]

2p TUE. PM

TUESDAY AFTERNOON, 2 NOVEMBER 1999

UNION E ROOM, 1:30 TO 4:35 P.M.

### Session 2pPP

## Psychological and Physiological Acoustics and Speech Communication: Honoring the Contributions of Robert C. Bilger: Bilger and Better Science

Lawrence L. Feth, Cochair

*Speech and Hearing Science, The Ohio State University, 110 Pressey Hall, 1070 Carmack Road, Columbus, Ohio 43210-1372*

Walt Jesteadt, Cochair

*Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131*

Chair's Introduction—1:30

### Invited Papers

1:35

**2pPP1. A Bilger journey.** Ira J. Hirsh (Central Inst. for the Deaf, 818 South Euclid, St. Louis, MO 63110)

Bob Bilger started (1954) his postdoctoral appointment at Central Institute with Ira Hirsh. The topic was recovery of auditory threshold for tones, following exposure to tones. They confirmed the results of Hirsh and Ward (1952) on recovery of click thresholds after similar exposures. The collaboration continued with papers on masking, remote masking, and additivity of different kinds of masking. The organization at Central Institute was such that he, like other research associates, learned about and participated in work in other laboratories. He participated in physiological studies; he branched out to time and speech; and his statistical expertise was much sought in clinical research. His move to Pittsburgh allowed him to continue his disregard for disciplinary barriers and his readiness for clinically related matters. Just before leaving Pittsburgh for Illinois, he undertook to write, organize and edit the impressive, monumental, first serious study of the effect of a single-channel cochlear implant on adult human subjects [Bilger *et al.* (1977)]. The Illinois stretch demonstrates that (1) speech and psychoacoustics are not very far apart, and (2) very successful Chairs of Speech and Hearing Departments need not be clinicians.

**2pPP2. Origins of the IWAIF model.** Lawrence Feth (Dept. of Speech & Hearing Sci., Ohio State Univ., Columbus, OH 43210, feth.l@osu.edu)

The intensity weighted average of instantaneous frequency (IWAIF) model evolved from early work on the processing of frequency-modulated (FM) tones suggested by Bob Bilger [Feth *et al.*, *J. Acoust. Soc. Am.* **45**, 1430–1437 (1969)]. This early work on FM signals led to the incorporation of Voelker's Unified Theory of Modulation [Proc. IEEE **54**, 340–353 (1966)] into the original envelope-weighted average of instantaneous frequency (EWAIF) model [L. L. Feth, *Percept. Psychophys.* **15**, 375–379 (1974)]. This talk will describe the development of the IWAIF model and describe its application in auditory signal processing such as formant tracking in speech processing algorithms and the perception of Doppler-frequency shifts produced by moving sound sources.

**2pPP3. Variability in forward masking and intensity discrimination.** Walt Jesteadt, Jason F. Reimer, and Huanping Dai (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, jesteadt@boystown.org)

Jesteadt and Bilger [W. Jesteadt and R. C. Bilger, *J. Acoust. Soc. Am.* **55**, 1266–1276 (1974)] noted greater individual differences for frequency discrimination than for intensity discrimination and summarized this effect in terms of analyses of variance. In this paper, we report similar comparisons for forward masking and intensity discrimination. Four subjects were tested in a 2IFC adaptive procedure. The forward masker and the pedestal in the intensity-discrimination task were broadband stimuli consisting of 18 components with third-octave spacing from 200 to 10 000 Hz. Thresholds were obtained for forward maskers from 20 to 80 dB SPL and for pedestals from 10 to 80 dB SPL in 10-dB steps, with eight 50-trial blocks per condition. When the results were analyzed in terms of signal power at threshold for forward masking and intensity discrimination, they showed smaller individual differences and less variability across repeated measurements in forward masking than in intensity discrimination. Psychometric functions reconstructed from the adaptive tracks were steeper for forward masking than for intensity discrimination. Thresholds obtained in forward-masking tasks are often considered to be highly variable, but may be more stable than those obtained in studies of intensity discrimination using comparable stimuli. [Work supported by NIDCD.]

**2pPP4. Physiological mechanisms of frequency discrimination.** Eric Javel (Dept. of Otolaryngol., Univ. of Minnesota, Minneapolis, MN 55455)

I recall a conversation with Bob Bilger that occurred around 1971, when I was a graduate student in Bioacoustics at Pitt. He said (translated into English from the original Bilgerian), "Did it ever occur to you that you can apply principles of signal detection theory in physiological studies? That way, you could examine neural and psychophysical performance using comparable measures and possibly shed light on underlying mechanisms." This idea, which predated its implementation in auditory research by at least 10 years, was initially lost on me. Much later, I recognized the possibilities and performed some pertinent experiments with various colleagues. We utilized 2AFC adaptive tracking procedures to investigate performance limits of single cat auditory nerve fibers and neural populations on pure-tone frequency discrimination tasks. For single-fiber responses, we found that: (1) decision strategies based on detecting differences in spike counts fail to account for perceptual findings by wide margins; and (2) strategies based on detecting differences in phase-locked activity produce data that match perceptual findings well at low frequencies but rapidly degrade at frequencies >1500 Hz. Examining neural population performance using a stochastic excitation pattern model, we found that a strategy which simply estimates the spatial location of the response peak accounts well for perceptual performance at most frequencies. See, Bob? Your idea works, and I WAS paying attention.

**2pPP5. Development of a screening version of the communication profile for the hearing impaired.** Marilyn E. Demorest (UMBC, 1000 Hilltop Circle, Baltimore, MD 21250), David J. Wark (Univ. of Memphis, Memphis, TN 38105), and Sue Ann Erdman (Baltimore, MD 21286)

The 163-item CPHI provides a diagnostic profile of scores on 25 scales that describe a client's adjustment to hearing impairment. Factor structure of the instrument shows that two important factors it assesses are communication performance (an aspect of hearing disability) and psychosocial adjustment to hearing impairment (an aspect of handicap). The goal of this study was to develop a brief instrument, to be used in conjunction with standard audiometric assessment, to screen for disability and handicap. A pseudo-random sample of 1000 cases was drawn from a large, heterogeneous clinical database. Item response theory was used to derive item characteristic curves, and item selection was based primarily on item discrimination. Nine items were chosen to screen for communication performance, and 11 were chosen to screen for psychosocial adjustment. Pass/Fail criteria were developed, and sensitivity and specificity were evaluated in a holdout sample of 319 cases.

**2pPP6. Forward masking recovery and peripheral compression in normal-hearing and cochlear-impaired ears.** David A. Nelson and Anna C. Schroder (Dept. of Otolaryngol., Univ. of Minnesota, 396 UMHC, 516 Delaware St. S.E., Minneapolis, MN 55455, dan@tc.umn.edu)

Iso-response temporal masking curves are obtained from normal-hearing subjects at a probe frequency of 1000 Hz for masker frequencies between 500 and 1200 Hz. Time constants calculated from the temporal masking curves varied with masker frequency, from around 70 ms for low off-frequency maskers (500–600 Hz) to around 36 ms for on-frequency maskers (close to the probe frequency). Continuing Bilger's earlier pursuits of nonlinearities in hearing, estimates of peripheral compression were calculated under the assumption that the response to a low off-frequency masker is linear at the probe frequency place. Average compression exponents



varied from close to 1.0 for remote off-frequency maskers (both below and above the probe) to below 0.4 for on-frequency maskers. Input-output transfer functions derived from the compression exponents were consistent with BM transfer functions recorded in animals with normal cochlear function. Comparisons of iso-response temporal masking curves in subjects with sizable cochlear hearing losses at the probe frequency yielded linear transfer functions consistent with BM data from cochlear-damaged animals. It is concluded that time constants for recovery from forward masking in ears with cochlear hearing loss are no different than those obtained from normal-hearing ears, once differences in peripheral compression are taken into account. [Work supported by NIH-NIDCD Grant DC00149 and the Lion's 5M International Hearing Foundation.]

3:35

**2pPP7. Selected aspects of across-frequency processing in binaural hearing.** Constantine Trahiotis, Leslie R. Bernstein (Surgical Res. Ctr., Dept. of Surgery (Otolaryngol.) and Ctr. for Neurological Sci., Univ. of Connecticut Health Ctr., Farmington, CT 06030), and Richard M. Stern (Carnegie Mellon Univ., Pittsburgh, PA 15213)

While considering the many contributions of Dr. R. C. Bilger to knowledge concerning auditory processing, it seemed both fitting and appropriate to honor him by discussing research in binaural hearing that is consistent with key aspects of what we perceive to be his style. Accordingly, the presentation will highlight the integration of theory, measurement, and empirical observation in selected aspects of across-frequency processing in binaural hearing. The topics addressed will be: (1) lateralization as a function of bandwidth, center frequency, and interaural time/phase disparities; (2) binaural interference; (3) the incorporation of peripheral compression, rectification, and low-pass filtering in an index that accounts for binaural detection across frequency.

3:55

**2pPP8. Context effects with limited spectral information.** Melanie L. Matthies (Boston Univ., 635 Commonwealth Ave., Boston, MA 02215)

Speech recognition has been established and unified as a construct through the systematic study of its domain [Bilger, 1984]. To investigate the effect of limiting spectral information, items from the Revised Speech Perception in Noise (SPIN) test [Bilger *et al.*, 1984] were impoverished to include two, three, or four channels of spectral information [cf. Shannon *et al.*, 1995] and administered to 62 normal-hearing, young adults. Subjects also identified consonants that likewise had been spectrally limited via signal processing. When consonant confusion matrices were analyzed, transmission of place of articulation was strongly affected by spectral limitations followed closely by effects on manner perception, while voicing and resonance (oral/nasal) cues were relatively robust. High-context (HC) SPIN items were significantly easier for the subjects to identify than low-context (LC) items for the four-channel sentences. The context advantage decreased for the two-channel task because listeners were unable to utilize the sentence cues. Across channel conditions, a regression line fit to normalized HC by LC data had a slope of 0.9. Reduced semantic and syntactic information in LC items, therefore, contributed to a perceptual uncertainty network (PUN) when spectral information was limited and this SPIN PUN should be explored further. [Work supported by MURI Grant Z883402.]

4:15

**2pPP9. Modeling closed-set phoneme and open-set word recognition by multi-channel cochlear implant users.** Ted A. Meyer, Mario A. Svirsky (Indiana Univ. School of Medicine, Dept. of Otolaryngol., Indianapolis, IN), Stefan Frisch (Univ. of Michigan, Ann Arbor, MI), Adam R. Kaiser, David B. Pisoni, and Richard T. Miyamoto (Indiana Univ. School of Medicine, Indianapolis, IN)

There has been phenomenal growth in research on speech perception by cochlear implant (CI) users since the printing of the "Bilger Report" [Bilger *et al.*, *Ann. Otol. Rhinol. Laryngol.* **86** (S38), 1-176 (1977)]. Undoubtedly, average speech perception performance with these devices has improved dramatically. However, despite advances in implant technology, CI users continue to demonstrate a wide range in the ability to perceive speech. Little progress has also been made in understanding how CI users actually perceive speech. Although many correlational analyses have been carried out, little research has focused on mechanisms of speech perception, and many clinical decisions are made on a trial-and-error basis. Using a new approach, we have developed a quantitative, psychophysically based model (Multidimensional Phoneme Identification, MPI) of phoneme perception by CI users [M. A. Svirsky and T. A. Meyer, *J. Acoust. Soc. Am.* **103**, 2977 (1998)]. The MPI model generates phoneme confusion matrices from performance on psychophysical tasks. In a complimentary line of work, we aim to predict open-set spoken word recognition from phoneme performance [S. Frisch and D. B. Pisoni, *Res. Spoken Lang. Proc.* 261-288 (1998)]. In both lines of research, we have found predictable relations between perception of phonetic features, phonemes, and words by CI users. [Work supported by NIH, AAO-HNS, DRF, NOHR.]

## Session 2pSA

## Structural Acoustics and Vibration: Characterization of Structural Properties

Jerry H. Ginsberg, Chair

School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

## Contributed Papers

1:15

**2pSA1. Measurements of the dynamic elastic moduli of viscoelastic materials with micro-inclusions.** R. Lance Willis, Lei Wu, and Yves H. Berthelot (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

An indirect method has been proposed [see *J. Acoust. Soc. Am.* **102**, 3549–3555 (1997)] to determine the complex, frequency-dependent, elastic moduli of polymers containing microscopic inclusions from laser-based measurements and finite element modeling. The method consists in measuring the surface dynamics of the sample under harmonic excitation by noncontact laser Doppler interferometry. The experimental results are then matched with numerical predictions in which the moduli are the adjustable parameters. The method is first validated at ambient pressure by measuring the dynamic moduli of a sample of known properties. Results obtained with samples of the same voided viscoelastic material but with different aspect ratios are presented. To assess the effect of static pressure on the moduli, the sample is placed inside a static pressure air chamber (0–500 psi) and the effect of pressure cycling is investigated by measuring the moduli during the first pressure cycle (0–500, and 500–0 psi) and after eight pressure cycles. [Work supported by ONR, code 334.]

1:30

**2pSA2. Determination of the complex shear modulus from torsional waves in viscoelastic bars.** Jacek Jarzynski (School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, jacek.jarzynski@me.gatech.edu) and John W. Doane (Georgia Inst. of Technol., Atlanta, GA 30332)

The torsional wave method for measurements of shear modulus, described by Garrett [*J. Acoust. Soc. Am.* **88**, 210–221 (1990)], is modified by the addition of metal end pieces to the sample. This allows measurements of the shear modulus at low frequencies (100–1000 Hz) for small samples. Measurements can be made with the sample either air-loaded or water-loaded. The system is calibrated using a laser vibrometer. Data will be presented for neoprene rubber and compared with measurements of Young's modulus using a resonant rod technique.

1:45

**2pSA3. Identification of multi-degree-of-freedom nonlinear vibratory systems consisting of unknown elastic forces.** Christopher M. Richards (Caterpillar, Inc., 100 N. E. Adams St., Peoria, IL 61629-9760, richards@iaonline.com) and Rajendra Singh (The Ohio State Univ., Columbus, OH 43210)

Nonlinear system identification techniques often require *a priori* knowledge of the nature and mathematical form of the nonlinearities. Unfortunately, for practical systems, this is not always possible. As a result, nonlinearities are often approximated and questions remain as to whether an accurate model can be determined. In addition, under experimental conditions, the amount of measurement noise present in the identification process must also be quantified. To address these issues, identification of discrete systems consisting of nonlinear elastic forces is examined in the presence of uncorrelated noise. It is assumed that the mathematical form of the nonlinearities is unknown but can be approximated by polynomials. Coherence functions are introduced which are based on a "reverse path"

spectral approach developed by the authors for multi-degree-of-freedom systems. These coherence functions, as calculated from conditioned spectra, indicate the extent of uncorrelated noise present and the accuracy of the assumed mathematical models. Using several example simulation systems, including a system with a continuous nonlinearity described by a noninteger exponent, both temporal and spectral identification techniques are employed to study the issues described above.

2:00

**2pSA4. Experimental characterization of nonlinear rubber isolators in a multi-degree-of-freedom system configuration.** Christopher M. Richards (Caterpillar, Inc., 100 N. E. Adams St., Peoria, IL 61629-9760, richards@iaonline.com) and Rajendra Singh (The Ohio State Univ., Columbus, OH 43210)

Experimental characterization has been investigated for three rubber isolators placed in a multi-degree-of-freedom system configuration. The configuration consisted of a rigid mass mounted to a flexible support beam via each of the rubber isolators. Random, sinusoidal, and sine sweep excitations were applied to the rigid mass and accelerations were measured. Data revealed the dependence of the isolators' properties on amplitude and type of excitation. Modeling based on the continuous system theory resulted in quasi-linear representations of the experimental system. Discrete system models consisting of nonlinear polynomial equations for describing the nonlinearities of the isolators were also developed. In addition, sound pressure levels were measured during sine sweep excitation for one of the isolators. Increased sound levels were observed at off-resonant regions of the vibration response. Finally, single-degree-of-freedom static and dynamic experiments were conducted. Discrepancies in the isolators' properties were found between the single- and multi-degree-of-freedom experimental results.

2:15

**2pSA5. Simple nondestructive quantization of specially orthotropic materials.** Curt Preissner and Thomas J. Royston (Univ. of Illinois at Chicago, MC 251, 842 W. Taylor St., Chicago, IL 60607, troyston@uic.edu)

Many different plate structures can be classified as specially orthotropic, such as cross-ply composites, unidirectional composites and quarter-cut wood boards. A simple method of determining the four elastic constants of a completely free, specially orthotropic plate has been developed. An improved Rayleigh expression, utilizing beam functions, is used to represent the deflection shape of the plate. A dimensionless objective function is formed from the frequency equation. The material properties are determined by minimizing the objective function with respect to modal analysis and geometry data. The technique has been successfully used on aluminum, composite and wooden plates. The ease of the method lends itself to use in NDE of composite plates and qualification of wood used in the construction of musical instruments.

2:30

**2pSA6. Maximum-likelihood estimation of wave components for structural vibrations.** Peter J. Halliday and Karl Grosh (Dept. of Mech. Eng. and Appl. Mech., Univ. of Michigan, Ann Arbor, MI 48109-2125)

In the analysis of structural acoustic systems, it is desirable to have a robust method for estimating the wave components (complex wave numbers and the corresponding wave amplitudes) of the structural response. A technique for evaluating steady-state wave propagation on single and multiply connected beam-like structures from noisy data is developed using previous modifications of Prony's method [K. Grosh and E. G. Williams, *J. Acoust. Soc. Am.* **93**, 836–848 (1993)] and a new maximum-likelihood scheme. An overdetermined exponential solution is fit to evenly spaced data points from which propagating and evanescent wave numbers and their associated amplitudes are determined for each frequency of interest. *In situ* estimation of the elastic modulus may be accomplished by using a least-squares fit of the analytic solution of the Timoshenko beam theory dispersion relationship to the dispersion curve estimated from experimental data. Structural intensity is evaluated from the wave component and material property information. Numerical experiments have shown the effectiveness of these methods even for relatively low signal-to-noise ratios, on the order of 30 dB.

2:45–3:00 Break

3:00

**2pSA7. Sensitivity of resonance-frequency shifts to defect location in adhesive-bonded joints.** Deborah Hopkins, Seiji Nakagawa, Kurt Nihei (MS 46A-1123, Eng. Div., Lawrence Berkeley Natl. Lab., 1 Cyclotron Rd., Berkeley, CA 94720, DLHopkins@lbl.gov), and Guillaume Neau (Univ. of Bordeaux1, Bordeaux, France)

Techniques based on frequency shifts and mode-shape analysis are being investigated to determine their feasibility for characterizing defects in adhesive-bonded joints in automotive structures. It is well known (Rayleigh–Ritz derivation) that introduction of a crack-like defect into a structure reduces its stiffness and results in a corresponding downward shift in resonance frequencies. Experimental and modeling results show that analysis of resonance-frequency shifts is much more complicated for bonded joints where the defect consists of a gap in the adhesive layer. Structures containing a defective joint sometimes exhibit higher-resonance frequencies than a structure with an undamaged joint. Such results have been observed in laboratory experiments on aluminum plates with adhesive-bonded T joints, and in finite-element simulations for a variety of structures. In all cases, the defects studied are gaps in the adhesive layer of the joints. While the reduction in mass associated with a gap in the adhesive layer is very small, the mass effect overwhelms the frequency-decreasing effect of the reduced stiffness when the defect is located in a low-stress region of the joint. Thus, the direction and magnitude of frequency shifts depend on the resonance mode and the location of the defect.

3:15

**2pSA8. Thermoelastic effects on surface acoustic wave propagation.** Zhongyu Yan and Peter B. Nagy (Dept. of Aerosp. Eng. and Eng. Mech., Univ. of Cincinnati, Cincinnati, OH 45221)

The effect of thermal stresses on ultrasonic surface wave propagation was investigated. Quasi-static and dynamic thermal stresses and deformations were considered in materials with and without fatigue cracks. The

perturbation of ultrasonic surface waves due to these thermal effects were studied both theoretically and experimentally. A long-duration infrared laser pulse was used to irradiate a series of aluminum and titanium specimens with fatigue cracks of known sizes between 0.5 and 1.0 mm. Preliminary numerical results from finite element simulation were found to be in good agreement with the experimentally observed laser-induced modulations of the ultrasonic surface wave. In particular, both the numerical and experimental results indicated the same characteristic transient behaviors of the thermo-optical modulations for cracked and intact materials. The different behaviors suggest a new promising method for ultrasonic nondestructive evaluation, which cannot only effectively distinguish fatigue cracks from other artifact scatters, but potentially can also provide a method for quantitative evaluation of crack features. [This effort was sponsored by the Defense Advanced Research Project Agency (DARPA) Multidisciplinary University Research Initiative (MURI), under Air Force Office of Scientific Research Grant No. F49620-96-1-0442.]

3:30

**2pSA9. A definition of a loss factor is unique, some definitions are not.** G. Maidanik (Carderock Div., Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

Noise control goals are often specified in terms of ratios of quadratic response quantities, among them a loss factor. Thus, the loss factor ( $\eta$ ) is defined as the ratio of the external input power density ( $\Pi e$ ) and the product of the stored energy density ( $E$ ) and the frequency ( $\omega$ ), i.e.,  $\eta = [\Pi e / (\omega E)]$ . A unique definition of the loss factor demands that the dynamic system be enclosed so that ( $\Pi e$ ) accounts for all the external input power density and ( $E$ ) accounts for all the stored energy density generated by ( $\Pi e$ ). Recently a number of acousticians in seeking noise control goals have defined the goals in terms of ratios that are assigned as loss factors. The definitions of these loss factors do not always conform to that uniquely defined loss factor. Those who maintain the conservative definition of a loss factor have been puzzled by reports of loss factors with values that exceed unity by several orders of magnitude. This paper attempts to examine a few examples of these less unique definitions of a loss factor and to show that a return to the unique definition is mandatory if the concept of a loss factor is not to be lost.

3:45

**2pSA10. Acoustical forced oscillation nondestructive testing method for meshing gears.** Leonid M. Gelman (Dept. of Nondestructive Testing, Natl. Tech. Univ. of Ukraine, 37, Peremogy pr., Kiev, 252056 Ukraine) and Alexandr S. Iievlev (SPU "Slavutich," Kiev-49, Ukraine)

Nondestructive testing and evaluation method of gears is considered. Differential equations of meshing gears forced oscillations are considered with time-variant piece-constant gear mesh stiffness and exciting force. New expressions of spectral density of mentioned oscillations are received for two cases: with defect and without one. For the first case modulation of gear mesh stiffness and exciting force is considered.

## Session 2pSC

## Speech Communication: Speech Processing (Poster Session)

Ashok K. Krishnamurthy, Chair

*Department of Electrical Engineering, The Ohio State University, Columbus, Ohio 43210**Contributed Papers*

To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m. To allow for extended viewing time, posters will be on display from 9:00 a.m. to 10:00 p.m.

**2pSC1. Telephone speech enhancement for elderly hearing-impaired listeners.** Amy E. Sheffield, Ashok Krishnamurthy (Dept. of Elec. Eng., The Ohio State Univ., Columbus, OH 43210, sheffield.20@osu.edu), Lawrence Feth, Stephanie Davidson, Evelyn Hoglund, and Lynette Roth (The Ohio State Univ., Columbus, OH 43210)

Many elderly persons with high-frequency hearing loss find telephone use frustrating due to lower intensity levels and reductions in acoustical information that can be useful in deciphering speech. The purpose of this project is to pre-process the speech signal before it is sent over the phone line and provide speech enhancement without the use of amplifying handsets or hearing aids at the receiving end. The enhancement technique takes into account the limited bandwidth of the phone line as well as the hearing characteristics of the user. Two pre-processing schemes, a single channel and a double channel approach, used to increase the intelligibility of speech in these situations are discussed. The single channel method performs amplitude compression of the entire signal. The two-channel method filters the incoming signal into high-frequency and low-frequency channels and performs independent compression on each before recombination. Results comparing the two speech enhancement schemes against no processing for a group of elderly hearing-impaired subjects are presented. [Work supported by a grant from the Franklin County Office on Aging.]

**2pSC2. High-frequency transfer function model of the vocal tract sections using FEM matrix condensation and transfer matrix techniques.** Samir El-Masri and Nobuhiro Miki (Grad. School of Eng., Hokkaido Univ., North 13, West 8, Kita-ku, Sapporo, 060-8628 Japan, selmasri@cho8-ei.eng.hokudai.ac.jp)

Using numerical tools such as the Finite Element Method (FEM) for acoustical modeling of the vocal tract would require a huge number of nodes and elements. In this paper, the matrix condensation and transfer matrix techniques [A. Craggs, *J. Sound Vib.* **132**, 393–402 (1989)] adapted to a three-dimensional model of the vocal tract section analysis will be introduced and discussed. Using these new FEM techniques, which use only the nodes at the input and output of the structure, the transfer function can be computed with a very small matrix comparing to standard FEM. The second goal of this research is to attempt to discover some geometrical parameters which could influence the transfer function. This investigation could be useful to make a new electrical vocal tract model which would consist of cascade transfer functions controlled by parameters. The study of the transfer functions will be carried out in low and high frequencies. [Work partly supported by Japanese CREST project.]

**2pSC3. Getting two birds with one phone: An acoustic sensor for both speech recognition and medical monitoring.** James D. Bass, Michael V. Scanlon, and Thomas K. Mills (Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783, jbas@arl.mil)

Automatic speech recognition (ASR) in noisy environments requires innovative use of disparate technology to overcome the special demands caused by multiple speakers and minimal signal-to-noise ratios (SNRs). Use of nonairborne acoustic sensors for ASR imposes special requirements on speech engines due to the changes in spectral information caused by alternative pickup locations. Specifically, the relative power of voiced and nonvoiced components is often reversed when compared to the relative power of these components collected with conventional microphone technology. Our research entails the evaluation of various ASR sampling configurations in conjunction with different body location points for the physiological sensor. Physiology provides clues to speakers' stress or cognition. For ASR, our goal is to build a suite of optimal sampling configurations for several strategic body locations (e.g., throat temple, thorax, etc.). The experimental design includes traditional word error rate and subjective task completion components. These experiments were conducted in environments with SNR ranges of 10, 3, 0, and  $-1$  dB. These SNR ranges cover the optimal commercial ASR environment of 10 dB to ranges where commercial ASR systems with conventional microphone technology are completely ineffective. Pilot studies indicate good performance below 0-dB SNR for a sensor that is throat located.

**2pSC4. Estimation of articulatory movement and its application to speech synthesis.** Jun Huang and Stephen Levinson (Dept. of ECE, Univ. of Illinois, Urbana, IL 61801, jhuang@ifp.uiuc.edu)

In this paper, two methods are investigated to estimate the movement of the vocal tract modeled by a three-dimensional articulator described by seven physical parameters. First, we use the cubic spline method to interpolate the articulatory parameters between consecutive phonemes. Second, the articulatory parameters are piecewise linearly interpolated and then passed by a cubic spline smoothing filter. The cubic spline smoothing filter is a low-pass filter with the maximum flatness property. The filter bandwidth can be easily adjusted by a control parameter. A graphical animation is built to visualize the articulatory movement of phonetic strings and to compare the results from the two different schemes. Some English phonemes are synthesized based on the articulatory model and the governing Webster equation in frequency domain considering viscous losses. An overlap-and-add method is used to convert the frequency domain spectrum to time domain speech signal. Finally, some example of the synthesized speech phonemes will be demonstrated. [We want to thank the National Science Foundation (NSF) for the support of our work.]



**2pSC5. Vocal-tract length normalization for acoustic-to-articulatory mapping using neural networks.** Sorin Dusan and Li Deng (Dept. of Elec. and Computer Eng., Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada, sdusan@speech2.uwaterloo.ca)

A new method of estimating the overall vocal-tract (VT) length and the normalization of acoustic parameters of different speakers is reported in this paper for acoustic-to-articulatory mapping. The main goal of this work was a high accuracy of VT length estimation from a short speech utterance. An articulatory model, originally developed by Maeda, was used as a reference female VT. Linear scaling was used to synthesize training data for VT lengths between 100% and 125% of the reference VT length (14.96 cm). These data had 250 utterances, resulted from different VT lengths, each containing six vowels. A neural network with two hidden layers was trained using vectors of 10 mel-frequency cepstrum coefficients and the corresponding VT lengths of these utterances. For the same VT length range, similar test data were synthesized using the training vowels but in different contexts. With the trained network, evaluation of this method on test data has shown an average error of less than 1% and a maximum error of 3.2% in estimating VT length from single test utterances. Frequency warping was used to normalize the cepstrum parameters according to estimated length factors ranging between 1.0 and 1.25. [This work was supported by NSERC.]

**2pSC6. Automatic ToBI prediction and alignment to speed manual labeling of prosody.** Ann K. Syrdal, Julia Hirschberg (AT&T Labs-Res., Florham Park, NJ 07932, syrdal@research.att.com), and Julie T. McGory (Ohio State Univ., Columbus, OH 43210)

ToBI (Tones and Break Indices) prosodic labeling of a speech corpus is a slow, labor-intensive process that typically takes from 100 to 300 times real time, even with experienced labelers. An experiment was conducted to determine: (1) whether manual correction of automatically assigned ToBI labels would speed up the labeling process; and (2) whether default labels introduced any bias in label assignment. A group of four graduate linguistics students previously trained in ToBI labeling were paid participants in the study. A large speech corpus of one female speaker reading several types of texts was labeled over a period of nine months. Half of each recording was labeled in the normal fashion "from scratch" without default labels, and the other half was presented with preassigned default labels for labelers to correct. Default ToBI labels were predicted from text using techniques developed for text-to-speech synthesis. Both labeling methods used standard Entropic waves+ tools developed for ToBI transcription. A log file was created during the labeling of each file that identified labeler, filename, method, and beginning and ending times of each session. Results indicate that labeling from defaults was faster than standard labeling, and that defaults had relatively little impact on label assignment.

**2pSC7. Evaluation of quality of speech enhanced by HMM and AR model-based systems.** Anisa Yasmin (Dept. of Elec. and Computer Eng., Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada, ayasmin@crg3.uwaterloo.ca), Paul Fieguth, and Li Deng (Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada)

Speech enhancement algorithms have demonstrated their application potential in a wide variety of speech communication contexts in which the quality or the intelligibility of speech has been degraded by the presence of background noise: hearing aids, cellular phones, public telephones, hands-free telephones, air-ground communications, etc. By far the two most popular choices for model-based speech enhancement are the Wiener-filter-based hidden Markov model (HMM), and the autoregressive (AR) model-based Kalman filter. Although researchers have been studying such enhancement systems for some time, relatively little has been undertaken in comparing the quality of enhanced speech produced by these two systems. This paper studies the Wiener/HMM and AR/KF models and conducts a comprehensive comparative study of the relative quality of enhanced speech, based on utterances from the TIMIT database contaminated by simulated and sampled empirical noises. The Wiener/HMM and

AR/KF comparison includes both subjective and qualitative evaluations: subjective assessments are based on mean opinion scores (MOS) and the inspection of temporal and spectrogram plots; objective evaluations are based on average and segmental signal-to-noise ratios. HMM enhanced speech has most of the noise removed, but with interruptions and discontinuities present due to the switched nature of the HMM. AR model-based enhancement possesses more audible background noise in the high-frequency region above 4 kHz, however the speech is smoother, with fewer discontinuities.

**2pSC8. Continuum mechanical model of the tongue and mouth floor.** Reiner Wilhelms-Tricarico (Institut de Commun. Parlee, INPG, Grenoble, France)

In this new model, the mouth floor and the tongue body are treated as blocks represented by tri-quadratic finite elements in which the muscles are modeled as continuous directed fields of the fiber direction. Collision with the hard palate and other rigid structures is modeled by subdividing the surface into small triangles for rapid collision detection and through calculating forces by a penalty method to enforce impenetrability of the rigid structures. A similar method is provided for the more complicated case of the collision between tongue body and soft floor of the mouth. The model can be coupled with a model of the jaw and hyoid where muscles may be modeled as contracting strings. The moving hyoid and jaw provide kinematic constraints for the tongue/mouth-floor model. Active muscle stress is generated by a Hill model of muscle shortening combined with a rational extension for muscle elongation, and a polynomial curve represents the relation between muscle length and stress. The stiffness matrix and damping matrix of the system (used for the implicit time stepping method solving the equations of motion) are derived directly from the continuum mechanical muscle model. An overview of the theory and numerical examples will be presented.

**2pSC9. Measurement error compensation using integration for hidden Markov-model-based speaker recognition.** Marie A. Roch (Dept. of Computer Sci., The Univ. of Iowa, Iowa City, IA 52242) and Richard R. Hurtig (The Univ. of Iowa, Iowa City, IA 52242)

Like all such measurements, the source data for speaker recognition is subject to errors in measurement due to transducer, channel, and quantization effects. When training and test equipment are known, one may calibrate the hardware and introduce a noise compensation procedure into the recognition process. In many applications, it is highly desirable for speaker recognition tasks to function with a wide variety of unknown equipment, making calibration impractical. A study is presented with the *a priori* assumption that for each feature vector  $\vec{o}$  observed with measurement noise, an error compensated vector  $\vec{\delta}$  lies within some uniformly distributed interval  $\pm \epsilon$  of the observed vector. A statistic derived from the observation set is computed and used to estimate an empirical interval in the neighborhood of each observation. An approximation to integration over the interval is carried out and is used in place of the density measurement at  $\vec{o}$ . Tests using a 15-component cepstral feature vector derived from telephone quality speech (King corpus, San Diego speakers, sessions 1-5) have shown reductions of error rate on the order of 15% as compared to a baseline system. Techniques to reduce the algorithmic cost of the integration will also be discussed.

**2pSC10. Detecting filled pauses in spontaneous speech.** Douglas O'Shaughnessy (INRS-Telecommunications, 16 Pl. du Commerce, Nuns Island, QC H3E 1H6, Canada)

Practical speech recognizers must accept normal conversational voice input (including hesitations). However, most automatic speech recognition work has involved read speech, whose acoustic aspects differ significantly from speech found in actual dialogues. Hesitations, filled pauses, and re-starts (after aborted utterances) are common in natural speech, yet few recognition systems handle such disfluencies with any degree of success.

Among other problems, filled pauses (e.g., “uhh,” “umm”), unlike silences, resemble phones as part of words in continuous speech. The work reported here further develops techniques to allow identification of filled pauses. A distinction is made between disfluencies in actual dialogs (e.g., in the Switchboard database of natural telephone conversations, which have poor recognition rates so far) and simulated ones (e.g., the ATIS Wizard-of-Oz-style database of airline travel inquiries). It appears that speaking with actual people influences disfluencies, e.g., filled pauses tend to be shorter and more variable in pitch patterns, although unfilled pauses adjacent to filled ones remain important in both styles. While most automatic recognition methods rely entirely on spectral envelope (e.g., low-order cepstral coefficients), identifying hesitation phenomena seems to require use of fundamental frequency and duration in addition to such spectral parameters.

**2pSC11. DSP circuit to improve speech intelligibility.** Michael J. Metz (Audiology Assoc., P. C., 14150 Culver Dr., Ste. 207, Irvine, CA 92604) and Neil A. Shaw (MenloScientific Acoustics, Inc., Topanga, CA 90290-1610)

Two studies, one equipment based and one clinically based, were attempted to verify the contention that a DSP circuit is capable of changing the intelligibility of speech in noise. The first used two intelligibility metrics (RASTI and %ALcons) to measure the change in metric score under the conditions of test signal with (1) no DSP speech signal processing of the test signal, and (2) with DSP speech processing of the test signal in the presence of pink noise in the background. The second study measured the articulation functions of normal ears using phonetically balanced word lists under conditions of (1) no speech signal processing, and (2) “maximum” DSP processing when each condition was varied in the presence of a set of level of pink noise. The algorithm used in this study is the “Voice Intelligibility Processor,” or VIP signal processor. The studies indicate an increase in the intelligibility metric scores and the measured articulation functions. For the articulation studies, when the processor is “normalized” to account for the perceived increase in loudness level produced by the processor algorithm, the mean shift in “50% point” of the articulation function was approximately 6 dB. For the RASTI study, similar improvement was also found.

**2pSC12. The use of spectral versus temporal cues to recognize speech.** Arindam Mandal, Laura J. Davis, Carol Espy-Wilson, and Melanie Matthies (Boston Univ., Boston, MA 02215)

Previous work has shown that listeners are able to use temporal information to identify spectrally limited consonants [Shannon *et al.*, Science 303–304 (1995)]. In this study, we used spectrally impoverished speech to: (1) investigate the performance of a temporal-based automatic classifier; (2) describe the performance of human listeners; and (3) compare the performance of the classifier with human listeners. For the automatic classifier, acoustic events are extracted from a cross-spectral temporal measure that computed onsets and offsets of acoustic energy. The strength and time difference between these acoustic events served as input to the classifier. The results from human listeners showed that their ability to identify nasality and voicing remained fairly good. In addition, recognition of manner for stops and fricatives remained high. As the speech signal was degraded, listeners often confused sonorants with fricatives. The temporal-based classifier did not make such confusions. The classifier is able to identify manner, voicing and nasality well and the results remain fairly consistent across different levels of degradation. Thus, these results show the importance of temporal information and they suggest that human listeners may not optimally combine all cues to identify spectrally impoverished consonants. [Work supported by ONR Supervised MURI grant Z883402.]

**2pSC13. Comparative study of  $F_0$  extractors for high-quality speech synthesis.** Hideki Kawahara (Faculty of Systems Eng., Wakayama Univ./ATR/CREST, 930 Sakaedani, Wakayama, 640-8510 Japan) and Parham S. Zolfaghari (ATR Human Information Processing Res. Labs., Kyoto, Japan)

Performance of a new  $F_0$  extraction algorithm based on fixed point analysis of filter center frequency to output instantaneous frequency [Kawahara *et al.*, Eurospeech'99] was compared with numbers of  $F_0$  extraction algorithms based on different definitions of fundamental frequency. The proposed method uses partial derivatives of the mapping at fixed points to estimate carrier to noise ratio of  $F_0$  information. It also enables integration of distributed  $F_0$  cues among harmonic components to provide a reliable  $F_0$  estimate. Objective evaluations were conducted using simulations and a speech database with simultaneous EGG (electroglottograph) recording. Subjective evaluations were based on reproduced speech quality assessment by a high-quality speech analysis/modification/synthesis method STRAIGHT [Kawahara *et al.*, Speech Commun. 27, 187–207]. Discussions about the relevance of various  $F_0$  definitions for high-quality speech synthesis will be presented based on these test results. It was also indicated that the proposed method is tunable to specific needs.

**2pSC14. A vowel synthesizer based on  $F_0$ -modulated formant sinusoids.** Ingo Hertrich and Hermann Ackermann (Dept. of Neurology, Univ. of Tuebingen, Hoppe-Seyler Str. 3, D-72076 Tuebingen, Germany)

A vowel synthesis algorithm was developed resembling parallel formant synthesizers. However, formants are computed as amplitude- and phase-modulated sinusoids instead of being represented by recursive filters or resonators. The fundamental frequency is imposed on the signal in the following way: Each pitch period starts with a short initial rise of a duration of  $\pi/2$  of the first formant frequency; at the end of this ramp phase all formants start as sinusoids at pre-defined phase angles and amplitudes, and successively undergo linear attenuation toward zero amplitude at the end of each pitch period. Adjacent pitch periods, thus, do not influence each other. In principle, the signals produced by this algorithm resemble the output of parallel formant synthesizers [e.g., D. H. Klatt, J. Acoust. Soc. Am. 67, 971–995 (1980)]. However, by incremental sample-by-sample computation of the formants' phase angles, formant frequencies are allowed to continuously change within single pitch periods. This algorithm produces vowels as well as formant transitions signaling stop consonant–vowel syllables such as /ba/, exhibiting a homogeneous, stereotypic voice quality. Because of its additive procedure, this method might allow for the implementation of a variety of further acoustic aspects of the human speaking voice in a well-controlled manner.

**2pSC15. Problems of signal processing in modern wireless terminals and its impact on speech quality.** Ekkehard Diedrich and Hans Wilhelm Gierlich (Goslarer Ufer 35, D-10589 Berlin, Germany)

Modern terminal applications in speech communications are more and more including signal-processing technologies, such as switching devices, echo cancellers, nonlinear processors, level control, and compander a.o. Apart from the benefits of such technologies, there are doubtless impacts on the speech transmission quality like echo, level fluctuations, and clipping. Even modern terminal developments, e.g., for mobile communications lead to shorter and compact handset designs with some negative influences on the speech transmission quality from an acoustical point of view. One example is echo because of the stronger internal coupling between microphone and loudspeaker in very short handsets and the need for echo control to avoid the negative impact on speech quality. New test and measurement procedures (subjective and objective) are described and discussed to give the possibility of classifying these impacts on speech transmission quality.

**2pSC16. Improvements to segmentation accuracy using a large speech database.** Alistair Conkie (AT&T Labs.-Res., 180 Park Ave., Florham Park, NJ 07932, adc@research.att.com)

Current text-to-speech (TTS) systems rely increasingly on large speech databases for both the raw speech necessary for concatenative synthesis and for material used for modeling prosody. It is time-consuming to label the databases phonetically so that they can be useful. It is nevertheless important for the quality of the resulting synthesis that the labeling be accurate. One currently used method involves some kind of automatic labeling – usually using hidden Markov modeling (HMM) techniques, followed by manual correction. This is still time-consuming. Recently, work has been done on post-processing the output from an HMM recognizer with a fuzzy logic system [Torre Toledano *et al.*, Third ESCA/COCOSDA Workshop on Speech Synthesis, Australia, November 1998] to improve accuracy. We examine this promising idea by applying post-processing to automatically generated segmentation of the speech signals in our TTS database. The effectiveness of the technique is evaluated both in terms of labeling accuracy in comparison to hand-labeled data, and with informal listening tests for synthesis quality.

**2pSC17. Flexible post-lexical processing for speech synthesis from a large unit selection database.** Mark Beutnagel (AT&T Labs.-Res., Shannon Lab., 180 Park Ave., Florham Park, NJ 07932, mcb@research.att.com)

Online unit selection from large speech databases provides an opportunity to essentially play back words, phrases, and even sentences which were included in a recorded corpus. This capability can be extremely useful for limited domains, e.g., application prompts. Without switching voices, such a synthesizer could integrate high-quality synthesis with near-perfect recorded material. However, traditional post-lexical processing (PLP) considers only the phoneme specifications and not the sequences which actually exist in the target database. Phonemes supplied by the dictionary are typically rewritten into a single sequence of phones with reduced vowels, flapped t's, etc. Given the enormous variability of human speech, any single sequence is unlikely to match an entire phrase or prompt as spoken and labeled. This paper addresses the use of flexible PLP, allowing multiple transcription possibilities which are essentially equivalent, at least for the speaker in question. By building the equivalences from the specific dictionary used by the synthesizer and the detailed phonetic labeling of a specific voice database, longer regions of the database can be selected, reducing the number of concatenation points in ordinary synthesis and increasing the odds of selecting complete recorded phrases.

**2pSC18. Distributed acoustic model of the entire respiratory tract.** Paul Harper (1285 Elec. Eng. Bldg., Purdue Univ., West Lafayette, IN 47907, pharper@ecn.purdue.edu), Hans Pasterkamp (Univ. of Manitoba, Winnipeg, Canada), Steve S. Kraman (VA Medical Ctr., Lexington, KY), and George R. Wodicka (Purdue Univ., West Lafayette, IN 47907)

The analysis of breathing sounds measured over the extrathoracic trachea has been used to detect and monitor respiratory tract changes such as those found in asthma and obstructive sleep apnea. To begin to link the attributes of these easily measured sounds to the underlying anatomy, a lumped-element model of the acoustic properties of the entire respiratory tract (supraglottal plus subglottal airways) with varying cross-sectional area, yielding walls, and dichotomous branching in the subglottal component was developed over the frequency range from 100 to 3000 Hz. The portions of the model above and below the larynx (supra- and subglottal) predict well the distinct locations of the spectral peaks from speech sounds such as /ah/ as measured at the mouth and the trachea, respectively, in healthy subjects. When combining the supra- and subglottal portions in a complete tract model, the predicted peak locations compare favorably with those of tracheal sounds measured during normal breathing in the same subjects. In addition, the model predicts both the direction and magnitude

of the relatively small spectral shifts that occur in tracheal sounds during a neck flexion/extension maneuver, indicating that measurable acoustic parameters are sensitively tied to respiratory tract dimensions.

**2pSC19. Effect of a tracheal tumor on breathing sounds: A case study.** Steve S. Kraman (VA Medical Ctr., 2250 Leestown Rd., Lexington, KY 40511, sskram01@pop.uky.edu), Paul Harper (Purdue Univ., West Lafayette, IN), Hans Pasterkamp (Univ. of Manitoba, Winnipeg, Canada), and George R. Wodicka (Purdue Univ., West Lafayette, IN)

A 67-year-old man came to the hospital with a history of shortness of breath on exertion and cough progressing to severe breathlessness at rest over several weeks. Physical exam revealed stridorous sounds in inspiration and expiration heard at the mouth and over the trachea. Lung function tests revealed severe airway obstruction in expiration and probably inspiration. Fiberoptic airway inspection revealed a large mass above and arising from the main carina (first airway bifurcation). This mass obstructed the distal trachea except for a narrow (~2 mm) circumferential space between the mass and the tracheal wall. Breathing sounds were recorded from in front of the mouth and the anterior neck overlying the trachea both before and after laser ablation of the mass. Before ablation, the tracheal sound exhibited unusual resonances and antiresonances that rose in frequency by nearly 100% during inspiration and fell during expiration. These variable sounds disappeared after laser ablation of the mass. It is postulated that the mass behaved as a dynamic acoustic obstruction and thereby altered the distal boundary condition and resonance behavior of the respiratory tract during breathing. This provides insights into the links between the airway anatomy and measurable acoustic properties of breath sounds.

**2pSC20. Speaker verification performance comparison based on traditional and electromagnetic sensor pitch extraction.** T. J. Gable, L. C. Ng, G. C. Burnett, and J. F. Holzrichter (Lawrence Livermore Natl. Lab., P.O. Box 808, Livermore, CA 94551)

This work compares the speaker verification performance between a traditional acoustic-only pitch extraction to a new electromagnetic (EM) sensor based pitch approach system. The pitch estimation approach was developed at the Lawrence Livermore National Laboratory (LLNL) utilizing Glottal Electromagnetic Micropower Sensors (GEMS, also see <http://speech.llnl.gov/>). This work expands previous pitch detection work by Burnett *et al.* [IEEE Trans. Speech and Audio Processing (to be published)] to the specific application of speaker verification using dynamic time warping. Clearly, a distinct advantage of GEMS is its insensitivity to acoustic ambient noise. This work demonstrates the clear advantage of the GEMS pitch extraction to improve speaker verification error rates. Cases with added white noise and other speech noise were also examined to show the strengths of the GEMS sensor in these conditions. The EM sensor speaker verification process operated without change over signal-to-noise (SNR) conditions ranging from -20 to -2.5 dB; the acoustic algorithms became unusable at SNR exceeding -10 dB. [Work supported by NSF and DOE.]

**2pSC21. The use of glottal electromagnetic micropower sensors (GEMS) in determining a voiced excitation function.** Gregory C. Burnett, John F. Holzrichter, Larry C. Ng, and Todd J. Gable (Lawrence Livermore Natl. Lab., P.O. Box 808, L-271, Livermore, CA 94551)

Recent experiments using a portable, extremely low-power electromagnetic motion sensor to detect the motion of the posterior tracheal wall during speech production will be presented. The motion of the wall may be related to the driving subglottal pressure through a lumped element circuit model, leading to an approximation to the voiced excitation function of the human vocal tract. Using the excitation and the recorded spoken audio, a stable and accurate transfer function of the vocal tract may be calculated every few glottal cycles in near real-time. The excitation func-



tion may be used to calculate very accurate pitch information at low cost, and the transfer functions may be employed as an additional feature vector to enhance the performance of a new class of speech recognizers and synthesizers. [Work supported by NSF and DOE.]

**2pSC22. Background speaker noise removal using combined EM sensor/acoustic signals.** Lawrence C. Ng, John F. Holzrichter, Gregory C. Burnett, and Todd J. Gable (Lawrence Livermore Natl. Lab., P.O. Box 808, L-054, Livermore, CA 94551)

Recently, very low-power EM radarlike sensors have been used to measure the macro- and micro-motions of human speech articulators as human speech is produced [see Holzrichter *et al.*, *J. Acoust. Soc. Am.* **103**, 622 (1998)]. These sensors can measure tracheal wall motions, asso-

ciated with the air pressure build up and fall as the vocal folds open and close, leading to a voiced speech excitation function. In addition, they provide generalized motion measurements of vocal tract articulator gestures that lead to speech formation. For example, tongue, jaw, lips, velum, and pharynx motions have been measured as speech is produced. Since the EM sensor information is independent of acoustic air pressure waves, it is independent of the state of the acoustic background noise spectrum surrounding the speaker. By correlating the two streams of information together, from a microphone and (one or more) EM sensor signals, to characterize a speaker's speech signal, much of the background speaker noise can be eliminated in real time. This paper presents several algorithms to demonstrate the added noise suppression capability of the glottal EM sensors (GEMS). [Work supported by NSF and DOE.]

TUESDAY AFTERNOON, 2 NOVEMBER 1999

GRANT-HARDING ROOM, 3:15 TO 5:15 P.M.

## Session 2pSP

### Signal Processing in Acoustics: Time-Frequency Applications in Acoustics II

Patrick J. Loughlin, Chair

*Department of Electrical Engineering, University of Pittsburgh, 348 Benedum Hall, Pittsburgh, Pennsylvania 15261*

Chair's Introduction—3:15

#### *Invited Papers*

3:20

**2pSP1. Reduced interference distribution application and interpretation for transient acoustic events.** William J. Williams (EECS Dept., Univ. of Michigan, Ann Arbor, MI 48109, wjw@eecs.umich.edu)

Transient acoustic events often present difficulties when using conventional time-series or spectral methods. The Reduced Interference Distribution (RID) and its derivatives have been under development [W. J. Williams and J. Jeong, *Time-Frequency Signal Analysis*, edited by B. Boashash (Longman Cheshire, New York, 1992)] for several years. The RID was specifically motivated by the need for a better analysis tool for transient acoustic events [ONR Grants N00014-90-J-1654, N00014-97-1-0072]. The deficiencies of the spectrogram for marine mammal sound analysis were early recognized [William Watkins, *Marine Bioacoustics* **2**, 15–43 (1966)]. Watkins' illustrations of spectrogram deficiencies and updated results obtained using the Wigner distribution (WD) and the RID will be presented. The RID represents an attempt to retain the desirable mathematical properties of the WD while reducing the often troublesome WD interference terms. Examples drawn from several acoustic transients such as marine mammal sounds, joint clicks and vortex formations in turbulent air flow will be used to illustrate the usefulness of the RID approach. The importance of retaining such properties as proper time and frequency marginals; time and frequency support along with proper group delay and instantaneous frequency will be discussed. The value of covariance with time and frequency shifts and scale changes will also be emphasized.

3:40

**2pSP2. Characterization of densities by moments.** Leon Cohen (Dept. of Phys., City Univ., New York, NY 10021)

For two-dimensional classical densities it is very often the case that the general properties of a density can be characterized by low-order joint moments and conditional moments. The advantage of this is that moments and conditional moments are constants and one-dimensional functions, respectively, and hence can be more effectively used for characterization, classification, and detection than the full density. In the time-frequency case unique problems arise. First, the definition of conditional moments is problematic and has not been fully investigated. Second, since a time-frequency density may go negative the methods of probability theory may not be taken over directly. We investigate these issues and show that time-frequency conditional moments can be defined in such a way that they appropriately characterize the general features of a time-frequency density. [Work supported by ONR.]

4:00

**2pSP3. Further applications of time-frequency methods in acoustics.** G. C. Gaunaud (Naval Surface Warfare Ctr., Carderock Div., Code 683, W. Bethesda, MD 20817-5700) and H. C. Strifors (Defense Res. Establishment, FOA, Stockholm, S-17290, Sweden)

Time-frequency ( $t$ - $f$ ) methods have been repeatedly established as effective tools to successfully analyze a variety of signatures from many types of sources or scatterers. Many such signatures have been acoustic in nature, and we have studied these most extensively in the past. We have presented several of these examples before [viz., G. C. Gaunaud and H. C. Strifors, *J. Acoust. Soc. Am.* **104**, 1746, 136th ASA Mtg. (1998)]. Here, we will cover acoustical applications that we have analyzed, but which were not adequately covered in 1998. We will also outline additional ones—more extensively discussed in *Appl. Mech. Rev.* **50**, 131–149 (1997)—and a recent radar application [i.e., *ibid.* *IEEE Trans. Antennas Propag.* **46**, 1252–1262 (1998)]. In this case, the present  $t$ -



*f* processing was shown capable of successfully identifying an air target that had been covered with a dielectric RCS-reducing layer. We believe that this general approach, jointly with the way we have used it to physically interpret broadband sonar/radar signatures, can lead to straightforward implementations for many practical target-classification problems associated with many types of sensors. [Work partially supported by the ILIR Program of the authors' Institutions and the ONR.]

4:20–4:30 Break

### Contributed Papers

4:30

**2pSP4. Time-frequency analysis of the backscattering by tilted finite cylindrical shells.** Scot F. Morse<sup>a)</sup> and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814, morse@lpsa2.nrl.navy.mil)

Previous experiments and analysis have shown that the backscattering from thick-walled finite cylindrical shells can be significantly enhanced in and above the coincidence frequency region over a large range of angles of the incident sound [S. F. Morse *et al.*, *J. Acoust. Soc. Am.* **103**, 785–794 (1998)]. In this presentation, impulse backscattering measurements performed with a wide bandwidth PVDF sheet source are analyzed in terms of a simple joint time-frequency representation. When plotted for a sequence of specific tilt angles, it is possible to identify and then observe the progression of individual ray contributions as the coupling conditions to the shell are changed. In the frequency range investigated, where  $ka < 45$  and  $kh < 3.4$ , several of the surface waves excited are highly dispersive and as a result are launched on the shell over a wide range of frequencies and tilt angles, including end-on incidence. The analysis described also confirms various features of the responsible scattering mechanisms predicted by ray theory. The results show the importance of time-frequency analysis to the identification of back scattering in a situation where no exact reference solution is available. [Work supported by the Office of Naval Research.] <sup>a)</sup>Currently with Naval Research Laboratory, Code 7136, 4555 Overlook Ave. SW, Washington, DC 20375.

4:45

**2pSP5. A new method for an almost-on-the-fly estimation of the instantaneous frequency and of the amplitude of a signal.** Gerard Girolami and Lucile Rossi (URA 20253 CNRS, SDEM, Equipe Ondes et Acoustique, Universite de Corse, Faculte des Sci., BP 52, 20250 Corte, France, girolami@univ-corse.fr)

The instantaneous frequency (IF) of a signal is a parameter whose estimation is of prime importance in a number of domains ranging from musical acoustics to seismics and astronomy. Its determination is usually based on one of the following three general methods: the derivation of the phase of the associated analytic signal, the Wigner–Ville transform and

others Cohen's class transforms, or the three-point algorithms like the Teager–Kaiser "energy operator." A new three-point method is described which is simple, fast, precise, robust to noise, and estimates the IF and the amplitude with a very small delay. This method gives excellent results on almost any kind of chirps (polynomials, hyperbolics, exponentials, etc.) and frequency modulations (sinusoidal, FSK, Gaussian pulse, etc.), whether or not associated with amplitude modulation (sinusoidal, Gaussian, etc.). The relative errors on the amplitude and on the IF are usually around 1% in the no-noise case, and increase to about 5% for a 20-dB signal-to-noise ratio, even when these different modulations are associated.

5:00

**2pSP6. Detection and identification for highlight echo using the time-frequency postaccumulation method.** Zhu-Ye and Gong-suying (Inst. of Acoust., Academic Sinica, P.O. Box 2712, Beijing 100080, PROC)

The highlight distribution of underwater target in active sonar is the main characteristic for target identification. The pulse compression with chirp signal is a current method for target echo detection. But sonar highlight echo is always time-frequency fading and the background reverberation is nonstationary, so that the matched filter is not optimum to detect the highlight echo. Based on that a linear-frequency modulation signal, the peak trace of the ambiguity function (AMF) or the Wigner–Ville distribution (WVD), is linearity, the accumulations along the peak trace (APT) (analog to Radon transform in image processing) of both the WVD and cross-AMF of multi-target echo are developed in this paper. The method is applied to a simulated multi-target echo and practical highlight echoes of the target. The detection characteristics and performances are comparison for both accumulations with matched filter. The results of the analysis and simulation are shown: (1) In detection performances, the APT–WVD and APT–CAMF are superior to the matched filter if the accumulating time is selected properly and especially if the background spectra is not even. (2) In multi-target resolution, the APT–WVD is superior to the APT–CAMF. (3) The APT–WVD need not know the reference signal which can be estimated by this method.

TUESDAY AFTERNOON, 2 NOVEMBER 1999

FAIRFIELD ROOM, 1:00 TO 5:20 P.M.

### Session 2pUW

### Underwater Acoustics and Engineering Acoustics: Robert J. Urick Memorial Session

David L. Bradley, Chair

*Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804*

Chair's Introduction—1:00

### Invited Papers

1:05

**2pUW1. Source level and Urick.** James E. Barger (BBN Technologies, 70 Fawcett St., Cambridge, MA 02138)

Source level (SL) is the first term in Bob Urick's sonar equation, and it represents the root-mean-square pressure of sound radiated at the target, in decibels. The fundamental measure of source level is the signal-energy level (SEL) represented by the time integral of sound pressure squared, to which the detection index is proportional, in decibels. Despite using a form of sonar equation more

natural for long-pulse transducers than for impulsive ones, he only described explosives in detail in his chapter on generation of underwater sound. When “Principles of Underwater Sound for Engineers” was published, there was a tendency to use impulsive sound sources to study sound interactions with the sea and to use narrow-band long-pulse transducers in sonar systems. The diminution of this tendency has been a very important development since Bob’s book was written, and it has occurred because of improvements in both kinds of transduction. Transducer bandwidth has been increased, both by materials with larger electromechanical coupling coefficients and by innovative designs. These include “hybrid” transducers that combine magnetostrictive and electrostrictive actuators and projector arrays designed to have no mechanical reactance. Improved impulsive sound sources such as air guns, sparkers, and explosive arrays now find sonar use.

### 1:35

**2pUW2. Transmission loss since Urick.** Charles W. Spofford (Science Applications Intl. Corp., 1710 Goodridge Dr., McLean, VA 22102)

Advances in understanding propagation of “underwater sound” are traced from the view expressed in Urick’s 1967 landmark text to our current state through two remarkable transitions. First, propagation in 1967 was an arcane phenomenon with annoying, questionably modelable “anomalies” from spherical spreading which forced extensive global measurements. It is now known to be a stable, well-understood property of the ocean whose predictability is limited only by our knowledge of the controlling environmental inputs. In fact, models are now being used to deduce that environment through inversions. Second, the “TL” term in the sonar equation has expanded to include coherence measures inextricably coupled to the “gain” terms. These are being routinely modeled, again limited mainly by our knowledge of the environmental inputs. The variety of approaches to modeling propagation hinted at in Urick has blossomed to form a powerful set of tools for designing and analyzing a wide range of sonars. Geo-political ASW concerns gave birth to “underwater sound” for engineering sonar systems of ever increasing complexity and challenged us in 1967 to understand the physics of deep-water propagation. In 1999 our challenge is to extrapolate that physics to the world’s enormously variable shallow-water areas.

### 2:05

**2pUW3. Target strength.** Louis Dragonette (Naval Res. Lab., Code 7132, Washington, DC 20375-5350)

Chapter 9 of Urick’s *Principles of Underwater Sound* is titled “Reflection and Scattering by Sonar Targets: Target Strength” [R. Urick, *Principles of Underwater Sound* (McGraw–Hill, New York, 1975), p. 263], and while this chapter gives an excellent review of target strength definitions, simple echo formation processes, and echo characteristics, it is the table, “Target Strength of Simple Forms,” compiled by Urick from several obscure sources, that is invaluable to those who deal with target strength issues on a routine basis. A typical first approximation to the target strength (TS) of a body of interest is made by approximating the scatterer as a combination of the simple shapes listed in the table, and then summing up the separate rigid-body responses to produce a first-order target strength prediction. Some common-sense principles that derive from such an exercise are discussed as well as comparisons between such first-order predictions and the results and understanding obtained from modern theoretical and experimental approaches. [Work supported by ONR.]

### 2:35

**2pUW4. Bob Urick and ambient sea noise.** Raymond C. Cavanagh (Science Applications Intl. Corp., 1710 Goodridge Dr., McLean, VA 22120)

Selected highlights of progress in the measurement and understanding of ambient sea noise are presented, from the time of publication of the first edition of Urick’s pioneering textbook (1967) through his noise monograph (1984) and beyond. Urick’s presentation of the fundamentals became the primer for a generation of ocean acousticians and the basis for many of today’s theories and predictive capabilities. Valuable insights were derived from his direct involvement in the measurement and interpretation of noise data, dating to the early 1950’s. His treatments of arctic and shallow-water noise anticipated research initiatives of the 1980’s and 1990’s, as did his framework for noise fluctuations. Some notable advances since Urick’s publications are discussed, including: mechanisms for wind-wave noise sources, predictive models for ship-generated noise properties, the noise “floor” and noise increases over time, directionality, and the effects of coastal sources. His attention to the details of metrics and units for ocean acoustic quantities is acknowledged as a special contribution to the field.

### 3:05–3:20 Break

### 3:20

**2pUW5. Properties of transducer arrays: Directivity index (Urick, 1967).** Arthur B. Baggeroer (MIT, Cambridge, MA 02139, abb@arctic.mit.edu)

The chapter, “Properties of Transducer Arrays: Directivity Index,” in Urick’s text includes not only the directivity index (DI) of an array, but also many of the important metrics of spatial performance which remain important in assessing the performance of a sonar system. Beampatterns, tapers and sidelobes, signal gain, noise gain and even superdirectivity are all included. The DI metric in the sonar equation indicates gain against 3-D isotropic noise; alternatively, it is used to convert a beam noise level to power per steradian. While easy to compute and appropriate when a field is diffuse, the DI is not a good metric for spatially discrete noise such as shipping which can often dominate a noise field. Recently, the signal and noise gain metrics are used more often. Signal gain is an

issue as sonar systems use very large apertures; noise gain is also an issue when the directionality of the noise fields is exploited. Certainly, the most significant evolution in spatial processing not included in Urick's chapter was the use of adaptive processing which has been enabled by the availability of high-speed DSP for real time processing. The presentation will review Urick's important contributions to describing the performance of arrays and how they have impacted sonars.

3:50

**2pUW6. Detection threshold.** Henry Cox (ORINCON Corp., 4350 No. Fairfax Dr., Ste. 470, Arlington, VA 22203)

Detection threshold or recognition differential is defined as the signal-to-noise ratio at the beamformer output required for detection. An introduction to this topic is provided in Urick's *Principles of Underwater Sound*. The subject is frequently a source of confusion due to the multiple possible definitions of signal-to-noise ratio, the effects of nonlinearities in the signal processing, and the overall complexity of the sonar operators job of signal recognition in real-world clutter. In this paper, a review of the basic theory of mathematical relationships are provided, followed by a discussion of practical limitations in applying simple formulas to performance predictions. The relationships of detection threshold to the signal-processing approach and background-noise statistics are discussed. Active sonars in both noise and reverberation, as well as passive sonars involving energy detection, spectral analysis, and cross correlators are discussed.

4:20

**2pUW7. Reverberation: Urick and beyond.** D. V. Holliday (Marconi, 4669 Murphy Canyon Rd., San Diego, CA 92123-4333)

In his classic textbook, *Principles of Underwater Sound for Engineers*, Urick dedicated more space to reverberation, than to any other subject except propagation loss. His characterizations of reverberation were practical, easily understood, and remarkably complete for the era in which they were written. The chapter on scattering, especially the references, was an invaluable starting point for both students and practicing designers of active sonars. Urick, the sonar designer, pragmatically asserted that reverberation was a "particularly obnoxious" phenomenon. However, Urick, the scientist, demonstrated an unbridled curiosity about the underlying causes of the phenomena that he observed in the sea. Near the end of his active career in acoustics, the patterns of research into sound scattering changed dramatically. Urick personally encouraged young scientists to pursue a detailed understanding of the causes of scattering. Reverberation, which clearly had been a part of the "noise" in the sea became a "signal" which could be used to study and describe the sea surface, the seabed, and the various particles, animals, and bubbles found in the water column. This paradigm shift and some of its consequences, e.g., the extension of measurements to both higher and lower frequencies will be examined. [Work supported by ONR Code 322BC.]

### Contributed Papers

4:50

**2pUW8. Perspectives on teaching: A tribute to Bob Urick.** Paul C. Etter (Northrop Grumman Corp., Oceanic and Naval Systems, P.O. Box 1488, M.S. 9115, Annapolis, MD 21404, paul\_c\_etter@md.northgrum.com)

Based on 5 years of teaching short courses with Bob Urick, an assessment is made of his teaching style and elements of content. Urick's influence on the underwater-acoustics community is exemplified by the popularity of his books and short courses. The fact that much of Urick's nomenclature became institutionalized within the international community provides further evidence of his authoritative command of the technical literature and his lucid conveyance of fundamental concepts. Specific examples are drawn from personal experience to demonstrate the effective aspects of his instructional methods, including both style and content. A synthesis of Urick's methods can instruct and inform contemporary educators in the field of underwater acoustics and related disciplines. His equally important, but less tangible, impact on the professional development of emerging sonar engineers is also discussed from the dual perspective of a co-instructor and former student. Finally, Urick's conception of the complementary roles of field measurements and numerical modeling is reviewed to characterize the proper relationship between measurements and modeling in modern ocean-acoustics research.

5:05

**2pUW9. Array signal-gain measurements in shallow-water.** William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215) and Peter G. Cable (BBN Technologies, New London, CT 06320)

Shallow-water transverse coherence lengths can be estimated by the measurement of the narrow-band coherence function, broadband correlation function, or the signal gain either from a direct measurement with a filled aperture or the steered beam response of a sparsely filled aperture. Signal-gain measurements with an array of sensors have a larger number of degrees of freedom, spatial filtering of unwanted noise sources, and higher signal-to-noise ratios. Given a Gaussian coherence function, the coherence length may be estimated from these gain measurements. This paper discusses signal-gain measurements performed in five shallow-water sandy-bottomed areas with known environmental conditions using horizontal-bottomed arrays, omnidirectional explosives, and continuous sources between 100 and 600 Hz to ranges of 40 km. The coherence lengths determined by using a Gaussian coherence function yield a consistent representative value of approximately 30 wavelengths at a range of 40 km and a frequency of 400 Hz, which decreases with an increase in frequency. Comparisons show these lengths are consistent with previous measurements. Analytical considerations indicate these lengths are due to the combined effects of the water column variability and bottom. Implications for shallow-water array processing are discussed.

## Session 3aABa

## Animal Bioacoustics: Detecting and Identifying Animals Using Acoustics II

David Helweg, Chair

SPAWARSSYSCEN-San Diego, Code D351, 53560 Hull Street, San Diego, California 92152-5435

Chair's Introduction—7:55

## Contributed Papers

8:00

**3aABa1. Signal processing techniques for passive acoustic location and tracking of animals.** Christopher W. Clark and Kurt M. Fristrup (Bioacoust. Res. Prog., Cornell Univ., Sapsucker Woods Rd., Ithaca, NY 14850)

Passive acoustic location and tracking of free-ranging animals has high value for population abundance estimation, mitigation of man-made activities, and understanding the biological significance of acoustic behavior. Applications have had limited success: e.g., in the Arctic for long-term study of an endangered bowhead population, in temperate waters with deep-diving sperm whales, and for behavioral studies on birds, dolphins, and seals. The Arctic situation is fairly unique, the whales produce discrete FM signals in the 100–400-Hz band, the water is isothermic, and propagation is relatively coherent. The task of locating and tracking marine animals in environments with more complex propagation conditions is more challenging, and different combinations of environment and species demand slightly different signal-processing solutions. Methods developed for fin, humpback, and sperm whales; elephants; and birds will be presented, and their relative merits will be discussed.

8:15

**3aABa2. Techniques for passive detection and localization of marine mammals using wide aperture arrays.** Joseph E. Bondaryk, Angela D'Amico, and Nicola Portunato (SACLANT Undersea Res. Ctr., Viale S. Bartolomeo 400, 19138 La Spezia, Italy, bondaryk@saclantc.nato.int)

SACLANTCEN is sponsoring a program entitled Sound Oceanography and Living Marine Resources Program, whose objective is to demonstrate dual use technology for passive detection and localization of marine mammals, which will assist in the implementation of acoustic risk mitigation policies when using high-power acoustic sources. A sea trial was conducted in August 1999 to evaluate marine mammal detection and localization techniques using a number of advanced passive arrays. Passive, bearing-only data from marine mammal vocalizations received on a towed horizontal line array and target motion analysis are used to localize whale position. The multipath arrival structure of received signals and a ray trace model are used to determine whale range and depth over time. Data from several stations, executed using three deployed vertical line arrays in the vicinity of diving whales, and triangulation of arrivals are used to track whale location in the horizontal plane, while beamforming of the vertical array data is used to track whale depth. Evaluation of these data and techniques will be used to formulate passive procedures for determining the presence of marine mammals prior to activation of high-power acoustic sources.

8:30

**3aABa3. Transient signal processing for marine animal localization.** Brian G. Ferguson, Jane L. Cleary, and Kam W. Lo (Defence Sci. and Technol. Organisation, P.O. Box 44, Pyrmont, NSW 2009, Australia)

A technique for locating a point source in a constant velocity medium using time delay measurements from a passive stationary sensor array is applied to the localization of underwater biological transient sources. The curvature of the spherical wave front emanating from a source is sensed at

a three-element linear receiving array by estimating the inter-sensor time delays, which are then processed to extract the range and bearing of the source. The technique is used to estimate the source positions of snapping shrimp sounds recorded in Sydney Harbour (Australia) during a high-frequency sonar experiment. Plotting the source positions for a series of acoustic transient events produces a spatial distribution of the local snapping shrimp population. Passive ranging errors caused by imperfect knowledge (noncollinearity) of the sensor positions are discussed. In addition, an acoustic method is presented for estimating the *in situ* sensor positions of the array by processing the sensor data for mechanical transients produced by spatially distributed sources with known positions.

8:45

**3aABa4. Rapid estimation of dolphin whistle bearings using a sparse towed hydrophone array.** Aaron M. Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093-0205), Thomas Norris, and Jay Barlow (Natl. Oceanograph. and Atmospheric Administration, La Jolla, CA 92038)

Acoustic bearings were obtained from modulated dolphin whistles (3–20 kHz range) using a three-element towed array with a 9-m aperture. This system has performed reliably on animals whose range was known to be greater than 3 km away. It also has worked well under Beaufort five sea states. The algorithm presented here is unusual in that the signals are beamformed in the frequency domain, instead of being cross-correlated in the time domain. Beamforming on such a sparse array is feasible because the signals are broadband, allowing beams to be computed at multiple frequencies. The grating lobes, or false locations, shift their bearings with frequency, while the true source position remains fixed and can thus be identified. Frequency beamforming may complement temporal cross-correlation methods in situations when the signals are heavily contaminated with noise and multipaths, when multiple animal calls are received simultaneously, or when long time sequences need to be processed and compactly stored to memory. [Work funded by U.S. Dept. of Commerce.]

9:00

**3aABa5. Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration.** John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 3440 Market St., Ste. 400, Philadelphia, PA 19104-3325, johnsr@sas.upenn.edu)

Calling animals are located using widely distributed receivers, and the sounds from the animals are used to map the sound speed and wind fields by means of tomography. In particular, two Red-Winged Blackbirds *Agelaius phoeniceus* are correctly located within a meter using recordings from five receivers spread over a 20×30-m<sup>2</sup> region. The demonstration hinges on two new developments. First, a new algorithm for blindly estimating the impulse response of the channel is shown capable of estimating the differences in the time of first arrivals at two receivers. Since it is known that the first arrivals travel along nearly straight paths, the difference in time constrains the animal's location to a hyperboloid, and the animal is located by intersecting hyperboloids from many pairs of receivers. Second, in order to accurately find the intersection point and map the sound speed and wind fields using tomography, a nonlinear equation is



solved. The new algorithm for blindly estimating the impulse response of a channel offers a new way for locating sounds and making tomographic maps of the environment without any requirement for a model for the propagation of sound such as is needed for focalization, matched field processing, and conventional tomography.

9:15

**3aABa6. The use of neural networks to classify echolocation calls of bats.** Stephen C. Burnett and W. Mitchell Masters (Dept. of Evolution, Ecology, and Organismal Biol., The Ohio State Univ., 1735 Neil Ave., Columbus, OH 43210, burnett.33@osu.edu)

The echolocation calls of bats contain substantial information about the animal producing the call. Previous studies have used discriminant function analysis (DFA) to classify calls to individual, to calling situation, to age, and to sex. DFA requires that the number of groups in the sample be known beforehand, and thus cannot be used to determine the group structure of novel sets of data. Echolocation calls (described by a variety of frequency, time, and amplitude variables) were given as inputs to artificial neural networks. The networks attempted to classify the calls without prior knowledge of group composition. A self-organizing map (SOM) neural network was used to estimate the number of bats that produced the set of calls on the basis of the variables describing the calls. Results suggest that this approach may yield a reasonable estimate of the number of bats in the sample even when *a priori* information is lacking. This may be useful for estimating the number of bats flying in an area based on recordings of their echolocation calls. Further analyses were conducted to determine the usefulness of using neural networks for other classifications. [Work supported by NIH.]

9:30

**3aABa7. Can the use of signal curvature improve identification of bats in flight using echolocation calls alone? A test using discriminant analysis and artificial neural networks.** Stuart Parsons and Gareth Jones (School of Biological Sci., Univ. of Bristol, Woodland Rd., Bristol BS8 1UG, UK)

Echolocation calls of bats are highly flexible and have been shown to vary within species due to habitat, foraging ecology, individual identity, family affiliation, sex, and age. Calls also vary between species due to morphology, habitat preference, and foraging ecology. Such plasticity in call design has made the development of species identification systems extremely difficult. Techniques employed previously include microphones linked to event recorders, manually tuning heterodyne bat detectors, zero-crossing analysis, discriminant function analysis (DFA) of time and frequency characteristics of calls, and synergetic analysis of spectrograms. To date, all have met with limited success. Research at the University of Bristol has combined estimates of curvature with "traditional" measures of echolocation calls (duration, start frequency, end frequency, frequency with most energy) to increase the precision of species identification. Call parameters were analyzed using a backpropagation neural network and

DFA. Results indicate that a correct identification rate of 96% was possible for 14 species of British bats using the neural network. This compares with a success rate of 80% using the same data analyzed by DFA and represents the highest identification rate yet achieved.

9:45

**3aABa8. Acoustic detection and identification of singing cicadas (*Homoptera: Cicadoidea*) in Europe and in S.E. Asia.** Matija Gogala (Slovenian Museum of Natural History, Presernova 20, SI-1001 Ljubljana, Slovenia)

Acoustic methods of detection and identification of species can be effectively used in the field for singing cicadas (*Homoptera: Cicadoidea*). In many cases they emit acoustic signals loud enough and containing such frequencies that they can be detected by the unaided human ear. Nevertheless, many species emit high-frequency songs, which are not easily audible. For detection of such signals ultrasonic detectors can be successfully used. In Europe calling songs of most species differ mainly in rhythmic or amplitude patterns. In Slovenia and Croatia all existing species can be with high certainty detected, recognized and located with the use of suitable recording equipment (ultrasonic detectors included). Completely different is the situation in tropical forests of S.E. Asia, where many species emit also frequency or frequency band modulated songs. There the main problem for using acoustic approach for detection and identification of single species is not in recording technique or in recognition of species specific patterns, but in identification of cicadas and in matching the species with recorded songs. These difficulties can be to some extent overcome with recording of caged specimens, stimulated with various methods to sing under such conditions.

10:00

**3aABa9. Individuality of advertisement calls in the Australian frog *Pseudophryne bibroni*: Potential for identification of individuals.** Daniel J. Rogers (Dept. of Environ. Biol., Univ. of Adelaide, Adelaide, South Australia 5005, Australia)

*Pseudophryne bibroni* is a small terrestrial frog inhabiting southeast Australia, the males of which call from burrows, both to attract females and maintain territories. This study investigates the capacity for using advertisement calls of *P. bibroni* to census and monitor frog individuals, through analysis of individually distinct call features. Discriminant analyses revealed that, for 40 individuals, 80% of individual calls were correctly classified, with the analysis of call bouts producing even higher resolution. Given the high level of call individuality in *P. bibroni*, the potential exists for advertisement calls to be used to identify frog individuals. A system based upon these results is currently being developed, which should provide a practical, field-based method of individual frog identification based on call, allowing the long-term monitoring of frog populations without the necessity for more invasive identification techniques. [I thank David Paton for assistance with supervision, and Nicki Mitchell for assistance with collection of data. This research received financial support from the Department of Environmental Biology Research Fund.]

3a WED. AM

## Session 3aABb

## Animal Bioacoustics: General Topics

Larry L. Pater, Chair

USACERL, 2902 Farber Drive, Champaign, Illinois 61821

Chair's Introduction—10:25

## Contributed Papers

10:30

**3aABb1. The digital archiving project at the Borror Laboratory of Bioacoustics.** Douglas A. Nelson, Sandra L. L. Gaunt, C. L. Bronson, Stephen C. Burnett, Gerald E. Hough II (Dept. of Evolution, Ecology, and Organismal Biol., Ohio State Univ., Columbus, OH 43210, gaunt.2@osu.edu), and Kim Beeman (Eng. Design, Belmont, MA)

The Borror Laboratory of Bioacoustics (BLB) at The Ohio State University is a research facility with an archive of recorded animal sounds collected primarily by BLB staff and associates. The 25 000 sound recordings are scientific data that require special treatment to ensure their longevity, and the BLB is, like other sound archives, dedicated to the preservation of these recorded sounds. Traditionally, sound recordings have been archived on analog 1/4-in. magnetic tape. However, magnetic tape is degraded by time, usage, and excess temperature and humidity. Additionally, access to data on analog tapes is slow. Facing loss of access to data, especially on tapes exceeding their 50-year life expectancy, we are copying the collection to digital format [compact disk recordable (CDR)] with the aid of funding from the National Science Foundation. Because digital technology has been tested and refined over nearly two decades, and CDR media for storage of digital data now sustains a viable commercial market, archives can today endorse their use. We present the limitations of analog technology, the limitations and benefits of digital technology and CD storage, and outline the production process and safeguards developed by us for this task. [Work supported by NSF DEB-9613674.]

10:45

**3aABb2. Comparative studies of masking by harmonic complexes in several species of birds.** Micheal L. Dent, Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, mdent@psyc.umd.edu), Marjorie R. Leek (Walter Reed Army Medical Ctr., Washington, DC 20307), Melonie R. Mavilia, and Bernard Lohr (Univ. of Maryland, College Park, MD 20742)

Studies of masking by harmonic complexes have been useful in identifying some of the mechanisms that may contribute to the processing of complex sounds. In humans, for instance, harmonic complexes constructed in positive and negative Schroeder phase are differentially effective as maskers even though they have identical temporal envelopes and long-term spectra — a result that may reflect the phase response of the mammalian basilar membrane. Recent experiments on budgerigars show significant differences from the responses observed in humans [Dent *et al.*, *J. Acoust. Soc. Am.* **104**, 1811 (1998)]. Here, we extend the generality of the budgerigar masking results to three other bird species. Masked thresholds for tones embedded in Schroeder-phase maskers were measured in budgerigars, canaries, zebra finches, and fire finches using operant conditioning and the method of constant stimuli. All species produced similar masking results to those reported earlier for budgerigars: none of the birds showed greater masking by the negative than the positive-phase waveforms, and in some cases the positive waveform was the more effective masker. These results may reflect general differences in the processing of complex sounds in avian versus mammalian cochleas. [Supported by NIH R01 DC00198 and NRSA DC00046.]

11:00

**3aABb3. Masked hearing thresholds and critical bandwidths for dolphins and a white whale at 20 and 30 kHz.** Carolyn E. Schlundt (Sci. Applications Intl. Corp., Maritime Services Div., 3990 Old Town Ave., Ste. 105A, San Diego, CA 92110, schlundt@spawar.navy.mil), James J. Finneran, Donald A. Carder, and Sam H. Ridgway (SPAWARSYSCEN SAN DIEGO, San Diego, CA 92152-6266)

Critical bandwidth is an important measure of the frequency selectivity within the auditory filters of the ear. In studies of masked hearing thresholds, the bandwidth of the masking noise must be greater than the subject's critical bandwidth for the frequency of interest. There have been few direct measurements of critical bandwidths for dolphins and none exist for white whales. Underwater pure-tone hearing thresholds were estimated for three dolphins, *Tursiops truncatus*, and one white whale, *Delphinapterus leucas*, at 20 and 30 kHz. Band-limited Gaussian masking noise projected at 80 and 90 dB *re*:  $1 \mu\text{Pa}^2/\text{Hz}$  was centered around the test frequencies. Six different noise bandwidths were used in threshold estimation and were presented in a quasi-random order. Bandwidths had  $Q$  values of 0.8, 1, 1.3, 2, 4 and 10, where  $Q$  is the ratio of the center frequency to bandwidth. Thresholds were estimated using the method of free response paradigm and an up-down staircase psychophysical procedure for stimulus presentation. Results of the present study add to the existing knowledge of critical bandwidths for dolphins and provide the first data on critical bandwidths for white whales. [Work supported by the Office of Naval Research.]

11:15

**3aABb4. Blueback herring respond to brief ultrasonic stimulus sounds.** Carl R. Schilt (ASCI Corp., Fisheries Eng. Team, P.O. Box 40, North Bonneville, WA 98639, schilt@saw.net) and John M. Nestler (U.S. Army Engr. Res. and Dev. Ctr., Waterways Experiment Station, Vicksburg, MS 39180)

Captive schools of Blueback Herring (*Alosa aestivalis*) were subjected to 1-h long treatments of short 118-kHz sounds applied from a piezoelectric transducer submerged outside a vinylized canvas pen that was suspended in Richard B. Russell Reservoir on the Savannah River. In the experiments reported here, the effect on herring distribution of 3.8-ms long bursts at 13 repetitions/s (5% duty cycle) was compared to the effect of both another 118-kHz sound treatment (with a different duty cycle but the same source level) and a sound-off treatment. Each experiment compared herring distributions in the pen with the three treatments applied in a cross-over design. The experiments lasted six days each with six 1-h treatments per day. All of the several combinations of pulse duration and repetition rate contained the herring near the end of the pen farthest from the sound source but hour-long continuous wave treatments produced only a brief startle response at stimulus onset. [Work supported by U.S. Army Engineer District, Savannah of the U.S. Army Corps of Engineers.]

11:30

**3aABb5. A robotic platform for investigating the usage of acoustic flow in cf-bats.** Jose M. Carmena and John C. T. Hallam (Inst. of Percept. Action and Behaviour, Div. of Informatics, Univ. of Edinburgh, 5 Forrest Hill, EH1-2QL Edinburgh, Scotland, UK)

In this work, a biomimetic sonarhead mounted on a mobile robot is used as a platform with which to test the type of acoustic flow in cf-bats considered by Muller [R. Muller and H.-U. Schnitzler, *J. Acoust. Soc. Am.* (submitted)]. Such acoustic flow, based on changes in Doppler shift and echo amplitude, provides crude localization information which suffices for tasks like obstacle avoidance in which high accuracy is not needed. Previous results with the biomimetic sonarhead [V. A. Walker, H. Peremans, and J. C. T. Hallam, *J. Acoust. Soc. Am.* **104**, 569–579 (1998)] showed the availability of dynamic cues in the form of frequency and amplitude modulations obtained by pinnae motion. Kuc's sensorimotor model for the fm bat [R. Kuc, *J. Acoust. Soc. Am.* **94**, 1965–1978 (1994)] has been adapted to the cf bat and implemented in the robot. Appropriate experiments, composed of a prey capture task during which obstacle avoidance by means of acoustic flow is performed, have been designed to exclude trivial static cues such as the echo delay correspondent to the

closest reflector. The results of this on-going work will be included in the final abstract if accepted.

11:45

**3aABb6. On-line cough recognizer system.** Annemie Van Hirtum, Jean-Marie Aerts, Daniel Berckmans (Dept. of Agricultural and Appl. Biological Sci., K. U. Leuven, Kard. Mercierlaan 92, 3001 Heverlee, Belgium, daniel.berckmans@agr.kuleuven.ac.be), Benoit Moreaux, and Pascal Gustin (Univ. of Liege, Liege, Belgium)

Until now the intelligent, controlled, and automated use of acoustic bio-responses in cattle houses as indicators of animal condition has been limited to human perception. Although the appearance of coughing, sneezing, etc., may point to an environmental deficiency in the animal habitat or the presence of diseases among the cattle, in order to use this kind of acoustic information as a bio-response signal in practice it is necessary to develop a simple on-line algorithm. In this paper an on-line registration technique for pig coughing is presented. The application of a distance function to a spectral sound analysis allows us to distinguish coughing from other animal and environmental sounds like metal clanging, grunting, noise, etc. The sound database was collected on a group of six healthy animals.

WEDNESDAY MORNING, 3 NOVEMBER 1999

KNOX ROOM, 9:00 TO 11:00 A.M.

### Session 3aBB

## Biomedical Ultrasound/Bioresponse to Vibration: Microbubbles in Medical Ultrasound

J. Brian Fowlkes, Chair

*Department of Radiology, University of Michigan Medical Center, 200 Zina Pitcher Place, Ann Arbor, Michigan 48109*

### Contributed Papers

9:00

**3aBB1. Shell resonance responses from ultrasound contrast agents.** John Allen, Dustin Kruse, Kathy Ferrara (Univ. of California at Davis, 1 Shields Ave., Davis, CA 95616-5294), and Jeffrey Jones (Univ. of Virginia)

Ultrasound contrast agents are used to increase the scattered signal from regions containing blood in order to differentiate them from tissue. These agents are microbubbles, 1–10  $\mu\text{m}$  in diameter, encapsulated by an elastic shell 10–250 nm thick. The agents are modeled as air-filled spherical shells and the scattered acoustic field is computed from a modal series representation of the solution. Reflectivity and angular scattering characteristics are computed for various shell thickness and material parameters. In addition to the monopole response, it is demonstrated that other resonance responses exist for agents of sufficient thickness. These additional responses are attributed to Lamb waves in the shell. A large dipole peak in the 15–40 MHz range for 2.5–3.5  $\mu\text{m}$  radius agents occurs for thick shell agents with elastic properties approximating albumin. The relative reflectivity from these dipole responses is comparable to the monopole response of a gas bubble in water. Potential clinical ultrasound applications for this work, including improved tissue perfusion studies, are discussed. [Work supported by NIH CA76062 and EY11468.]

9:15

**3aBB2. Vaporization of micrometer size droplets in simulated *in vivo* environments.** Oliver D. Kripfgans (Dept. of Radiol., Univ. of Michigan, 200 Zina Pitcher Pl., Ann Arbor, MI 48109-0553, greentom@umich.edu), Douglas L. Miller, J. Brian Fowlkes, and Paul L. Carson (Univ. of Michigan, Ann Arbor, MI 48109-0553)

Droplets were generated for a range of albumin (bovine) concentrations. Resulting emulsions were examined for size distribution, number density and yield. Albumin concentrations as low as 1 mg per ml were capable to form stabilized droplets. Changes in number density and mean size were observed. Reducing the amount of albumin by a factor of 16 resulted in an increase of the number density by a factor of 40 while the mean droplet size decreased by a factor of 2.5. The measured droplet yield was approximately 30%. Droplets were vaporized using short ultrasonic tone bursts for frequencies ranging from 1.5 to 8 MHz. Two 1-cm Zerdine 81 layers of acoustic attenuation of 0.5 dB/MHz/cm were overlaid on the flow tube used. Droplets were phase transitioned in saline and in blood. Pressure thresholds for vaporization were acquired for conditions with and without Zerdine as well as saline or blood as bulk fluid. It has been seen that the threshold in saline increased by approximately 25% by introducing the Zerdine layers (1.5 MHz) and 60% for the Zerdine-blood combination. [Work supported in part by PHS Grant No. R01 HL54201 from the National Heart, Lung and Blood Institute.]

9:30

**3aBB3. Observations of contrast agents during insonation with a high-speed imaging system.** Paul Dayton, James Chomas, and Katherine Ferrara (Univ. of California, Davis, One Shields Ave., Davis, CA 95616)

With the use of a high-speed imaging system, fluctuations in microbubble radius and asymmetries leading to destruction at clinical ultrasonic frequencies can be evaluated. The extraordinary temporal resolution necessary to obtain these observations was achieved using a 10-W copper-vapor laser in conjunction with a high-speed digital camera. This combination produces an effective shutter speed of 50 MHz, which allows for imaging of the fluctuations in radius over the acoustic pressure cycle. Using this imaging technique in combination with image-analysis software, radius versus time curves of bubble expansion and contraction during the acoustic pressure cycle can be produced. It is observed that contrast agent microbubbles can expand and contract by a large fraction of their resting diameter (130%) during insonation while remaining intact at low acoustic pressures (180 kPa). At higher acoustic pressures (2.3 MPa), bubbles can expand up to 500% during the rarefactional half-cycle, after which they usually fragment into smaller bubbles. Since the echo from a bubble is determined by the radial fluctuation and resulting wall velocity, mapping such changes over time is essential if the radiated pressure waveform is to be predicted. For Biomedical Ultrasound/Bioresponse to Vibration Best Student Paper Award.

9:45

**3aBB4. Ultrasound contrast agents retain acoustic activity post phagocytosis by leukocytes.** Paul Dayton, James Chomas, Katherine Ferrara (Univ. of California, Davis, One Shields Ave., Davis, CA 95616), Jonathan Lindner, and Matthew Coggins (Univ. of Virginia, Charlottesville, VA 22908)

It has been observed that ultrasound contrast microbubbles are phagocytosed intact by activated leukocytes in regions of inflammation. In order to determine if microbubbles remain acoustically active following phagocytosis, the physical responses of phagocytosed and free microbubbles to acoustic pulses were measured optically and acoustically. The rate of microbubble destruction at each acoustic pressure is lower for phagocytosed compared to free microbubbles. Destruction of phagocytosed microbubbles at  $-1600$  kPa occasionally results in disruption of the leukocyte cell membrane. The phenomena of leukocyte destruction by insonified microbubbles was assessed with a high-speed imaging system and appeared to be due to rapid, extreme fluctuations in microbubble diameter (up to 300%) during pressure rarefaction. Analysis of the echoes returning from phagocytosed microbubbles demonstrated similarity to the nonlinear responses produced by free microbubbles. Primary radiation force is observed to displace phagocytosed microbubbles a distance of  $100 \mu\text{m}$  ( $-240$  kPa,  $10$  kHz PFR), thus providing further evidence of acoustic activity. It can be concluded that phagocytosed microbubbles maintain their susceptibility to acoustic destruction and can generate nonlinear echoes. These results indicate that contrast-enhanced ultrasound may be used to identify and assess regions of inflammation by detecting acoustic signals from microbubbles that are phagocytosed by activated leukocytes. For Biomedical Ultrasound/Bioresponse to Vibration Best Student Paper Award.

10:00

**3aBB5. Noninvasive ultrasonic interruption of contrast agents for monitoring flow in hepatic vasculature.** Richard Rhee, J. Brian Fowlkes, Jonathan Rubin, and Paul Carson (Univ. of Michigan Medical Ctr., 200 Zina Pitcher Pl., Ann Arbor, MI 48109-0553)

An experimental setup has been designed to assess differential liver perfusion due to the hepatic artery (HA) and the portal vein (PV) using continuous intravenous infusion of ultrasound contrast agents (CA) and transcutaneous ultrasound applied to target selected vessels. A perfluorocarbon-based CA was continuously infused intravenously in the ear vein of New Zealand White rabbits. A 1-in.-diam single element medi-

cal ultrasound transducer was applied transcutaneously to interrupt the CA in the abdominal aorta. Real-time feedback from the ultrasound scanner showing the clearing of CA in the target vessel indicated successful positioning of the interruption transducer. The clearing of all CA from the hepatic vasculature was accomplished by a sufficiently long pulsed insonification of the interruption transducer. The interruption transducer was turned off, producing a sharp bolus of CA. Rapid bolus rise times on the order of  $0.1$  s were observed in the abdominal aorta. Sharp demarcations in vascular response between the HA and PV to the applied CA bolus was observed. Differential arrival times of the bolus between the HA and PV on the order of  $7$  s was measured by analysis of spectral Doppler images. Through this technique, liver perfusion due to HA and PV may be discriminated.

10:15

**3aBB6. Control of cavitation in lithotripsy by waveform shaping.** Benjamin C. Davenny and Robin O. Cleveland (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02115, robinc@bu.edu)

Cavitation has been proposed as a possible mechanism in lithotripsy. Results presented by Bailey *et al.* [J. Acoust. Soc. Am. **104**, 2517–2524 (1998)] using an electrohydraulic lithotripter indicate that cavitation field created by a lithotripsy shock wave can be dramatically affected by the shape of the incident pressure waveform. We consider a piezoelectric lithotripter with 170 elements which allows for great flexibility in creating pressure waveforms. Pressure pulses are constructed from both predicted and measured waveforms. The Gilmore–Akulichev formulation is used to model the response of a cavitation bubble to the pressure waveforms. Preliminary numerical results have been obtained using single pulses and multiple pulse sequences. Results from the bubble dynamics model indicate that careful construction of the pressure pulse can lead to either enhanced or diminished collapse of cavitation bubbles. [Work supported by NIH-DK43881.]

10:30

**3aBB7. Bacterial stress response due to acoustic cavitation.** Ann M. Willman, Amy C. Vollmer (Dept. of Biol., Swarthmore College, Swarthmore, PA 19081), and E. Carr Everbach (Swarthmore College, Swarthmore, PA 19081)

*E. coli* bacteria that have been genetically engineered to emit visible light when they are stressed provide a means to study the mutual interactions of ultrasound, bubbles, and bacterial cells [Vollmer *et al.*, Appl. Environ. Microbiol. **64**, 3927–3931 (1998)]. Stresses such as protein damage, DNA damage, oxidative stress, and membrane perturbation produce differential light output in different strains of bacteria exposed to similar acoustic cavitation fields. The relative severity of each of these stressors, as determined by the luminosity per surviving cell, suggests the relative importance of each damage mechanism. Besides giving insight into cavitation mechanisms, the results suggest the parameters most useful for acoustic purification of water.

10:45

**3aBB8. Integrated diffusion and ultrasound therapy unit, for membrane transport studies.** Leonard J. Bond, Aaron A. Diaz, Kayte M. Judd (Pacific Northwest Natl. Lab., P.O. Box 999, MS K5-25, Richland, WA 99352), and Lawrence K-Y. Ng (Univ. of Colorado Health Sci. Ctr., Denver, CO 80262)

The application of ultrasound has been shown to enhance transport through a variety of membranes. This paper will present the design, calibration and initial results for a custom integrated ultrasound therapy and diffusion cell for use with a system that is a model for the blood brain barrier. The unit is designed to operate at frequencies between  $300$  kHz and  $35$  MHz, with initial trials and data taken at  $1$  MHz. Power levels are comparable with those commonly used in physical therapy ( $0.5$ – $10$



W/cm<sup>2</sup>). Unit calibration has included ultrasound field and point measurement, and both thermal and electrical measurements to calibrate power delivered. The unit is designed to hold standard snap-well cell culture supports, which are used with a cell culture system that consists of pri-

mary cultures of bovine microvessel endothelial cells. Preliminary data with a prototype-cell provide an enhancement in diffusion of up to a factor of 3. This paper will present new data, for the new cell, which has significantly improved acoustic characteristics.

WEDNESDAY MORNING, 3 NOVEMBER 1999

HARRISON ROOM, 8:30 A.M. TO 12:00 NOON

### Session 3aEAa

## Engineering Acoustics: Generation, Radiation and Scattering

Ahmet Selamet, Chair

Department of Mechanical Engineering, The Ohio State University, 206 West 18th Avenue, Columbus, Ohio 43210-1107

### Contributed Papers

8:30

**3aEAa1. The analysis, interaction, and measurement of loudspeaker far-field polar patterns.** Justin E. Baird and Perrin S. Meyer (Res. and Development Dept., Meyer Sound Labs., Berkeley, CA 94702)

This paper discusses some issues related to the measurement and representation of loudspeaker far-field polar pattern data (also known as directivity patterns). The mathematical framework of this method is based on approximating the time-harmonic far-field complex acoustic pressure field using the sum of spherically weighted acoustic point sources. One issue that is discussed in detail is a method where the far-field polar pattern is estimated by rotating a loudspeaker on a turntable and measuring the response at equally spaced intervals. It is often asserted that in order for the far-field data to be most useful, this "point of rotation" must correspond to an "effective acoustical center." However, this location is often not known *a priori*, and often a single "effective acoustical center" does not exist or varies with frequency. A simple mathematical transform is developed to move the "point of rotation" of a measured far-field polar pattern to a different location, simplifying the choice of a rotation point. Experimental evidence showing the usefulness of this transform is presented from careful measurements of a model loudspeaker rotated in an anechoic chamber.

8:45

**3aEAa2. Phase-contrast optical imaging of transient acoustic signals.** Gregory Clement,<sup>a)</sup> and Stephen Letcher (Univ. of Rhode Island, Kingston, RI 02881)

An optical imaging technique is described that utilizes a phase-contrast Fourier filtering technique to reconstruct acoustic signals in space at constant times. The system is similar to the pulsed, central-order schlieren method, but uses a Fourier filtering technique that images the actual acoustic pressure field, where the former technique reconstructs only the absolute value of the field. Signals resulting from a low-amplitude, 500-kHz, single-cycle sinusoidal pulse input to an underwater piezoceramic transducer are presented. Relative pressure fields are tomographically reconstructed in space over a series of times after the source excitation. A Hankel-transform-based algorithm exploits the cylindrical symmetry of the field for efficient reconstruction. The data are compared with synthetic signals calculated using the impulse response approach. Theory and limitations of the phase contrast system are discussed.<sup>a)</sup>Currently at Harvard Medical School, Boston, MA.

9:00

**3aEAa3. A MATLAB(R) version of CHIEF (Combined Helmholtz Integral Equation Formulation) for solving acoustic radiation and scattering problems.** Steve Forsythe (NUWC, Newport, RI 02841, forsythese@npt.nuwc.navy.mil)

The CHIEF program was developed during 1965–1988 at the Naval Ocean Systems Center. It was written in FORTRAN. The program uses the surface integral formulation of the Helmholtz equation to solve for acoustic pressure and velocity over a surface subject to surface geometry, boundary conditions, and incident field pressure on the surface. CHIEF uses a file interface to specify geometry, boundary conditions, etc., and writes its solutions for surface pressures to an output file. Thus, other programs are required for surface visualization and interpretation of results. Further, much of the original CHIEF FORTRAN code is devoted to bookkeeping imposed by machine limitations. As part of a larger modeling effort, CHIEF was ported to MATLAB. This makes the CHIEF methodology available as part of the MATLAB workbench approach to programming. Bookkeeping code that is no longer needed is discarded. Visualization tools are included as part of the new CHIEF environment. MATLAB functions are used where appropriate, combined with C++ code for speed when performing operations that are not vectorizable by MATLAB. The object-oriented features of MATLAB CHIEF are discussed as they apply to problem definition, visualization, and clarity and conciseness of programming. [Work supported by ONR 321 as Acoustic Measurement Methodology.]

9:15

**3aEAa4. Application of a new post-processing method for predicting acoustic radiation when using incompressible, viscous computational fluid dynamics.** Sheryl M. Grace and Caroline K. Curtis (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, sgrace@bu.edu)

A new framework is being developed for post-processing results from computational models of incompressible, viscous flows. The post-processor is a computational scheme focused on producing acoustic predictions using the incompressible flow quantities as inputs. The post-processing method has three distinguishing features: it is applicable to any flow field; it allows for the computation of the acoustic pressure *within* complex flow regions; and it extends the applicability of existing CFD codes in a manner easily implemented by those with CFD experience. The method is based on the classical splitting of a field into three components; vortical, acoustic, and entropy modes; and the method accounts for viscous and thermal dissipation. When simplified, the method reduces to solving Lighthill's equation computationally with the source term supplied by the incompressible CFD calculation. Here, the new method is applied

to the problem of a two-dimensional cavity subject to grazing flow. The incompressible, viscous results are calculated using FLUENT. The subsequent acoustic predictions are then compared to experimental and computational results found in the literature and are shown to agree well.

9:30

**3aEAa5. Acoustic radiation impedance of rectangular pistons on prolate spheroids.** J. E. Boisvert and A. L. Van Buren (Naval Undersea Warfare Ctr. Div. Newport, 1176 Howell St., Newport, RI 02841-1708, boisvert@lego.npt.nuwc.navy.mil)

The self- and mutual impedances for rectangular piston(s) arbitrarily located on a rigid prolate spheroidal baffle are formulated. The pistons are assumed to vibrate with uniform normal velocity and the solution is expressed in terms of a modal series representation in spheroidal eigenfunctions. The prolate spheroidal wavefunctions are obtained using computer programs that have been recently modified to provide accurate values of the wavefunctions at high frequencies. Results of the normalized self- and mutual radiation resistance and reactance are presented over a wide frequency range at various locations for several different sizes of pistons and spheroidal baffles. [Work supported by ONR Code 321 and NUWC ILIR program.]

9:45

**3aEAa6. Harmonic distortion of sound produced by a directional, airborne, parametric source array.** Anthony J. Brammer and David I. Havelock (Inst. for Microstructural Sci., Natl. Res. Council, Montreal Rd., Ottawa, ON K1A 0R6, Canada, tony.brammer@nrc.ca)

A parametric source in which sound is generated by nonlinear interaction between collinear ultrasonic beams forms an attractive alternative to a broadband, end-fire array of conventional loudspeakers. The acoustic output of an amplitude-modulated, plane-circular piston ultrasonic source has been recorded in an anechoic chamber by a microphone positioned on, and off, the axis of axial symmetry. Attention has been paid in the experiments to minimize spurious signals generated by radiation pressure (from the intense ultrasound), or by distortion within the microphone, or receiver electronics. The ultrasonic source was operated at 30 kHz, and possessed a beam width of 8 deg and bandwidth of approximately 4 kHz ( $-3$  dB). Acoustic signals were generated by pure-tone amplitude modulation of the ultrasonic drive signal at frequencies from 0.25 to 2 kHz. Parametric generation of sound was confirmed by measurements of the decay in sound pressure with axial distance from the ultrasonic source, the change in sound pressure level with the level of ultrasound, and the directionality patterns at 4 m. The generation of second harmonic distortion will be discussed as a function of modulation depth and drive frequency, within the limitations imposed by the bandwidth of the ultrasonic source.

10:00

**3aEAa7. Aerodynamic sound from a high-speed transitory wave by Rayleigh's formula.** Alan Powell (Dept. of Mech. Eng., Univ. of Houston, Houston, TX 77204-4792)

Lighthill's theory of aerodynamic sound generation and the vortex theory are not well-suited for application to high-speed periodic flows. Another method, used for very low Mach number flows, employs Helmholtz's formula, for which the very near-field (incompressible) perturbations on the enclosing boundary can be determined by any useful method, vortex or otherwise. This notion is carried over to a screeching supersonic jet, for which Schlieren photographs show evanescent waves that appear to be the source region: they extend far beyond the jet boundary into the ambient fluid. For rectangular jets, Rayleigh's formula can be reduced to a line integral through the evanescent field just exterior to the jet. The jet

instability can be represented by a convecting sinuous transitory wave in the jet boundary that grows and then decays spatially according to some envelope function. The establishment and the decay of the associated evanescent field result in radiation. Because of very effective cancellation, the radiation is very weak except at convective Mach numbers approaching unity, when a single downstream maximum occurs. The intensity is insensitive to the envelope shape and does not follow any simple power law. Plausible assumptions lead to realistic intensity estimates.

10:15-10:30 Break

10:30

**3aEAa8. Aerodynamic sound from a high-speed modulated transitory wave: Application to screech.** Alan Powell (Dept. of Mech. Eng., Univ. of Houston, Houston, TX 77204-4792)

Screech occurs in supersonic jets characterized by a periodic cell structure (cell length  $s$ ) in which the jet velocity varies periodically in the downstream direction. The transitory wave (Mach number  $M$ , wavelength  $\Lambda$ ) consequently becomes space-modulated, upsetting the otherwise high degree of cancellation and yielding a stationary phased array, resolvable into *long* and *short* propagating waves, of wavelengths  $\Lambda s/(\Lambda - s)$  and  $\Lambda s/(\Lambda + s)$ . The radiation of the latter is comparatively very weak. The former radiates with a single peak in the upstream direction, maximizing with perfect reinforcement when  $\Lambda/s = (1 + M)$ , the peak shifting away from the axis for higher Mach numbers. Plausible assumptions yield realistic screech intensity estimates (that exceed those of the downstream radiation of the *unmodulated* wave). The first harmonic of the cell structure interacts with that of the transitory wave to produce another single directional maximum, at near normal to the jet direction, again in accordance with experiment. In this treatment the simplest of assumptions have been used. In particular, the cell structure has been taken to be fixed, whereas Schlieren photographs reveal that sometimes very large displacements occur. This nonlinear refinement is now being investigated.

10:45

**3aEAa9. Computer simulation of a plane progressive wave within a sphere produced by directional sources on the surface of the sphere.** Alan H. Lubell (Lubell Labs., Inc., 21 N. Stanwood Rd., Columbus, OH 43209)

A 34-year-old paper by the author gave a prescription for a closed surface array based on the Helmholtz Integral Theorem [Alan H. Lubell, "Use of Directional Probes for Nearfield Measurements," USNJUA 1150-1152, June 1965]. The referenced paper stopped short of physically describing the required transducers. The present paper gives a generic description of a directive source in terms of orientation, spacing, and feeding of a pair of small nondirectional reciprocal transducers. Directive pattern and required time delay for a given position on the sphere is specified. Since these transducers operate in a compliance controlled mode, they are amenable to simple computer simulation. 916 total sources making 408 directional sources are carefully positioned on two concentric and closely spaced spherical surfaces and maps made of the contained field. Excellent results are obtained over a wide range of frequencies both as regards uniformity of the contained field and the null property outside the sphere. This proves the value of the array for simulating a distant receiver for the purpose of measuring the transmitting patterns of a large directional transducer in a facility of limited size.

11:00

**3aEAa10. Acceleration of integral equation based transient analysis of acoustic scattering phenomena using the plane wave time domain algorithm.** A. Arif Ergin, Balasubramaniam Shanker, and Eric Michielssen (Dept. of Elec. and Comp. Eng., Univ. of Illinois at Urbana-Champaign, 1406 W. Green St., Urbana, IL 61801, [ergin@decwa.ece.uiuc.edu](mailto:ergin@decwa.ece.uiuc.edu))

The numerical analysis of transient scattering from homogeneous bodies residing in an unbounded medium is often performed using integral equation (IE) based methods. The major advantages of these methods over differential equation based techniques are that (i) the radiation condition is implicitly imposed in the formulation, and that (ii) the number of spatial unknowns,  $N_s$ , scales as the surface area of the scatterer. However, IE techniques are notoriously costly as their computational complexity scales as  $\mathcal{O}(N_s^2)$  per time step. This complexity can be reduced considerably by using the recently introduced plane wave time domain (PWT) algorithm [A. A. Ergin *et al.*, *J. Comput. Phys.* **146**, 157–180 (1998)]. In this presentation, the PWT algorithm will be briefly reviewed and its incorporation into an IE-based transient solver for analyzing scattering from rigid bodies will be elucidated. It will be shown that the computational cost of a transient analysis using two-level and multilevel PWT enhanced solvers scales as  $\mathcal{O}(N_s^{1.5} \log N_s)$  and  $\mathcal{O}(N_s \log^2 N_s)$  per time step, respectively. Numerical examples that validate these complexity estimates and demonstrate the efficacy of the proposed methods in analyzing scattering from realistic structures will be presented.

11:15

**3aEAa11. Exact global boundary condition with reduced computational complexity for grid truncation in acoustic finite difference time domain simulations.** A. Arif Ergin, Balasubramaniam Shanker, and Eric Michielssen (Dept. of Elec. and Comp. Eng., Univ. of Illinois at Urbana-Champaign, 1406 W. Green St., Urbana, IL 61801, [ergin@decwa.ece.uiuc.edu](mailto:ergin@decwa.ece.uiuc.edu))

The numerical analysis of transient acoustic radiation by and scattering from inhomogeneous bodies is often carried out using the finite difference time domain (FDTD) method. However, FDTD grids must be truncated with absorbing boundary conditions when the structure of interest resides in an unbounded medium. Local boundary conditions have to be imposed on a convex outer boundary, thereby often inflating the problem size. This drawback is alleviated if the grid is truncated using exact boundary conditions that rely on a retarded-time boundary integral (RTBI) [L. Ting and M. J. Miksis, *J. Acoust. Soc. Am.* **80**, 1825–1827 (1986)]. Unfortunately, for a 3-D computational domain comprising of  $\mathcal{O}(N^3)$  FDTD cells, where  $N$  denotes the number of cells per unit length, the evaluation of the RTBI using classical methods requires  $\mathcal{O}(N^4)$  operations per time step, which renders the RTBI approach impractical. In this presentation, it will be shown that this computational bottleneck can be alleviated by employing

the plane wave time domain algorithm [A. A. Ergin *et al.*, *J. Comput. Phys.* **146**, 157–180 (1998)] using which the RTBI can be evaluated in  $\mathcal{O}(N^2 \log^2 N)$  operations. Numerical results that demonstrate the efficacy and reduced complexity of the method will be presented.

11:30

**3aEAa12. Photoacoustic localization of small gas leaks on cylindrical surfaces.** Serdar H. Yönak and David R. Dowling (Dept. of Mech. Eng. and Appl. Mech., Univ. of Michigan, Ann Arbor, MI 48109-2121, [serdar@engin.umich.edu](mailto:serdar@engin.umich.edu))

Detecting and localizing leaks in or on manufactured components that contain pressurized gases or liquids is a critical quality control process. Previous work has shown that gas leaks on a flat surface with a leak rate on the order of one cubic centimeter per day can be reliably detected and accurately localized using photoacoustic signals, four microphones, and incoherent broadband matched-field processing (MFP). However, leaks may not lie on flat surfaces. This work considers a case in which the surface where the leak resides is cylindrical. Photoacoustic sound is generated by rapidly scanning a carbon dioxide laser tuned to 10.6 micrometers ( $\mu\text{m}$ ) over the suspected leak area. Sulfur hexafluoride, which has strong spectroscopic absorption near 10.6  $\mu\text{m}$ , is the tracer gas. Localization results obtained using Bartlett MFP and minimum variance MFP are compared. The physics behind photoacoustic signal generation is discussed. Experimental results are presented. [Sponsored by Ford Motor Company.]

11:45

**3aEAa13. Modeling of transient scattered field.** P. Lesniewski and A. Nafalski (School of Phys. and Electron. Systems Eng., Univ. of South Australia, Mawson Lakes Blvd., Mawson Lakes SA 5098, Australia)

The spatial impulse response approach has been used to formulate an analytical model which could be helpful in predicting transient distortion by rigid scattering objects with regular/simple forms. The transient field is obtained as a linear convolution of the spatial impulse response (related to the Huygen's construction and object's geometry) with the free-field acoustic signal. The time-domain description is attractive since it allows for an easy separation of the incident and scattered field, for a strict representation of the wavefront and it may also allow for a closed—rather than integral—form of the solution for the scattered wave. Experimental results in which transients scattered by a large metal sheet were recorded at various positions near its edge in an anechoic chamber and which are in agreement with the classical harmonic integral description have been compared with approximate transient solutions obtained for a rigid half-plane scatterer. Possible closed-form solutions for the above scatterer are also discussed and related to the boundary conditions concluded from the experiment. [Acknowledgments: to Professor Harry Green for his intellectual support of this research and to Professor Richard Thorn for his long-term support and advice.]

3a WED. AM

## Session 3aEAb

## Engineering Acoustics: Engineering Acoustics Poster Session

George S. K. Wong, Chair

*Institute for National Measurement Standards, National Research Council, Montreal Road,  
Ottawa, Ontario K1A 0R6, Canada*

## Contributed Papers

To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon. To allow for extended viewing time, posters will be on display from 9:00 a.m. to 10:00 p.m.

**3aEAb1. Ultrasonic investigation of acoustic lenses.** Petre Petculescu (Ovidius Univ., 124 Mamaia Ave., Constanta, RO 8700, Romania)

In order to be efficient, acoustic lenses must meet the following requirements: the lens material must have low-ultrasound attenuation and be defect-free and the concave focusing surface must be as smooth as possible to minimize aberrations. In this paper, we analyzed two types of plane-concave lenses by using the immersion C-scan method. To acquire and process the data, we used a last-generation ultrasonic flaw detector interfaced with a computer which also drives the mechanical immersion scanning setup. The software also enabled us to perform B-scans on the lens for depth visualization and analysis. Thus we were able to realize both depth and profile maps of the lenses with considerable precision.

**3aEAb2. Neural network-based ultrasonic C-scan imaging of austenitic structures.** Petre Petculescu (Ovidius Univ., 124 Mamaia Ave., Constanta RO-8700, Romania, petculescu@ovidius.ct.ro) and Dorin Ciobanu (Inst. of Nuclear Res., Pitesti, Romania)

Results of new studies of C-scan imaging on austenitic welds, used in conjunction with artificial neural network techniques, are reported. The C-scan technique provides high-resolution images of subsurface zones which are inaccessible by conventional A-scans. Based on the anisotropy-induced perturbation of ultrasonic signals, the anisotropic regions in various steels were detected by C-scan imaging. C-scan images of both homogeneous samples of austenitic steels and heterogeneous samples of austenitic welds allowed the neural network-based "training" of the measurement and analysis system. The locations and approximate compositions of the various inhomogeneities present in the welds were determined. Image analysis was used to identify three austenitic structures (W.4541, W.6903 and HP 50) in the welds under study.

**3aEAb3. A study of loss in piezoelectric ceramic resonators.** Robert Pastore, Jr., Arthur Ballato (U.S. Army CECOM, I2WD, Bldg. 600, AMSEL-RD-IW-TI, Fort Monmouth, NJ 07703, pastore@mail1.monmouth.army.mil), and H. L. Cui (Stevens Inst. of Technol., Hoboken, NJ 07030)

Piezoelectric resonators and filters are important components in many military and commercial communication systems. Standard piezoelectric materials such as quartz have low coupling values and are quite expensive

to fabricate. Newer materials like piezoceramics have been developed that are cheaper, have higher values of coupling constant, but can be considerably more lossy than single-crystal materials. The purpose of this work is to explore loss mechanisms through the use of complex elastic coefficients in the piezoelectric constitutive relations and the differential equations for a simple one-dimensional (1-D) resonator. Results will be presented that show how the complex elastic coefficients modify the acoustic wave velocity, the coupling constant, and the resonant frequencies of the resonator. An equivalent circuit model of a resonator will also be presented with and without loss to compare with the calculated results. An experiment has been designed to better characterize loss in devices by applying an exponentially decaying sine wave. In theory this type of signal should be able to excite the complex resonance of the device. Results will be presented for thickness-excited PZT resonators, and the results will be compared to the circuit models and the 1-D calculations.

**3aEAb4. Computer modeling and analysis of a cylindrical piezoceramic transducer with interior solid filling.** Sergey M. Balabaev and Natalia F. Ivina (Far East State Tech. Fishery Univ., ul. Lugovaya, 52-B, Vladivostok, 690600 Russia)

A computer model of cylindrical piezoceramic transducer with interior solid filling based on the mixed finite element—boundary element method is examined. The work of the transducer with calculation of acoustic radiation to water is considered. The transmitting voltage responses and beam patterns of the transducer are presented. If the filling has small acoustic resistance, the transmitting voltage response of this transducer has two maxima, thus the transducer with interior solid filling is similar to free-flooded transducer. The first (low-frequency) maximum is connected with the resonance of the interior filling, the second maximum is connected with the resonance of the piezoceramic cylinder. If the filling has great acoustic resistance, the transmitting voltage response of this transducer has one maximum only. The optimum length of the transducer is determined. At the optimum length the frequency characteristic of the radiation-voltage response becomes uniform and wide band; i.e., at this length, the responses at the volume and radial resonances are approximately equal, and the response flatness (the difference between the values at the two resonances and at a minimum between them) is within 3 dB.



**3aEAb5. Intelligibility improvement of cellular phone speech in a noisy environment.** Mutsumi Saito (Fujitsu Kyushu Digital Technol. Ltd., 3-22-8, Hakata-Ekimae, Hakata-ku, Fukuoka, 812-0011 Japan, saito@qdt.ts.fujitsu.co.jp), Kimitoshi Fukudome, and Takashi Tsumura (Kyushu Inst. of Design, Fukuoka, 815-8540 Japan)

In a noisy environment cellular phone users often have difficulties in hearing far-end speech because the ambient noise disturbs the user's understanding of the conversation. To improve speech intelligibility in noise, speech enhancement is a feasible solution. In this paper the effects of two speech enhancement methods, multiband amplitude compression and formant enhancement, are reported. The multiband amplitude compression is a typical method for hearing aids that divides speech signal into some frequency bands and conducts amplitude compression independently. On the other hand, the formant enhancement is an LPC analysis-synthesis technique that provides for formant frequency shaping. The performances of these methods on the speech intelligibility were evaluated in the simulated real-world noise circumstances. Recorded noise sounds were presented from loudspeakers in a soundproof room and the speech samples were reproduced from real cellular phone handset which was fixed on the subject's head. VCV (vowel-consonant-vowel sequence) units were used as test speech samples. The results show that multiband amplitude compression makes significant improvement in speech intelligibility and does not degrade the subjective impression of speech.

**3aEAb6. Design and development of a parametric sounder for sodar.** Mukesh Chandra and B. S. Gera (Natl. Physical Lab., Dr. K. S. Krishnan Marg, New Delhi 110012, India)

Sodar is an acoustic remote sensing technique for monitoring the atmospheric boundary layer. Conventional sodar uses a parabolic dish with a transducer at its focus to transmit acoustic burst at 2 kHz. System performance depends upon the antenna directivity and transducer sensitivity which in turn are functions of the dish diameter and operational frequency. To achieve directional beam with nearly zero sidelobes, even at lower frequencies, a parametric sounder has been designed and developed. The parametric acoustic signal is realized by nonlinear interaction of two high-power ultrasonic waves of frequencies in the vicinity of 20 kHz and of acoustic pressure of 140 dB (approx.). The directional and propagation characteristics of the resulting audio signal have been studied to ensure parametric nature of the virtual source. The resulting audio beam is narrow with a semi beam-angle of less than 3 deg at 3-dB down points. The directional pattern shows the absence of side-lobes. The transmitting system is capable of sending a coded signal in any desired direction. The performance of the new and inexpensive parametric sounding system for sodar application has been evaluated and discussed in the light of comparative observation.

WEDNESDAY MORNING, 3 NOVEMBER 1999 GRANT-HARDING ROOM, 9:55 A.M. TO 12:00 NOON

### Session 3aED

## Education in Acoustics: The Use of Multimedia in Acoustics Education

P. K. Raju, Chair

*Mechanical Engineering Department, Auburn University, Ross 201, Auburn, Alabama 36849-5341*

Chair's Introduction—9:55

### Invited Papers

10:00

**3aED1. Penn State World Campus courses in noise control engineering.** Courtney B. Burroughs (Grad. Prog. in Acoust., The Penn State Univ., State College, PA 16802)

One of the first offerings in the Penn State World Campus is a series of four graduate-level courses in noise control engineering. These courses use computer-based technology to deliver the courses to students at a distance. Therefore students can take the courses from any location on a very flexible schedule, making the courses available to students who are committed to a career and/or family. Course material is sent to students on CD-ROM and interactions between students and with faculty are conducted via an e-mail server. Computer technology permits illustrative animations in the text. Assignment involves interactive animations and virtual experiments. Students collaborate on realistic noise control projects which involve measurements and analyses conducted using virtual instruments. Descriptions of the courses, the multimedia used to develop and deliver the courses, and the interactions between students and faculty are presented.

10:30

**3aED2. Animations for teaching more advanced acoustics topics.** Daniel A. Russell (Sci. and Math Dept., Kettering Univ., Flint, MI 48504, drussell@kettering.edu)

Computers and the World Wide Web are quickly replacing traditional methods of presenting animations of acoustic-wave phenomena in the classroom. A previous paper [V. W. Sparrow and D. A. Russell, *J. Acoust. Soc. Am.* **103**, 2987 (1998)] outlined how MATHEMATICA could be used to generate animations which can be converted to QUICKTIME movies or animated GIF images for insertion into HTML documents. The mathematical prowess of a program like MATHEMATICA allows for animating more complex wave phenomena while ensuring that the physics and mathematics are correct. At the same time, primitive graphical capabilities allow for visually pleasing results. This paper is an extension of previous work and will include a gallery of new animations which have been used in a senior-level acoustics course for engineering and physics majors. Animations will include some theoretical (sound fields of complex sources, particle motion in waves, and multiple-dof vibrating systems) as well as some experimental (modes of vibration of acoustic and electric guitars) animations. Some hints of MATHEMATICA tricks to generate animations will be shared, as well as methods for producing animated GIF files.

**3aED3. A multi-media courseware that applies vibration theories to real-world problems.** P. K. Raju and Chetan Sankar (Auburn Univ., AL 36849, pkraju@eng.auburn.edu)

Many engineering students lack the breadth of knowledge and skills that are fundamental to the practice of their profession. There is now a growing realization among educators of the need to put a greater emphasis on imparting higher-level cognitive skills. The learning experience must move from lecture as a dominant mode to include a significant level of active learning approaches. In order to fulfill these needs, written case studies were developed and used to bring real-world problems into engineering classrooms. New multi-media information technologies provide access to vast information sources, support safe discovery-based educational experiences, and more aggressively support peer-to-peer education. Studies report that the nonlinear systems associated with multimedia technologies are superior to linear paper-based systems in improving problem solving skills and increasing user satisfaction. Therefore, the written case studies were enhanced by adding video, audio, photographs, and competency material to create a CD-ROM so that the students are provided a rich set of materials that describes a real-world engineering problem involving control of excessive vibration in a power plant. Testing of this CD-ROM in vibration courses has shown that students applied the theories of vibration and rotordynamics in solving this real-world problem. [Work supported by NSF DUE #9752353.]

### *Contributed Papers*

11:30

**3aED4. Teaching acoustics courses on the Internet.** Paul A. Wheeler (Utah State Univ., UMC-4120, Logan, UT 84322)

This paper describes a general education course in acoustics taught over the internet. This makes the course available to colleges and universities not currently teaching courses in acoustics. The paper discusses challenges faced in offering on-line courses, such as homework, labs, and exams. This course uses a software package called WEBCT, which provides an administrative backbone for the course. The backbone includes: web navigation, exam base, threaded discussions, and tools to track students. Exam questions include formula-based questions for which each student receives a different input variable which is automatically calculated and compared to the student's answer. One difficulty in creating a web-based course is providing a hands-on lab experience. Traditional labs were converted into web activities and home activities. Web activities use the internet to supplement information found in the text. Home activities allow students to perform experiments at home which are later submitted through the web. All submissions and grading is done on the web, including homework, exams, web and home activities. The course works well for handling the large number of students commonly found in general education courses while providing access to nonconventional students taking the course through distance education.

11:45

**3aED5. Development of computer-based interactivities in hearing science.** Patrick M. Zurek and Nathaniel I. Durlach (Sensimetrics Corp., 48 Grove St., Somerville, MA 02144)

Work is underway to develop a CD-ROM collection of interactivities to supplement undergraduate and graduate instruction in hearing science. No attempt is made to provide complete coverage of hearing science in a stand-alone course. Rather, the approach emphasizes and exploits those features that can make multimedia computer-based presentation so valuable as a teaching adjunct: animation, audio output, and interactivity. With these capabilities, moving graphical images can represent physical or conceptual objects to help convey ideas and relationships that are difficult to illustrate with static images. In addition, the student is given extensive opportunity to hear the stimuli and experience the relevant psychoacoustic phenomena directly. Most important, in the current approach, unlike what can be achieved simply with prerecorded materials, students are given sufficient control over the signals, parameters, and procedures, to conduct a wide variety of experiments. To a significant degree, the materials provide each student with a personal laboratory. The CD-ROM now being developed, which includes interactivities in basic auditory anatomy, physiology, and perception, will be useful both for classroom demonstrations by an instructor and for work by individual students. [Work supported by NIDCD.]

**Session 3aMU****Musical Acoustics and Architectural Acoustics: Daniel W. Martin Tribute Session**

Douglas H. Keefe, Cochair

*Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131*

Dan Clayton, Cochair

*128A Greenacres Avenue, White Plains, New York 10606-3121*

An organ concert and demonstration is scheduled for Tuesday evening, 2 November, from 8:00 p.m. to 9:30 p.m. at the First Congregational Church in Columbus. A sign-up sheet is available at the registration desk. Please see section 36 in the Meeting Information for full details.

**Chair's Introduction—9:00*****Invited Papers*****9:05**

**3aMU1. Dan Martin and musical acoustics.** Douglas H. Keefe (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, keefe@boystown.org)

Dan Martin published research findings on musical acoustics in the *Journal of the Acoustical Society of America* in the years 1941–1983. Among the areas of inquiry were the acoustics of musical instruments including the brass wind instruments, piano, organ and electroacoustic instruments, subjective musical acoustics, and the complex interactions between the instrument, the concert hall and the listener. A retrospective examination of these publications highlights the progress made in musical acoustics as well as questions that remain the concern of current and future research.

**9:20**

**3aMU2. Dan Martin . . . architectural acoustician.** William J. Cavanaugh (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776)

Dan Martin's interest in architectural acoustics was aroused early in his academic years and remained for his entire professional life. While a graduate student and teaching assistant in the Physics Department at the University of Illinois, 1937–41, he worked closely with the pioneering acoustical consultant, Dr. Floyd R. Watson, on acoustical testing at the Purdue Music Hall, the Indianapolis Coliseum, the Indiana University Auditorium among others. During a period of eight years while working on research assignments for various divisions of RCA, Dan did part-time consultations on hundreds of applications of sound amplification equipment in motion picture theaters and auditoriums as well as applications in broadcast and recording studios. During Dan's long career at the Baldwin Piano and Organ Company, from which he retired in 1983 as Director of Engineering and Research, he consulted on the development and installation of Baldwin electronic organs in a multitude of architectural spaces and was Baldwin's premier in-house consultant on all matters acoustical. After Baldwin, Dan continued and intensified his consulting in architectural acoustics and noise control joining the National Council of Acoustical Consultants in 1983 at the urging of his many consulting colleagues. During his long and prolific consulting career, circa 1957 through 1998, Dan had over 500 projects many in worship and music performance spaces for which he had special interest. For Dan Martin, consulting in architectural acoustics was challenging and enormous fun. This paper highlights some of those joyful moments and it is hoped that any and all who have recollections of Dan's consulting experiences over the years will spontaneously share them with us at this special tribute session to a friend and colleague who has marked our Society so indelibly.

**9:50**

**3aMU3. Dan Martin as JASA editor.** Robert T. Beyer (Dept. of Phys., Brown Univ., Providence, RI 02912)

Dan Martin's interest in JASA goes back to his college days, but his major role has been in the position of patent review editor, in which role he worked faithfully for 20 years and, of course, as editor-in-chief for the past 14 years. Like Lindsay before him, he took the post just after serving as President of the Society, and presided over major changes in the journal, including a steady expansion of its size and the plunge into electronic publishing. Dan conducted these operations with skill, patience, and efficiency, and the Society's debt to him is great.

**10:20–10:30 Break**

10:30

**3aMU4. Temporal fine structure of organ sounds in churches.** Jürgen Meyer (Berglustrasse 2a, D-38116 Braunschweig, Germany, Juergen.Meyer@ptb.de)

The impressive sound of an organ is based on a typical frequency and time structure of the sound field which is created by the instrument's properties as well as by the room. Flue pipes have onset times of about 100 ms for middle pitch increasing to 1 s for the lowest pipes. Additionally, the onset time of the "full organ" may be prolonged by the building-up time of the wind pressure. Often, the definition of the tonal onset is supported by a short articulation noise. As the 60-dB decay time is extremely short, the organ particularly needs the reverberation of the room. The reverberation time of most historic churches is not only rather long but shows a typical frequency characteristic depending on the architectural style: representative examples include churches playing an important role in J. S. Bach's musical life. The long reverberation leads to a rather long onset time of the statistical sound field in the room and often, the early reflections arrive divided into two groups: higher frequencies coming from columns and (later on!) lower frequencies coming from sidewalls. This effect additionally prolongates the onset of low tones and creates the characteristic "heavy and sacral" sound of the organ.

11:00

**3aMU5. Architectural requirements of pipe organs.** Ewart A. Wetherill (Paoletti Assoc., 649 Mission St., San Francisco, CA 94105)

Dan Martin invested a lifetime in the design and building of both pipe and electronic organs and in their spatial requirements for optimum performance in places of worship. He combined the still-evolving discipline of room acoustics with a musical instrument old enough to be mentioned in the book of Genesis. As editor-in-chief of the Journal of the Acoustical Society, he actively supported education in acoustics and broader understanding of architectural responsibility in the design of organ spaces. There could be no more appropriate tribute to his work than to continue his efforts in an era when worship space design is in constant flux and when the very need for any type of organ is often in question. This paper summarizes the fundamental requirements of successful organ installation with the hope of promoting heightened awareness and understanding by architects, builders, and the people who populate building and budget committees.

11:30

**3aMU6. An examination of the acoustic characteristics of four churches.** L. Gerald Marshall (Marshall/kmk Acoustics, 59 S. Greeley Ave., Chappaqua, NY 10514)

For over 15 centuries, the axial-plan basilica church and its cruciform derivative have comprised the dominant architectural forms employed for Christian churches. The author has selected four churches with which he is intimately familiar, ranging in size from "cathedral" to "meeting-house," to illustrate acoustic attributes and differences associated with specific geometric qualities, including choir/organ placement. Personal experience with the buildings includes participation as a congregant, as a performing musician in concerts, and/or as an acoustic consultant. Measured data include RT, C50 for speech intelligibility, C80 for music clarity, and organ/choir sound distribution.

WEDNESDAY MORNING, 3 NOVEMBER 1999

MCKINLEY ROOM, 8:15 A.M. TO 12:00 NOON

### Session 3aNS

## Noise and the Federal Interagency Committee on Aviation Noise (FICAN): Preservation of Natural Quiet

George A. Luz, Chair

*U. S. Army Center for Health Promotion and Preventive Medicine, Environmental Noise Program,  
Aberdeen Proving Ground, Maryland 21010-5422*

**Chair's Introduction—8:15**

### *Invited Papers*

8:20

**3aNS1. National park service noise issues.** Wesley R. Henry, William B. Schmidt, and Rick Ernenwein (Natl. Park Service, 1849 C St., NW, MS3223, Washington, DC 20240)

The National Park Service (NPS) is an agency established, in part, to protect the resources and values associated with the properties assigned to its care. The NPS has found itself increasingly involved in responding to and dealing with human-caused noise and its impacts on the parks. This paper traces some of the more significant developments in the recent NPS past related to noise and discusses the current NPS issues related to noise.



8:40

**3aNS2. Loss of natural soundscapes within the Americas.** Bernie Krause (Wild Sanctuary, Inc., 13012 Henno Rd., Glen Ellen, CA 95442, www.wildsanctuary.com)

Species specific calls within an ecosystem evolve to adjust to the physical acoustic constraints of that ecosystem as well as the constraints imposed by the spectral content of the calls of other species occupying that ecosystem. The result is a rich texture of sound spread over both the audible (human) and inaudible spectrum with changing patterns at different times of the day. Based on 30 years of experience in recording natural soundscapes, the author documents the loss of natural soundscapes within the U.S. and other countries of the Americas. The results are demonstrated with field recordings.

9:00

**3aNS3. Guidelines for the measurements and assessment of low-level ambient noise.** Gregg G. Fleming (Acoust. Facility, Volpe Ctr., Kendall Square, Cambridge, MA 02142)

The Federal Aviation Administration's Office of Environment and Energy (FAA/AEE) with support from the Acoustics Facility at the John A. Volpe National Transportation Systems Center has developed a detailed methodology for characterizing ambient noise in low-level environments such as the National Parks. The presentation will include a general overview of the method, along with a brief historical perspective. Specific details of characterizing a study area's ambient noise will be outlined, including measurement site selection, instrumentation, and general field measurement procedures. Particular attention will be given to the data reduction and analysis methodology, including guidance on assigning actual measured ambient noise to specific locations within a study area. In addition, several examples of how this methodology has actually been utilized in a practical field measurement application will also be discussed.

9:20

**3aNS4. Measurement of the natural soundscapes in south Florida National Parks.** Micah Downing, Christopher Hobbs, and Eric Stusnick (Wyle Labs., Inc., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202)

The National Park Service is in the process of developing noise management plans for one or more of its parks in south Florida—Everglades National Park, Biscayne National Park, and Big Cypress National Preserve. A central concept is the definition of the natural (ambient) soundscapes as a resource to be managed as authorized by the NPS Organic Act of 1916 and other relevant mandates. The key to this concept is the development of a credible and defensible description of that resource. Previous studies have involved sound level monitoring with manned observations over relatively short time periods of 1–3 h in which all natural and intruding sounds were identified. This study evaluated the use of unmanned monitors to extend the measurement of the natural soundscapes to time periods of several days duration. Results from unmanned measurements will be compared to previous manned studies and methods for quantifying the natural soundscape and the effect of intrusions will be discussed. [Work supported by The National Park Service.]

9:40

**3aNS5. Challenges of modeling aircraft noise in national parks.** Kenneth J. Plotkin (Wyle Labs., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202, kplotkin@arl.wylelabs.com)

Aircraft noise in national parks has become a matter of concern in recent years. Modeling that noise for purposes of impact assessment and decision making poses challenges that are generally not present for conventional airport noise analysis. People present in parks are not residents, so conventional annual DNL analysis cannot be assumed to be appropriate. The concept of natural soundscape as a resource requires that noise be quantified on a much broader basis than for the conventional people-animals-structures perspective. Parks can also have a variety of ground cover and terrain, with terrain quite extreme in places. Despite the differences, and obvious extended capabilities required, any noise model must have a solid relation to existing standard practices. This paper reviews the issues which must be addressed in a model which is suitable for modeling of noise in national parks.

10:00–10:20 Break

10:20

**3aNS6. Using visitor responses to rank-order national park soundscapes.** Nicholas P. Miller (Harris Miller, Miller & Hanson, Inc., 30 New England Executive Park, Burlington, MA 01803)

Simultaneous collection of acoustic data and surveying of visitors at sites in national parks has permitted development of "dose-response" curves. These curves show what percent of visitors reported annoyance or interference with "natural quiet" from tour aircraft noise, as a function of the aircraft sound. Two measures of aircraft sound correlated reasonably with the visitor responses: percent of time aircraft were audible and difference between the aircraft and the nonaircraft produced equivalent level,  $L_{eq}$ . Using the relationships between these two visitor responses and these two measures of aircraft sound, a framework is developed that permits a rank ordering of park locations in terms of visitor reaction, once these two metrics have been measured at the location.

3a WED. AM

10:40

**3aNS7. Respondents' interpretations of impact measures for dose-response studies.** Robert Baumgartner (Hagler Bailly Consulting, 455 Science Dr., Madison, WI 53711)

Dose-response studies conducted at several National Park Service areas showed that respondents reported higher levels of impact for interference with the appreciation of natural quiet and sounds of nature than for the traditional annoyance measure of were you bothered or annoyed by aircraft noise during your visit to [SITE]? These differences led to questions about respondents' interpretation of the two measures and the appropriateness of each for policy decisions. To probe this issue, 21 cognitive interviews were conducted with respondents to a dose-response survey sponsored by the USAF at White Sands National Monument. The interviews probed respondents' question understanding and their interpretation of key phrases (e.g., bothered or annoyed and interference with appreciation of the sights and sounds of nature). In addition, they were asked how they selected scale points for each response and what types of events or conditions would prompt them to change their selected point. These qualitative data indicated that interference and annoyance were related concepts, but measured different types of impacts. Interference was interpreted as more event-specific and transitory, whereas annoyance was a summary evaluative term indicating a threshold level of negative reactions.

11:00

**3aNS8. Educating national park users on preserving natural soundscapes.** Rick Ernenwein, Wesley R. Henry, and William B. Schmidt (Natl. Park Service, 1849 C St., NW, MS3223, Washington, DC 20240)

The National Park Service (NPS) is the custodian of both some of the sites tied intimately to the history of powered flight and some of the quietest places in this country. Recognizing the need to both educate its personnel and the public about the sounds of nature and to properly deal with visitor opportunities like air tours that generate noise that masks the sounds of nature, the NPS has put together an educational package. This presentation will use many of those materials to show how the NPS is dealing with the protection of the sounds of nature, the natural soundscape of the parks.

11:20–12:00

#### Panel Discussion

#### Lessons Learned from Current and Past Efforts to Preserve Natural Soundscapes

WEDNESDAY MORNING, 3 NOVEMBER 1999

MADISON ROOM, 8:55 A.M. TO 12:15 P.M.

### Session 3aPA

## Physical Acoustics, Engineering Acoustics and Structural Acoustics and Vibration: Acoustic Nondestructive Evaluation: New Directions and Techniques, Part III

Philip L. Marston, Chair

*Physics Department, Washington State University, Pullman, Washington 99164*

Chair's Introduction—8:55

#### Invited Papers

9:00

**3aPA1. Hertzian contact Lamb wave sensors for NDE.** F. Levent Degertekin, Jun Pei, and B. T. Khuri-Yakub (E. L. Ginzton Lab., Stanford Univ., Stanford, CA 94305)

Lamb waves have long been used for ultrasonic testing and evaluation of platelike structures with success. Emerging applications of *in situ* defect detection and process monitoring in high-temperature, harsh environments impose significant challenges for conventional techniques. Hertzian, dry contact transducers offer an effective and economic solution to some of these challenges. In this paper, selective excitation of the lowest-order Lamb waves using Hertzian contact transducers is analyzed using the normal mode theory and the results are compared with experiments. Using silicon crystal plates as the test case, it is shown that these transducers can be used to measure the phase velocity of Lamb waves very accurately. In addition to materials characterization, several other NDE applications are also discussed. An ultrasonic temperature sensor for rapid thermal processing of semiconductor wafers and a thin film thickness sensor implemented in an aluminum sputtering station are given as process monitoring applications in high-temperature, vacuum environments. The defect detection results are presented in the context of pipe erosion corrosion monitoring and composite plate delamination detection. The dual thickness-distance inversion and tomographic imaging capabilities of the sensors are also shown on stainless steel and aluminum plate samples.

9:30

**3aPA2. Acoustic methods for classification of fluids in sealed containers for identification of hazardous chemicals.** Dipen N. Sinha and Gregory Kaduchak (Los Alamos Natl. Lab., Electron. and Electrochemical Mater. and Devices Group, MS D429, Los Alamos, NM 87545)

There are many situations where it is necessary to be able to determine information about the liquid contents inside sealed or otherwise inaccessible containers. Typically for classification of a liquid filler, one needs to determine some simple properties of the liquid contents, such as the sound speed, density, and attenuation of the liquid. This information is often sufficient to identify and characterize various subclasses of chemicals. In cases where the containers are to contain hazardous chemicals (e.g., chemical warfare agents, highly toxic chemicals, highly flammable liquids, etc.) it may be unsafe for operators to make measurements using traditional techniques which require direct contact with the contents. This talk will address several techniques which address fluid classification for situations where direct contact with the fluid is not feasible. It will concentrate on methods which couple acoustic energy into the interior container cavity (by means of direct contact or stand-off methods) to extract frequency dependent acoustic information to serve as classification clues to identify the filler.

10:00

**3aPA3. Acoustic target characterization for inertial confinement fusion.** Thomas J. Asaki and Thomas C. Hale (Los Alamos Natl. Lab., MS K764, Los Alamos, NM 87545)

Numerous hurdles mark the path leading to successful inertial fusion energy and these tasks are being addressed in a multinational effort. Much work at Los Alamos National Laboratory (LANL) has focused on addressing theoretical target design, target fabrication, and target characterization. Favored target designs incorporate millimeter-size beryllium or plastic shells filled to near-critical density with hydrogen isotopes. This fuel is then solidified at cryogenic temperatures and allowed to symmetrize through the natural process of beta layering. Implosion physics constraints demand very strict design standards on these targets in terms of layer sphericity, concentricity, and surface smoothness. Design-size targets for the National Ignition Facility (NIF) have recently been manufactured at LANL. Resonant ultrasound spectroscopy (RUS) is now being implemented as a valuable tool in many aspects of target characterization and it is especially useful for examining the interior of opaque objects. RUS has now been used to determine general aspects of target construction, fill densities and pressures, and the shape of interior cavities. Studies continue in the areas of direct observation of beta layering in opaque shells and supporting fluids properties measurements. Improved sensitivity and practical usability of ultrasound techniques have come through the implementation of optical vibration sensing.

10:30–10:45 Break

10:45

**3aPA4. Temperature dependence of diffuse field phase.** Richard L. Weaver and Oleg Lobkis (Theoret. and Appl. Mech., Univ. of Illinois, 104 S. Wright St., Urbana, IL)

We report on a new method for ultrasonic interrogation of solids. Diffuse fields are usually examined by means of the time evolution of their spectral energy density. The phase information is usually discarded as resisting analysis. The phase, while unpredictable, is, however, robust; according to theory it remains constant if source, medium, and receiver are not disturbed. Nevertheless, slow drifts of phase are observed over time scales of minutes. It is hypothesized that the phase changes are due to temperature fluctuations. A wideband 1-MHz transient source in an aluminum block was used to generate a complex waveform whose phase changes on cooling from 40 °C to room temperature were monitored. The waveform was found to undergo a virtually pure dilation, by a factor of order 0.00025/°C, a figure which is in accord with published data on the temperature dependence of ultrasonic velocity in aluminum. Due to the low dissipation in the sample, the precision with which changes are measured exceeded a part in ten million, a figure which could be improved by employing a stronger source. The possibilities that this parameter is a measure of residual stress or other NDE parameter, is explored.

### Contributed Papers

11:15

**3aPA5. Analysis of surface instability in an elastic anisotropic cone by the use of surface waves of weak discontinuity.** Marina V. Shitikova (Dept. of Struct. Mech., Voronezh State Acad. of Construction and Architecture, ul.Kirova 3-75, Voronezh 394018, Russia, MVS@vgasa.voronezh.su) and Yuriy A. Rossikhin (Voronezh State Acad. of Construction and Architecture, ul.Kirova 3-75, Voronezh 394018, Russia)

The problem of surface instability of a right circular hexagonal cone is investigated. The surface of the cone is free from stresses, but in the near-surface layer initial constant tensile or compressive stresses act in the hoop direction and in the direction of the cone's generators. Surface instability is analyzed by the use of weak nonstationary disturbances which propagate along the conic surface in the form of the nonstationary Rayleigh wave polarized in the sagittal plane and the nonstationary wave of the "whispering gallery" type polarized perpendicular to the sagittal plane. The analysis is carried out using the theory of discontinuities based on the conditions of compatibility; in so doing the velocities of surface wave propagation and their intensities have been obtained. It has been

found that the characteristics of nonstationary Rayleigh waves and surface waves of the "whispering gallery" type are very sensitive to the level of the tensile or compressive stresses in the near-surface layer of an elastic cone. These properties allow one to use these types of surface waves for the analysis of surface instability of elastic bodies and for nondestructive testing of residual stresses in the vicinity of surfaces of different bodies. [Work supported by RFBR.]

11:30

**3aPA6. Vibration measurements of radio frequency micro-scale acoustic devices.** Scot F. Morse,<sup>a)</sup> Joseph F. Vignola,<sup>b)</sup> Brian H. Houston, Andrew R. McGill (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375, morse@lpsa2.nrl.navy.mil), Jeremy Greenblatt (Israel Inst. for Biological Res., Ness-Ziona 70450, Israel), and Russell Chung (Geo-Ctrs., Fort Washington, MD 20744)

Current micromachining and IC manufacturing technology has enabled the development of a wide range of acoustic sensors and devices operating at frequencies up to several GHz. Accurate measurements of the

vibrational behavior of these devices is often complicated by their small size, intricate geometry and high operating frequencies. To address these issues a measurement system based on a fiber delivered heterodyne laser interferometer has been developed to perform spatially dense measurements of the surface displacement at high frequencies. With this system, noncontact displacement measurements can be made with sub-angstrom sensitivity for a spot several micrometers in diameter. Evaluation of the system is made with 2-D scans of the normal surface displacement of a 110-MHz delay line surface acoustic wave (SAW) chemical vapor sensor. Measured displacement levels of the uncoated quartz substrate for this device are typically less than 1 nm for acoustic wavelengths of 29  $\mu\text{m}$ . Limitations of this method are discussed, including the effect of spatial filtering brought about by the finite optical spot size. [Work sponsored by the Office of Naval Research.] <sup>a)</sup>National Research Council-NRL Research Associate. <sup>b)</sup>Also with SFA, Inc., Landover, MD 20785.

11:45

**3aPA7. Physical modeling and computer software for acoustic ecology problems.** Stanislav A. Kostarev (Lab. of Acoust. and Vib. Tunnel Assoc., 21 Sadovo-Spasskaya St., Moscow, 107217 Russia) and Sergey A. Makhortkykh (Russian Acad. of Sci., Pushchino, Moscow reg., 142292 Russia)

In the report physical modeling and the computer system for acoustic ecology problems based on it are presented. An investigation of acoustic fields and modeling of wave propagation in the case of media with a complex heterogeneous structure has been carried out. Both dissipation and nonhomogeneities were considered. In a particular case of soil, models were formulated and their computer analysis, including the problem of acoustic waves propagation, was carried out. Both bulk and shift waves were taken into account. Several types of elastic shells as the sources, merged in soil, were considered (including particular cases with realistic parameters of underground transport tunnel inner). Numerous measurements on the ground were carried out for the verification of proposed models. Computer realizations of described technology for acoustic and

geophysics fields analysis have been obtained. In particular a software package for analysis of vibration and acoustic situation on the ground surface and estimations of possible measures for vibration reduction is proposed. A new combined digital-analytic method for physical problems investigation has been used. It is based on the adaptive expansions of considered functions in an adjustable orthogonal function basis. Obtained results were applied for modeling of the vibration field from underground railways and ground surface building operations.

12:00

**3aPA8. Inverse problem of geologic parameters estimation for vibration calculations in city conditions.** Sergey A. Makhortkykh (Inst. of Mathematical Problems in Biol., Russian Acad. of Sci., Pushchino, Moscow reg., 142292 Russia and Lab. of Acoust. and Vib. Tunnel Assoc., 21 Sadovo-Spasskaya Str., Moscow, 107217 Russia) and Samuil A. Rybak (N. N. Andreev Acoust. Inst., Moscow, 117036 Russia)

In the report a new method of the dynamic and dissipative ground parameters estimation (velocities and damping factors of elastic waves) in various geological conditions for the modern city is presented. Knowledge of the specified parameters is necessary for prediction of vibration levels in the ground and buildings near underground railway lines. The approach is based on handling of a medium response on calibrated vibration action. The character of a signal makes possible a very compact representation in a classical orthogonal functions system (Hermite, Laguerre, etc.). Estimated parameters of a model are three velocities of elastic waves: longitudinal, transversal and surface waves and accordingly three factors of dissipation. The obtained nonlinear task of a system identification based on a known experimental signal is solved by means of each component expansion in a series on the chosen basis functions. For satisfactory evaluation, e.g., of three absorption factors, it is enough to take into account only the first terms of a series (signal envelope). The used approximation allows stable results to be received for much smaller experimental points number than in a case with the traditional exponential representation of a signal.

WEDNESDAY MORNING, 3 NOVEMBER 1999

GRANT-HARDING ROOM, 8:15 TO 9:45 A.M.

### Session 3aPPa

## Psychological and Physiological Acoustics: Auditory Physiology

Richard R. Fay, Chair

*Psychology Department, Loyola University, 6525 North Sheridan Road, Chicago, Illinois 60626*

### Contributed Papers

8:15

**3aPPa1. Fluid pressure distribution due to outer hair cell excitation in a 3-D dual chamber model of cochlea.** Anand A Parthasarathi (Dept. of Biomed. Eng., Univ. of Michigan, Ann Arbor, MI 48109), Karl Grosh (Univ. of Michigan, Ann Arbor, MI 48109), Tianying Ren, and Alfred L Nuttall (Oregon Hearing Res. Ctr., OHSU, Portland, OR 97201-3098)

In this study, basilar membrane (BM) velocity and the total fluid pressure distributions arising from exciting only the outer hair cells (OHCs) at a particular location in the cochlea are modeled. The cochlea is modeled using standard idealizations such as rectangular cross-section and linear inviscid fluid satisfying Laplace's equation. Two fluid chambers, one each above and below the cochlear partition (CP), are used. A simplified micromechanical model consisting of the BM, OHC, stereocilia, and tectorial membrane (TM) represents the CP. The formulation is solved using a hybrid finite element method. A localized excitation of the OHCs causes fluid pressure disturbances, which propagate toward the apex and stapes

(emissions). Both steady state and transient responses, and also the influence of boundary conditions at the apex, round window, and the oval window will be presented. Emission results will be compared to electrically evoked otoacoustic emissions for gerbils to determine the ability of the model to predict these experimental results, such as propagation velocities. [Work supported by NIH NIDCD RO1 DC00141, RO1 DC00078, and RO1 DC04084-01.]

8:30

**3aPPa2. Is the cochlea a wavelet filterbank?** Timothy A. Wilson (Dept. of Elec. Eng., The Univ. of Memphis, Memphis, TN 38152-6574)

Cochlear place and log-frequency responses are sufficiently similar that frequency-domain measurements are sometimes transformed to place-domain ones, using the cochlear map to relate cochlear place and sinusoidal frequency. Such similarity has been attributed to the cochlea's implementation of a wavelet filterbank. For such a filterbank, place responses at different frequencies are shifted versions of each other, not only for the



response magnitude, but also for its phase. According to cochlear macro-mechanical modeling, however, the cochlear partition velocity at the base leads the stapes velocity by 90 deg, regardless of frequency, which is inconsistent with the place responses at different frequencies being shifted versions of each other. In an alternative formulation, derived from the WKB approximate solution to cochlear dynamic equations, cochlear place and frequency responses are expressed in terms of a prototypical place, not frequency, response. Place responses at different frequencies are scaled shifted versions of the prototype, while frequency responses are derived from the prototype in a more complicated manner. Examples are given for several prototype place response, including a gamma-function envelope with exponential phase.

8:45

**3aPPa3. Cat auditory nerve fiber recovery functions.** Eric Javel (Dept. of Otolaryngol., Univ. of Minnesota, Minneapolis, MN 55455)

The literature contains few examples of auditory nerve fiber response recovery functions. To provide more information, hazard functions were compiled from 10–120 s samples of spontaneous and driven activity. Responses formed two groups that differed only in the amount of activity at short (<1.5 ms) inter-spike intervals. The larger group possessed hazard functions described by an equation with one time constant, i.e.,  $R(t) = 3D \cdot 1 - e^{-(S-t)/t}$ , where  $R(t)$  is the amount of response recovery at time  $t$ ,  $S$  is the starting time of the recovery process (analogous to absolute refractoriness), and  $t$  is the time constant. The other, smaller group possessed hazard functions similar to that reported by Gaumond *et al.* [J. Neurophysiol. **48**, 856–873 (1982)], in which a portion of the response recovers rapidly at the end of the absolute refractory period and the remainder recovers more slowly. This behavior was described well by a two-process equation, i.e.,  $R(t) = 3DM \cdot (1 - e^{-(S_e - t)/t_e}) + (1 - M) \cdot (1 - e^{-(S_l - t)/t_l})$  where  $S_e$  and  $S_l$  are the starting times of the “early” and “late” recovery processes,  $t_e$  and  $t_l$  are their time constants, and  $M$  is the amount of recovery attributable to the early process.  $S$ ,  $S_e$  and  $S_l$  ranged from 0.7–1.3 ms,  $t_e$  ranged from 1.2–1.5 ms,  $t$  and  $t_l$  ranged from 3–6 ms and usually decreased somewhat with level, and  $M$  ranged from 0.2–0.4.

9:00

**3aPPa4. Mechanisms of directional hearing in toadfish (*Opsanus tau*).** Richard R. Fay and Peggy L. Edds-Walton (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

Neurophysiological experiments on toadfish indicate the different strategies by which toadfish localize sources in azimuth and elevation. The saccules of the toadfish respond with great sensitivity to acoustic particle motion in the manner of inertial accelerometers. The directional response properties of otolithic afferents were studied in the horizontal and mid-sagittal planes using a 3-D shaker system. Most primary afferents have a cosine directionality response pattern. In the horizontal plane, an afferent’s best azimuthal direction is determined primarily by the orientation of the otolith organ in the head, and azimuth could be encoded as interaural response differences deriving from each organ’s gross directionality. The elevation of the axis of particle motion could be represented by a within-

organ pattern of response magnitude over the afferent array that is determined by the patterns of hair cell orientation. In humans, source elevation can be estimated from the head-related transfer function, a monaural computation based on tonotopy. In toadfish, the saccule is not organized tonotopically, but directly with respect to the elevation of most effective particle motion. Thus, sound localization in fishes could be based on mechanisms of monaural and binaural computation substantially similar to those operating in other vertebrates. [Work supported by NIH and NIDCD.]

9:15

**3aPPa5. A cortical map for tempo.** Carolyn Drake, Matthieu Adenier, Renaud Brochard (Lab. de Psych. Exp., CNRS UMR 8581, Univ. Ren Descartes, 28 rue Serpente, 75006 Paris, France), Antoine Ducorps, and Richard Ragot (CNRS-UPR 640-LENA Hpital de la Salpêtrière, 75651, France)

Tonotopy is a well-known property of the brain which consists of a linearly organized representation of the tonal frequencies on the auditory cortex. A similar organization for the rate of auditory events (tempo) is demonstrated. Subjects listened to isochronous, isotonic sequences of pure tones varying in tempo (100-, 200-, 250-, 300-, 400-, 800-, and 1600-ms IOI) and frequency (400, 1000, and 2500 Hz). Brain magnetic activity (MEG) was recorded over the whole skull with a 150-sensor array. For slow tempi (above 300-ms IOI), sources of brain magnetic activity occurring 100 ms after stimulus onset (M100) were localized in the auditory cortex of both hemispheres. Source position varied systematically as a function of tempo and frequency: (1) on a transverse line for different tempi, more laterally in both hemispheres for the slower tempi (“tempotopy”), and (2) on an anteroposterior line for different frequencies, more anterior for the higher frequencies (tonotopy). For fast tempi, the main observed MEG response occurred at 70 ms. The presence of these two distinct neural activation patterns corroborates previous psychophysical data.

9:30

**3aPPa6. Developing a physical basis for acoustic perception. Part I: Introduction to endogenous microwaves, formation of critical bands.** Philip L. Stocklin (Consulting Physicist, 439 Blue Jay Ln., Satellite Beach, FL 32937)

The physical theory of brain-generated standing microwaves was first proposed in 1979 [P. L. Stocklin and B. F. Stocklin, T.-I.-T. J. Life Sci. **9**, 29–51 (1979)] and their existence experimentally verified in the mid-1980’s [see Session BBB, A.S.A. 114th National Meeting, Fall 1987]. Generation of such waves is one effect of energy changes in neural transmembrane proteins during normal brain activity. During exposure to acoustic stimuli, the primary auditory cortex [AI] is identified as the major microwave source. Properties of the standing microwaves resulting from the anatomy and cochleotopicity of the AI result in the perception known as critical bands, as well as providing a physical basis for other perceptive phenomena. Following a synopsis of endogenous microwave properties, the physical basis for critical bands is discussed as a particularly clear example.

3a WED. AM

## Session 3aPPb

## Psychological and Physiological Acoustics: Potpourri (Poster Session)

Babette L. Eaton, Chair

*The Ohio State University, 110 Pressey Hall, 1070 Carmack Road, Columbus, Ohio 43210*

## Contributed Papers

To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon. To allow for extended viewing time, posters will be on display from 9:00 a.m. to 10:00 p.m.

**3aPPb1. Leaky integrators, Fletcher curves, and period histograms: A new link.** Bruce G. Berg (Dept. of Cognit. Sci., Univ. of California, Irvine, CA 92697, bgberg@uci.edu)

Fletcher curves, thresholds from tone-in-noise detection tasks plotted as a function of masker bandwidth, are thought to reflect characteristics of peripheral filters. This notion, however, is compromised by the finding that a roving level procedure affects thresholds only for very narrow bandwidths [Kidd *et al.*, *J. Acoust. Soc. Am.* **86**, 1310–1317 (1989)]. Alternatively, a three-stage, leaky integrator (i.e., initial, relatively broad band-pass filter, followed by half-wave rectification and low-pass filtering) yields frequency-dependent, level-invariant Fletcher curves, thus entertaining the possibility that sensitivity to envelope cues underlies detection at all masker bandwidths. The decision statistic is a measure of similarity (Mahalanobis distance) between the observed amplitude spectrum of the leaky integrator output and expected spectra with and without a signal. Regarding physiological plausibility, the model also describes the temporal response of auditory nerve fibers to complex stimuli. Horst *et al.* [*J. Acoust. Soc. Am.* **85**, 1898–1901 (1985)] used multicomponent, periodic stimuli to obtain period histograms from auditory nerve recordings. Fourier transforms of the histograms were used to assess temporal responses of fibers. For various stimulus configurations, the spectra of the period histogram and leaky integrator output are nearly identical. Theoretical implications are discussed. [Work supported by ONR.]

**3aPPb2. Ipsilateral acoustic reflex detection using wideband acoustic reflectance, admittance, and power measurements.** Douglas H. Keefe (Boys Town Natl. Res. Hospital, Omaha, NE 68131, keefe@boystown.org) and M. Patrick Feeney (Ohio Univ., Athens, OH 45701)

Progress has been achieved in measuring contralateral acoustic reflex shifts to tonal and noise activators in wideband reflectance, admittance, and power [Feeney and Keefe, *J. Speech, Language, and Hearing Research* (in press)], but no measurements have been made of the corresponding ipsilateral reflex shifts. Because the thresholds of these wideband contralateral reflex shifts were at least 8 dB lower than those obtained using standard clinical techniques, and because clinical research requires the use of both contralateral and ipsilateral testing, it is of interest to formulate a wideband technique to measure the ipsilateral reflex. Such an ipsilateral technique has been developed based on the observation that the major component of the contralateral reflex is below 1.6 kHz. Probe signals were designed with bandwidths from 0.25 kHz up to 1.6 or 2 kHz, and calibrated to measure reflectance (and related responses). The reflex activator signals consisted of tonal stimuli in the range from 2–4 kHz, with levels varied to measure reflex input–output functions. Ipsilateral

reflex shifts have been detected using this technique. Work is in progress to compare ipsilateral and contralateral reflexes measured using the same probe and activator signals. [Work supported by NIDCD.]

**3aPPb3. Gap detection for similar and dissimilar gap markers.** John H. Grose, Joseph W. Hall III, Emily Buss, and Deb Hatch (Div. Otolaryngol./Head & Neck Surg., Univ. of North Carolina, 610 Burnett-Womack, CB#7070, Chapel Hill, NC 27599-7070, jhg@med.unc.edu)

Gap detection for acoustic stimuli is most acute for iso-frequency gap markers and becomes poorer when the two markers are shifted apart in frequency. Gap detection for electrical stimuli presented on a single (iso-frequency) electrode channel in cochlear implant users is most acute when the two markers of the gap are perceptually similar and becomes poorer for perceptually dissimilar markers [M. Chatterjee, Q. Fu, and R. V. Shannon, *J. Acoust. Soc. Am.* **103**, 2515–2519 (1998)]. This raises the question of whether the decline in performance for acoustic stimuli with disparate frequencies reflects poor across-frequency temporal processing or the effects of a perceptual discontinuity between the two markers. To test this, gap detection was measured in a group of normal-hearing listeners using markers that were either perceptually similar or dissimilar. The types of markers were pure tones, am tones, and fm tones. The two markers bounding a gap were either the same or differed in type, frequency, and/or modulation rate. Results indicated that gap detection performance declined for frequency-disparate markers but did not uniformly deteriorate for perceptually dissimilar markers. [Work supported by NIH NIDCD R01-DC01507.]

**3aPPb4. Steady state response as an indicator of basilar membrane travel time.** Saber Ghiassi, Emily Buss, John Grose, and Joseph Hall III (Div. Otolaryngol./Head and Neck Surg., Univ. of North Carolina, 610 Burnett-Womack Bldg. CB#7070, Chapel Hill, NC 27599)

It has been suggested that neural firing in response to a broadband stimulus can be maximized by a frequency sweep that counteracts the frequency-specific delays of basilar membrane propagation. To test this, synchrony of neural discharge across frequency was estimated by measuring steady state responses (SSRs) elicited by harmonic complexes whose components were in either negative or positive Schroeder phase. These stimuli reduce to periodic frequency glides, ramping upward (negative Schroeder) or downward (positive Schroeder) in frequency, but having unfiltered envelopes that are identical. The data from three normal-hearing listeners support the basic notion that an upward frequency glide can lead

to improved neural synchrony. Higher levels of neural synchrony occurred at higher rates for higher center frequencies, consistent with shorter travel times as a function of frequency for higher center frequencies. The derived travel times were consistently faster than those typically predicted via passive models of basilar membrane wave propagation. Interpretation of these data require caution in light of possible effects of unbalanced evoked responses from the various frequency regions stimulated, nonlinear summation of responses across channels, or nonlinearity of frequency-dependent delays in basilar membrane propagation. [Work supported by NIH NIDCD RO1-DC01507.]

**3aPPb5. Magnetic field responses of auditory cortex to produced speech.** Srikantan Nagarajan (Scientific Learning Corp., 1995 University Ave., Ste. 400, Berkeley, CA 94704-1074, sri@scilearn.com), John Houde, and Michael Merzenich (Univ. of California, San Francisco, CA 94143-0732)

Several behavioral and brain imaging studies have demonstrated a significant interaction between speech reception and speech production. In this study, the dynamics of modulation of auditory cortical responses to speech during self-production and feedback alteration were examined. Magnetic field recordings were obtained from both hemispheres of subjects who spoke while hearing controlled versions of their feedback via earphones. These responses were compared to magnetic field recordings made while subjects listened to a tape recording of their production. The amplitude of the tape playback was adjusted to be equal to the amplitude of produced speech. Successful recordings of evoked responses to both self-produced and tape-recorded speech were obtained free of movement-related artifacts. In all subjects, a progressive differentiation in the dynamics of the responses to self-produced speech and to recorded speech was observed. In several subjects, a clear bilateral suppression of amplitude of the peak response 100 ms after stimulus onset was also observed. Feedback alteration enhanced the evoked response to produced speech. [Work supported by the McDonnell-Pew Foundation, the Coleman Fund, HRI, and Scientific Learning Corp.]

**3aPPb6. Hearing in the speech range as a key to a puzzle of loudness constancy and loudness adaptation.** Tracy Maguire, Ernest M. Weiler, Suzanne Boyce, Hongwei Dou, and Laura Kretschmer (ML #379, Hearing Lab., CSD, Univ. of Cincinnati, Cincinnati, OH 45221, Ernest.Weiler@uc.edu)

Loudness adaptation has been a controversial phenomenon, partly due to differences in subject behavior at different frequencies, and partly to conflicting results from studies using different methodology. In this study, parallel experiments were run contrasting two methodological paradigms, the Ipsilateral Comparison Paradigm (ICP) and Simple Adaptation (SA). Both used magnitude estimation to assess loudness adaptation—decline—at 60-dB ANSI. Forty students with normal hearing (ages 18–45) were divided into four groups, tested at 1, 4, 6, or 8 kHz, with ICP vs SA as the within-groups variable. Results showed significantly greater adaptation in the ICP as opposed to the SA condition at 1, 4, and 6 kHz. There was no significant difference at 8 kHz. Thus mean differences between the two methods were greatest at the frequencies most closely coinciding with the meaningful frequency range of speech, as previously hypothesized [E. M. Weiler *et al.*, *J. Acoust. Soc. Am.* **101** (1997)]. It is suggested that differences in subject behavior for speech and nonspeech frequencies may be modulated by differences in active processes of perception. These results will be discussed in the light of recent related work.

**3aPPb7. Peripheral factors in loudness adaptation with OAE effects.** Hongwei Dou, Ernest M. Weiler, Laura Kretschmer, David E. Sandman (ML #379, Hearing Lab., CSD, College of Allied Health Sci., Univ. of Cincinnati, Cincinnati, OH 45221, Ernest.weiler@uc.edu), and Thomas Goldman (Jewish Hospital, Cincinnati, OH)

Tone decay (TD), the ipsilateral comparison paradigm (ICP), and the monaural simple adaptation (SA) procedure assess loudness adaptation differently; this may all be peripheral but it has not been previously compared in one study. Transient evoked oto-acoustic emissions (TEOAE) have their source in the outer hair cells of the cochlear periphery. A significant correlation ( $r = -0.60$ ) between contralateral suppression of transient evoked OAEs and tone decay adaptation ( $r = -0.60$ ) has been reported [Collet *et al.*, *Audiology* **31**, 1–7 (1992)]. A significant correlation ( $r = -0.36$ ) between ICP adaptation and contralateral suppression of TEOAEs has been found [Ernest M. Weiler, Hongwei Dou *et al.*, *J. Acoust. Soc. Am.* **103**, 3052 (1998)]. Comparison of TD, ICP, and SA with suppression of TEOAEs used repeated measures testing of 75 students (20–35 yrs) of mixed nationalities. Significant correlations were found, but not at all values, between TD, ICP, SA, and suppression of TEOAEs at 80, 60, and 70 dB. A principle components factor analysis found factor I based on suppression, and factor II based on monaural adaptation. It was concluded that the ICP, TD, and SA are primarily peripheral in origin. TD and ICP each include variation in stimulus intensity which may play a role in triggering perception of loudness adaptation.

**3aPPb8. Perceptual segregation of competing speech sounds: The role of spatial location.** Ward Drennan, Stuart Gatehouse, and Catherine Lever (MRC Inst. of Hearing Res., Scottish Section, Glasgow Royal Infirmary G31 2ER, UK)

Culling and Summerfield [*J. Acoust. Soc. Am.* **92**, 785–797 (1995)] showed that listeners could not use ITDs (defined as differences in interaural phase) to perceptually segregate the sounds. The present studies tested a free-field analog of their conditions. Listeners' ability to identify one of two competing whispered "vowels" was determined. The "vowels" consist of narrow bands of noise, which represent the first and second formants of "ar," "ee," "er," and "oo." A target noise band pair (vowel) was presented at various angles on the listener's left and a distracter was presented on the listener's right. Listeners correctly identified the vowels in the free-field well above chance. To control for head movements, the stimuli were recorded through a head and torso simulator and presented to the listeners over headphones. Performance remained well above chance. The results show that, although a phase shift alone is not sufficient to achieve segregation, cues provided in real-world listening (ITD, ILD and onset-asynchrony) can provide sufficient information for this purpose. Further experiments systematically remove these cues evaluating their contribution to the task.

**3aPPb9. Do bilinguals perceive the tritone paradox as monolinguals do?** Magdalene H. Chalikia (Psych. Dept., Moorhead State Univ., Moorhead, MN 56563, chalikia@mhd1.moorhead.msus.edu) and Jyotsna Vaid (Texas A&M Univ., College Station, TX 77843)

When a two-tone pattern in which the tones are related by a half-octave interval (tritone) is played in any key, listeners hear it as descending or ascending, on the basis of an individual pitch class template (Deutsch, 1986). Perception of this pattern has been reported to vary with the geographical region in which the listener grew up, and it has been suggested (Deutsch, 1994; Ragozzine and Deutsch, 1994) that the characteristics of ones pitch class template may vary among speakers of different dialects and languages. This hypothesis was tested by asking how bilingualism may influence the perception of the tritone pairs. Spanish/English bilinguals were tested and their performance was compared with that of a group of English speakers. All listeners were from Texas. The results indicate that bilinguals perceive the patterns differently than the monolinguals. Envelope influences will also be discussed.

**3aPPb10. Perception of the tritone paradox by listeners in Texas: A re-examination of envelope effects.** Magdalene H. Chalikia (Psych. Dept., Moorhead State Univ., Moorhead, MN 56563, chalikia@mhd1.moorhead.msus.edu) and Jyotsna Vaid (Texas A&M Univ., College Station, TX 77843)

Deutsch (1986, 1987) first reported that when listeners are presented with pairs of octave-complex tones related by a half-octave interval (a tritone), they hear the pattern as descending or ascending, according to an individual pitch class template. Individual pitch class effects were found to be stable across tests that use tones with different spectral envelopes. Repp (1994, 1997) has found envelope effects but, in a recent study, Giangrande (1999) has not. Perception of the tritone pattern can vary with the geographical region in which the listener grew up, even though similarities have been reported between listeners in California and South Florida. These similarities suggest a canonical manner of perceiving the tritone stimuli within the U.S., probably propagated by the media (Ragozzine and Deutsch, 1994). The present study tested the hypothesis that a canonical manner of perceiving the tritone stimuli exists, by presenting tritone tones to listeners who grew up in Texas. The influence of the spectral envelope on the perception of these stimuli was also tested. The results suggest that (a) the overall pattern of responses is different from that of Californians, and (b) there are envelope effects.

**3aPPb11. Interstimulus interval as an effective measure of auditory spatial attention.** Michael A. Skelly, Richard E. Pastore, and Edward J. Crawley (Dept. of Psych., Binghamton Univ., Binghamton, NY 13903)

Previous experiments investigating auditory spatial attention have focused on the importance of the stimulus onset asynchrony (SOA) between the spatial cue and target. These experiments have reported accelerated target discrimination judgments for longer SOAs. Although this measurement provided insight into the working mechanisms of spatial attention, there typically was a confound between SOA and the interstimulus interval (ISI). The present experiment was designed to evaluate the importance of SOA, ISI, and cue duration by systematic manipulation of these variables (where SOA is composed of cue duration and ISI). Each trial began with a binaural cue that provided no information about spatial (ear) location of the monaural target, but alerted the participant to the anticipated presentation of the target. Listeners responded by indicating the target ear. The results indicate that longer SOAs are not required to accelerate target discrimination. Instead, the presence of a silent gap between the presentation of the cue and target stimulus (ISI) is far more important for accelerating performance. Furthermore, listeners seem adept at using available information in developing effective attentional strategies.

**3aPPb12. Gap detection: Effects of temporal location and stimulus duration.** Edward J. Crawley, Richard E. Pastore, and Michael Skelly (Dept. of Psych., Binghamton Univ., Binghamton, NY 13902-6000, br00437@binghamton.edu)

Three experiments evaluated a recent finding that gap detection thresholds vary nonmonotonically as a function of the temporal location of the gaps relative to the onset of the carrier noise burst [Brown and Nicholls (1997)]. Specifically, gaps presented in the temporal center of the noise burst required shorter durations to be detected relative gaps presented either near the onset or offset of the noise burst. This finding contrasts with a previous experiment that, although directly testing for such an effect, did not find an effect of temporal location [Forrest and Green (1987)]. The current experiment 1 replicates, and formally presents, the temporal location effect reported by Brown and Nicholls. Experiment 2 rules out bias, uncertainty, and task demands as possible contributing factors to the observed effect. Experiment 3 then demonstrates that the size of this effect varies with the overall duration the noise burst thus leading to a possible reconciliation of the findings of the previous studies. The current

results will be discussed in terms of their ability to reconcile a number of apparently inconsistent findings in the gap detection literature as well as their similarity to effects observed in a simultaneous masking task.

**3aPPb13. Gap discrimination and speech perception in noise.** Marc A. Fagelson (Dept. of Communicative Disord., E. Tennessee State Univ., Johnson City, TN 37614, fagelson@etsu.edu)

The relation between discrimination of silent gaps and speech-in-noise perception was measured in 20 normal-hearing listeners using speech-shaped noise as both the gap markers and the noise source for speech testing. In the gap discrimination experiment, subjects compared silent gaps marked by 60 dB SPL 250-ms noise bursts to standards of either 5, 10, 20, 50, 100, or 200 ms. The gap results were most similar to those reported by Abel [S. M. Abel, *J. Acoust. Soc. Am.* **52**, 519–524 (1972)] as  $\Delta T/T$  decreased non-monotonically with increased gap length. In a second experiment, the California Consonant Test (CCT) was administered at 50 dB HL via CD in three conditions: quiet, +10 S/N, and 0 S/N. Results from both experiments were correlated and the association between  $\Delta T/T$  and CCT scores was generally negative. Listeners who discriminated the gaps with greater acuity typically had higher speech scores. The relation was strongest for the smaller gap standards at each S/N, or when performance for any gap duration was compared to the CCT results obtained in quiet.

**3aPPb14. The effect of age on temporal integration.** Sara Elizabeth Gehr and Mitchell S. Sommers (Dept. of Psych., Washington Univ., One Brookings Dr., Campus Box 1125, St. Louis, MO 63130)

The present study was designed to assess the effects of age, independent of hearing loss, on temporal integration of sinusoidal signals. Thresholds for detecting a 1000-Hz sinusoid as a function of signal duration were measured for signal durations ranging from 20–500 ms. The signal had a rise/fall time of 10 ms. Subjects were younger (18–22) and older (over age 65) adults with normal hearing (thresholds less than 20 dB HL) for frequencies of 4 kHz and below. At the shortest signal duration (20 ms), thresholds for both groups were elevated approximately 16 dB relative to their thresholds for the 500-ms signal. Both older and younger adults also exhibited similar decreases in threshold as a function of signal duration. Overall, the findings suggest that unlike most temporally based psychoacoustic tasks, temporal integration of sinusoids remains relatively unchanged as a function of age.

**3aPPb15. Effect of intensity on sensory dissonance.** Nandini Iyer, Bret Aarden, Evelyn Hoglund, and David Huron (School of Music, 1866 College Rd., Ohio State Univ., Columbus, OH 43210)

Sensory dissonance is known to be related to the critical band [Greenwood (1961); Plomp and Levelt (1965)]. The maximum dissonance between two pure tones has been estimated to arise when the tones are separated by roughly 40% of a critical band [Greenwood (1991)]. Sensory dissonance disappears when the tones are separated by more than a critical band. Experimental work by Kameoka and Kuriyagawa (1969) has further demonstrated that dissonance judgments are affected by intensity. Since the size of critical bands are known to increase with increasing intensity [Moore and Glasberg (1987)], it follows that listeners should locate maximum dissonance at larger-frequency separations for higher-intensity tones. Similarly, the point at which dissonance disappears should involve larger frequency separation for higher intensity. The results of two experiments are reported where dissonance judgments were explicitly examined in the context of intensity-induced changes in critical bandwidth. In the first experiment, listeners adjusted the frequency of one tone away from a fixed



tone to the point of maximum dissonance. In the second experiment, listeners adjusted the tone to the point of just-not-noticeable dissonance. Results will be reported for 10 musician and 10 nonmusician listeners.

**3aPPb16. Parametric investigation of the virtual glide.** Nandini Iyer, Babette Eaton (Dept. of Speech and Hearing Sci., 1070 Carmack Rd., Ohio State Univ., Columbus, OH 43210), and Ramkumar Sridharan (College of Eng., Columbus, OH 43210)

This study examined the perceptual spectral integration effect using dynamic signals called “virtual glides” (Lublinskaya, 1996; Anantharaman *et al.*, 1997). A virtual glide was generated by modulating the amplitude of two sinusoids at fixed frequencies ( $f_1$  and  $f_2$ ) which were centered at  $f_c$ , with a frequency separation of  $\Delta f$ . Initially, the level of  $f_1$  was higher than the level of  $f_2$ . Over the duration of the signal, the level of  $f_1$  decreased linearly while the level of  $f_2$  increased at the same rate. This produced the percept of a frequency glide; hence the name “virtual glide.” The subjects’ task was to match the rate of change of a real glide to that of a virtual glide using Jesteadt’s (1980) double staircase matching procedure. The  $f_c$ ,  $\Delta f$ , and duration of the stimuli remained constant within a block. The  $f_c$  of the virtual glide was varied (500, 1000, 2000 Hz), as was  $\Delta f$  (2, 4, 5, 6, 8 ERB). Further, the stimuli were presented at three different durations (125, 250, 500 ms). The slope of the linear glide was varied adaptively to obtain the 50% point of subjective equality (PSE). We investigated the range of  $\Delta f$  and signal duration over which listeners could perform this task.

**3aPPb17. Just discriminable change of velocity of a simulated moving sound source.** Sarah M. Hassett and Lawrence L. Feth (Dept. of Speech and Hearing Sci., Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210)

The Doppler effect refers to the frequency change observed for a moving sound source. The listener will observe a higher than emitted frequency in front of the moving source and a lower frequency behind it. Although the existence of the Doppler effect has been long documented, relatively little experimental work has investigated the human perceptual response to frequency and intensity changes due to moving sound sources. This experiment determined the just discriminable change of velocity of a computer simulated moving sound source. Subjects were tested at 500, 1000, and 4000 Hz for reference velocities of 1, 2, 4, and 8 m/s. Signal duration was 500 ms. A 2-Q,2-AFC adaptive procedure determined the listener’s velocity difference limen (DL). At 500 Hz, the smallest DL was 10.46 m/s at reference velocity 8 m/s. The DL increases as the reference velocity decreases. This pattern is evident at all tested source frequencies. Weber fractions were calculated for both velocity DLs and frequency deviation. The results indicate that both velocity DL and frequency deviation increase as reference decreases, signifying that Weber’s law does not hold for source velocity and frequency. [Work supported by a College of Arts and Sciences Honors Scholarship and a grant from the College of Social and Behavioral Sciences.]

**3aPPb18. A comparison of two auditory models using monaural, Doppler-like signals.** Mark A. Ericson (Air Force Res. Lab., WPAFB, OH 45433-7901) and Lawrence L. Feth (The Ohio State Univ., Columbus, OH 43210)

Detection threshold predictions of an excitation pattern and an intensity-weighted average of instantaneous frequency (IWAIF) model were compared with empirical data. Monaural changes in intensity and frequency of a moving sound source were simulated over headphones. A pair of 400-ms duration signals, separated by a 100-ms inter-stimulus interval, were utilized in an adaptive two-alternative, forced-choice procedure. Detection thresholds of velocity induced frequency and distance induced intensity changes were measured on seven listeners. Sound source motion was simulated along linear trajectories with three minimal dis-

tances of 1, 10 and 100 m at the point of closest passing. Two sets of stimuli were chosen to test the models. The first set consisted of pure tones at frequencies of 200, 1000 and 5000 Hz. The second set contained three-tone, harmonic complexes with a fundamental frequency of 1000 Hz and harmonic frequency separations of 100, 200 and 400 Hz. With the pure-tone stimuli, the excitation pattern model underestimated the detection threshold at 5000 Hz. With the harmonic stimuli, the IWAIF correctly predicted slightly degraded thresholds with increasing frequency separation of the harmonics, whereas the excitation pattern model predicted a near constant detection threshold for all harmonic frequency separations.

**3aPPb19. Testing time, occlusion effects, and required masking for insert earphones versus supra-aural headphones.** Donald Hayes (Comm. Sci. and Disord., Univ. of Cincinnati, ML 379, Cincinnati, OH 45221)

In a recent survey of audiologists [Martin *et al.*, *J. Am. Acad. Aud.* **9**, 95–100 (1998)] the authors reported that only 24% of respondents used insert earphones in any capacity. Surprisingly few clinicians use insert earphones even though they have several desirable qualities relative to supra-aural headphones. Insert earphones are reputed to eliminate ear canal collapse, reduce occlusion, and increase interaural attenuation. The purpose of this study was to determine the extent to which insert earphones might increase the speed and ease with which hearing tests are obtained. Audiometric assessments were obtained on 91 individuals in the course of their standard clinical investigations. ER-3A insert earphones were used to assess 46 of them and the other 45 were tested using TDH-39 supra-aural headphones. The audiometric tests of the two groups were compared on the basis of: testing time, earphone-related occlusion effects, necessity for masking the nontest ear, collapsed ear canals, and occluded or collapsed insert earphones. The insert earphones were found to be superior in every category. It is strongly recommended that ER-3A insert earphones should become the earphones of choice for audiometric assessment.

**3aPPb20. Loudness functions of older people for pure tones and broadband noises.** Kenji Kurakata and Yasuo Kuchinomachi (Nat. Inst. of Bioscience and Human-Tech., 1-1 Higashi, Tsukuba, Ibaraki, 305-8566 Japan, kurakata@nibh.go.jp)

Loudness functions for pure tones and low-pass filtered noises were obtained by the magnitude estimation to develop a method for evaluating loudness of older people. The subjects were young adults in their twenties with normal hearing and old adults of 60 and over with age-related hearing loss. The results of the measurements are as follows: (1) There is no significant difference in the slope of loudness functions for pure tones between both age groups, suggesting most of the older subjects do not have large loudness recruitment. (2) The loudness of filtered noises judged by the older subjects decreases as the cutoff frequency of the noises increases. Considering the audiograms of the subjects, the declined ability of hearing at higher frequencies seems to be the cause of the decrease of loudness. (3) The decrease of loudness may be corrected by taking into account the amount of hearing loss of older people because they have no significant loudness recruitment.

**3aPPb21. Relative weights for three different psychophysical tasks.** Matthew R. Matiasek and Virginia M. Richards (Dept. of Psych., 3815 Walnut St., Univ. of Pennsylvania, Philadelphia, PA 19104, matiasek@cattell.psych.upenn.edu)

Relative weights for two profile analysis and one detection task were estimated. For the profile analysis “bump” task the standard was the sum of  $N$  equal-amplitude sinusoidal components ranging in frequency from 200 to 5000 Hz. The signal for this task was an increment in level to the 1000-Hz component of the standard. The number of components  $N$  ranged

from 4 to 50. For the profile analysis “down-up” task the standard was composed of  $N$  equal amplitude tones ranging in frequency from 200 to 5000 Hz. The signal to be detected led to a spectrum that varied down-up . . . down-up. The detection task used the same standard as the profile analysis down-up task, but the signal to be detected was an added 1000-Hz tone. Comparisons of the relative weights revealed that profile analysis bump and the detection of a tone added to a notched masker rely on similar processing strategies. By contrast it was apparent that auditory processing in the profile analysis down-up task depended on a different strategy than the other two tasks. Finally, rough estimates of the auditory filter bandwidths suggest invariance with regard to task. [Work supported by NIH.]

**3aPPb22. Frequency-discrimination performance using place-domain techniques.** Rachod Thongprasirt and Timothy A. Wilson (Dept. of Elec. Eng., The Univ. of Memphis, Memphis, TN 38152-6574)

A classic analysis [W. M. Siebert, Proc. IEEE **58**, 723–730 (1970)] of performance in frequency discrimination tasks used a Poisson-process model for auditory neuron firings in a simple model of the auditory periphery to infer that frequency-discrimination performance can be explained by place mechanisms (or, alternatively, that periodicity information, if used, is used inefficiently). The analysis consisted of determining the variance of the maximum-likelihood estimate of sinusoidal frequency for a single neuron, using the firings over the duration of the stimulus, then averaging those estimates over the independent neurons. An alternative approach is to consider the pattern of auditory nerve firings in the place-time plane first as a function of place, then averaging over time. To that end, the frequency-discrimination performance of a computational model of the auditory periphery was examined by determining maximum-

likelihood frequency estimates from the spatial pattern of firings at an instant, then averaging over the duration of the stimulus. Those results, and likelihood-ratio ones, are compared to the analytical results mentioned above, to computational implementation of the Siebert approach, and to psychophysical observations.

**3aPPb23. A new system of environmental noise identification and subjective evaluation.** Hiroyuki Sakai, Shin-ichi Sato, and Yoichi Ando (Grad. School of Sci. and Technol., Kobe Univ., Rokkodai, Nada, Kobe, 657-8501 Japan)

A new system of environmental noise identification including aircraft noise or traffic noise and its subjective evaluation based on the model of the human auditory-brain system [Y. Ando, *Architectural Acoustics—Blending sound sources, sound fields, and listeners* (AIP Press/ Springer-Verlag, New York, 1998)] is proposed. In order to discuss the subjective effects of environmental noise, following factors are calculated by use of the autocorrelation function (ACF) and interaural crosscorrelation function (IACF) of binaural signals; From the ACF analysis: (1) energy represented at the origin of the delay,  $\Phi(0)$ ; (2) effective duration of the envelope of the normalized ACF,  $\tau_e$ ; (3) the delay time of the first peak,  $\tau_1$ ; and (4) its amplitude,  $\phi_1$  are extracted. From the IACF analysis: (5) IACC; (6) interaural delay time at which the IACC is defined,  $\tau_{IACC}$ ; and (7) width of the IACF at the  $\tau_{IACC}$ ,  $W_{IACC}$  are extracted. For example, loudness is related with  $\Phi(0)$  and  $\tau_e$ , and directional information may be obtained from  $\tau_{IACC}$ . Especially, a noise source may be identified by the all ACF factors are timbre. [Work was partially supported by the Research and Development Applying Advanced Computational Science and Technology, Japan, Science and Technology Corporation (ACT-JST), 1999.]

WEDNESDAY MORNING, 3 NOVEMBER 1999

MORROW ROOM, 9:30 TO 11:30 A.M.

### Session 3aSA

## Structural Acoustics and Vibration: Vibration Control

Thomas J. Royston, Chair

*Department of Mechanical Engineering, University of Illinois at Chicago, Chicago, Illinois 60607*

### Contributed Papers

9:30

**3aSA1. The effect of internal point masses on the radiation of a ribbed cylindrical shell.** Martin H. Marcus and Brian H. Houston (Naval Res. Lab., 4555 Overlook Ave., S.W., Washington, DC 20375, marcus@astro.nrl.navy.mil)

The radiation associated with complexity is examined by use of finite element–infinite element models of framed fluid loaded shells for frequencies up to  $ka=10$ . The vibro-acoustic response of the shell with no internal complexity is compared to an identical structure with point masses attached to the frames. The simple ribbed shell is seen to have pass and stop band structure with radiation physics that is dominated by supersonic components. The addition of point masses to the structure gives rise to major changes in the response that include a coupling of radiating and otherwise nonradiating circumferential orders, resulting in a significant increase in radiation (about 10 dB) over a broad frequency range. These results are in general agreement with nearfield acoustical holography (NAH) experimental results. [Work supported by ONR.]

9:45

**3aSA2. An energy finite element optimization process for reducing high-frequency vibration in structures.** George A. Borlase and Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109-2145)

The energy finite element analysis (EFEA) offers an attractive alternative to the established statistical energy analysis (SEA) for simulating high-frequency structural vibration [N. Vlahopoulos *et al.*, “Numerical Implementation, Validation, and Marine Applications of an Energy Finite Element Formulation,” accepted by the Journal of Ship Research]. One of the main benefits of utilizing EFEA in a simulation based design process is the capability to assign local damping properties on selected sections of a component. In this work, an optimization routine is integrated with an EFEA solver in order to optimize the distribution of damping over a structure. The formulation of the optimization process, the definition of the design variables, the constraints, and the objective function will be discussed. The validation and an application case will be presented. [Research supported by Michigan Seagrant and the U.S. Coast Guard.]

10:00

**3aSA3. Numerical and experimental investigations on the isolation of structure-borne sound using rubber mounts.** Margareta I. Bittner (Inst. of Machine Elements and Machine Acoust., Tech. Univ. of Darmstadt, Magdalenenstr. 4, D-64289 Darmstadt, Germany, bittner@memak.tu-darmstadt.de)

Rubber mounts are frequently used for the isolation of structure-borne vibrations in the low-frequency range. However, not much research has been done in the acoustic frequency range. In this essay the sound isolation behavior of rubber mounts is investigated by means of the finite element method. The static preload is described by a nonlinear elastic material model. For dealing with the dynamic behavior, a linear viscoelastic model with frequency-dependent parameters is used. The influence of different material parameters on the transmissibility and on the dynamic stiffness of different resilient elements is examined. The calculations are compared with experimental data. For this a measurement apparatus has been set up based on ISO 10 846. Further on, the interaction between resilient elements and a structure supported on them is investigated. The supported structure is chosen to be a plate made of steel. The plate is not regarded to be rigid. The influence of different mounts on the structure-borne sound of the plate is examined. Additionally, it is researched how the sound isolation changes when eigenfrequencies of the supported plate and of the elastic mounts are approximately the same.

10:15

**3aSA4. Modeling and compensation of hysteresis in piezoceramic actuators for vibration control.** Soon-Hong Lee, Thomas J. Royston (Dept. of Mech. Eng., Univ. of Illinois at Chicago, Chicago, IL 60607, troyston@uic.edu), and Gary Friedman (Univ. of Illinois at Chicago, Chicago, IL 60607)

It has been established that the Maxwell resistive capacitor (MRC) hysteresis model accurately represents hysteretic behavior in the dielectric domain of PZT-based piezoceramics. Methods for MRC and inverse MRC online model identification have been developed along with an MRC-based framework for calculating continuous hysteretic energy loss for arbitrary loading histories. Building on these developments, an adaptive feedforward compensation technique for attenuating the effect of hysteresis in vibration control applications is proposed and evaluated experimentally via studies on a 1–3 piezoceramic composite. Also, indirect electrical excitation is compared to direct voltage drive of PZT devices for reduced hysteretic behavior in vibration control applications. [Work supported by the Office of Naval Research.]

10:30–10:45 Break

10:45

**3aSA5. Concurrent design concepts for adaptive structures.** G. Clark Smith and Robert L. Clark, Jr. (Dept. of MEMS, Duke Univ., Durham, NC 27708-0302, gcs@duke.edu)

For an adaptive structure there are two critical design components: temporal compensation and spatial compensation. The influence of the controller on the sensor signals characterizes the temporal compensation of the adaptive structure and the dynamic coupling of the actuator and sensor through the structure characterizes the spatial compensation. Traditionally, techniques for adaptive structure design have concentrated on one component and considered the other fixed. This work proposes a

method of concurrent design for both the temporal and spatial compensation of an adaptive structure. Concurrent design concepts are realized by simultaneously executing controller optimization and spatial optimization routines. The structural acoustic control performance and robustness of an adaptive plate structure is improved significantly with concurrent design. The results show the advantages of concurrent design concepts over traditional methods and the effectiveness of the proposed approach.

11:00

**3aSA6. Analysis of periodically varying gear systems with Coulomb friction using Floquet theory.** Manish Vaishya (The Ohio State Univ., 206 W. 18th Ave., Columbus, OH 43210-1107) and Rajendra Singh (The Ohio State Univ., Columbus, OH 43210)

Realistic gear pairs may undergo significant time-varying mesh stiffness and damping changes during one mesh cycle. The instantaneous number of teeth in contact governs the load distribution and sliding resistance acting on individual teeth. The geared system responds to a combined excitation of transmission error due to profile deviations, periodic variations in system parameters, and time-varying sliding friction. Such complexities have led prior researchers to adopt numerical approaches or approximate analyses, with many modeling simplifications. This article presents a new analytical model of a gear mesh with spatially varying mesh stiffness and viscous damping parameters. Unlike previous models, the excitation consists of two separate terms, namely the unloaded transmission error and the periodically varying Coulomb friction force. Hill's equation with nonproportional damping is employed to determine the forced system response. Exact integrals are calculated for the state transition matrix, instead of using Fourier series expansion, thus eliminating the mode truncation errors. Floquet theory provides an efficient analytical tool for such time-varying systems, especially for the dynamic response over long time intervals.

11:15

**3aSA7. Theoretical treatment and design of vibration damping constructions in the upper path structure for underground railways.** Samuil A. Rybak (N. N. Andreev Acoust. Inst., Moscow, 117036 Russia and Lab. of Acoust. and Vib. Tunnel Assoc., 21 Sadovo-Spasskaya Str., Moscow, 107217 Russia) and Stanislav A. Kostarev (Lab. of Acoust. and Vib. Tunnel Assoc., Moscow, 107217 Russia)

Perspective vibration damping devices, applicable for underground conditions, are considered. The physical model of elastic damping designs for the upper structure of a railway is formulated and on its base the numerical analysis of their effectiveness for underground railways conditions is carried out. The possible realizations of a design as continuous elastic mats and separate elements (shock absorbers) are considered. In the second case, the arrangement of the damping elements under a ballast stratum and under the carrying elements of a rail path (cross ties, concrete blocks, etc.) is possible. General conclusions are following: the most critical frequency band for underground railways' practice is an octave with mean geometric frequency 31.5 Hz; for correct choice of the vibration isolating measure a simulation of multi-unit oscillation system is needed; usually a two-unit system is more efficient than one-unit both for low and high frequencies; the most efficient mechanism for vibration reducing in 31–63-Hz frequency bands is elastic coefficient lowering and mass enhancing of a unit with maximal vibration levels; damping coefficient increase leads to vibration level reduction for the resonant frequency and weak influence on the amplitude far from the resonant frequency.

**Session 3aSC****Speech Communication and Signal Processing in Acoustics: Bionic Ears: Influence on Perception and Production**

Rosalie M. Uchanski, Chair

*Central Institute for the Deaf, 818 South Euclid Avenue, St. Louis, Missouri 63110***Invited Papers****8:15****3aSC1. Introduction to cochlear implants.** Emily Tobey (Callier Ctr., 1966 Inwood Rd., Dallas, TX 75235)

This brief overview will introduce cochlear implants and describe their basic features. The invited presentations will focus on performance outcome measures of multichannel cochlear implants, primarily in children.

**8:25****3aSC2. Individual differences in effectiveness of cochlear implants.** David B. Pisoni (DeVault Otologic Res. Lab., Indiana Univ. School of Medicine, Indianapolis, IN 46202)

The efficacy of cochlear implants has been firmly established. However, the effectiveness of cochlear implants varies widely and is influenced by demographic and experiential factors. First, there are enormous individual differences in both adults and children. Some patients show large changes in speech perception whereas others show only modest gains. Second, age of implantation and length of deafness affect outcome measures. Younger children do better than older children and children who have been deaf for shorter periods of time do better than children who have been deaf for longer periods of time. Third, communication mode affects outcome measures. Children from "oral-only" environments do much better than children who use "total communication." Fourth, there are no preimplant predictors of outcome performance. The underlying abilities and skills "emerge" after implantation. Finally, there are no differences among devices or processing strategies. The pattern of results suggests that processes such as perception, attention, learning, and memory may play an important role in explaining the enormous individual differences. Investigations of the content and flow of information in the central nervous system and interactions between sensory input and stored knowledge may provide new insights into individual differences. [Work supported by NIH-NIDCD Grants DC00012 and DC00111.]

**8:55****3aSC3. Psychophysical and cognitive limitations to speech perception by cochlear implant users.** Mario A. Svirsky, Adam R. Kaiser, Ted A. Meyer, Ashesh Shah, and Peter M. Simmons (Dept. of Otolaryngol.—HNS, Indiana Univ. School of Medicine, msvirsky@iupui.edu)

Deaf listeners with cochlear implants (CIs) achieve significant levels of speech recognition, but their performance range remains wide. Unfortunately, our understanding of the speech perception mechanisms employed by CI users is still incomplete. In particular, we do not know the exact combination of acoustic cues that are employed by CI users to understand speech, nor do we understand how sensory information is represented and combined, and how that information is used to perform speech identification. We have attempted to address this issue by developing mathematical models (Multidimensional Phoneme Identification or MPI models) that aim to predict phoneme identification for individual cochlear implant users based on their discrimination along specified acoustic dimensions. Mathematically, the MPI model is a multidimensional extension of the Durlach-Braida model of loudness perception. The MPI model can explain most of the vowel pairs or consonant pairs that should be more frequently confused by groups of CI users. In this presentation we will discuss individual data suggesting that speech perception by CI users may be limited by two kinds of factors: psychophysical (i.e., limited jnd's along relevant acoustic dimensions) and cognitive (related to imperfect integration of different acoustic cues). [Work supported by NIDCD (R01-DC03937), NOHR and DRF.]

**9:25****3aSC4. Issues relevant to the evaluation of speech intelligibility with cochlear implant recipients.** John W. Hawks (School of Speech Pathol. and Audiol., Kent State Univ., Kent, OH 44242) and Marios S. Fourakis (Ohio State Univ., Columbus, OH 43210)

In a series of experiments aimed at identifying factors that influence speech perception by cochlear implant (CI) users, we are currently investigating the reception and utilization of relevant perceptual information available in monosyllables and steady-state vowels. Normal-hearing subjects and adults fitted with Clarion and Nucleus cochlear implant (CI) devices provided identification responses to both natural and synthetic speech. The first experiment examines intelligibility differences of a standard word identification test (NU-6) recorded in three different contexts by two talkers (male/female): (1) individual words spoken in typical citation format, and the same words excised from a narrative read at (2) a typical rate and (3) a rapid rate. The second experiment demonstrates the potential diagnostic value afforded from mapping the acoustic vowel space using steady-state synthetic tokens. The last experi-



ment explores the effects of harmonic richness and first formant ( $F_1$ ) emphasis on synthetic vowel identification by creating tokens varying only in  $F_0$ , such that a harmonic of  $F_0$  would coincide with the  $F_1$  center frequency in (1) a male (harmonically rich)  $F_0$  and (2) a female (harmonically sparse)  $F_0$  range, and (3) an intermediate  $F_0$  which was harmonically noncoincident with  $F_1$ . [Work supported by OSU Office of Research.]

9:55

**3aSC5. Factors contributing to speech perception in children implanted before age 5.** Ann Geers (Central Inst. for the Deaf, 818 So. Euclid Ave., St. Louis, MO 63110)

This report describes results obtained from 92 8- and 9-year-old children who received a Nucleus-22 implant before age 5. Educational and rehabilitation history was obtained from questionnaires completed by parents, cochlear implant centers, and therapists. Speech perception measures included tests of phoneme discrimination (VIDSPAC), closed set word recognition (ESP, WIPI), open set word (LNT) and sentence (BKB) recognition, and visual enhancement (CHIVE). A multivariate analysis was conducted to predict results on speech perception measures from educational variables, while controlling for effects of IQ, socioeconomic status, age at onset of deafness, age at implant, and type of speech processor and map characteristics such as number of active electrodes, dynamic range, and growth of loudness. Results indicate about half of the variance in auditory speech perception is accounted for by factors related to characteristics of the implant itself. However, significant additional variance attributable to rehabilitation factors is observed. Children educated primarily in oral settings showed a significant advantage in speech perception. Amount and type of intervention following implication has a significant impact. [Work supported by NIH.]

10:25–10:40 Break

10:40

**3aSC6. Speech production factors in children using cochlear implants.** Emily Tobey (Callier Ctr., 1966 Inwood Rd., Dallas, TX 75235)

This presentation will summarize speech intelligibility, sound development, and durational characteristics of 92 children who use multichannel cochlear implants. All children had at least 5 years experience with the device and whose chronological age was 8 or 9 years. Half of the children used auditory-oral modes of communication and the other half used total communication. Speech intelligibility was assessed using 36 sentences of 3, 5, and 7 syllables in length [McGarr (1981)]. Samples were recorded, digitized, edited, randomized, and played to listeners (72 listeners per child). Samples also were phonetically transcribed by three independent raters. In addition, recordings were made of CV syllables in isolation or as the final syllable of a two-syllable word. Data indicated speech intelligibility was significantly higher in the children with auditory-oral modes of communication ( $p < 0.01$ ). Sound development also was significantly higher in children with auditory-oral modes of communication ( $p < 0.01$ ); however, the pattern of sound development was similar in the two groups. Children with less intelligible speech also experienced difficulty in producing CV syllables in final syllables as indicated by their acoustic measures. Syllables displayed inappropriate durations and formant values approximating schwas. [Work supported by NIH.]

11:10

**3aSC7. Relations between acoustic measures and the intelligibility of speech for children using cochlear implants.** Rosalie Uchanski, Gina Torretta, Ann Geers (Central Inst. for the Deaf, 818 S. Euclid, St. Louis, MO 63110, ruchanski@cid.wustl.edu), and Emily Tobey (Callier Ctr., Dallas, TX 75235)

Speech was recorded from approximately ninety 8- and 9-year-old children who had used a cochlear implant for 4–6 years. Each child produced two sets of sentences for two separate purposes. First, 36 McGarr (McGarr, JSJR, 1983) sentences were recorded for subsequent intelligibility testing by listeners with normal hearing. Second, 11 newly constructed sentences were recorded for acoustic analysis. Acoustic measurements included: (1)  $VOT_i - VOT_d$ : average voice-onset time for word-initial /t/ minus that for /d/; (2)  $F_{2i} - F_{2d}$ : average second formant frequency for /i/ minus that for /ɔ/; (3) vowel/word: average ratio of vowel-to-word duration; and (4) differences in spectral moments for /s/-/ʃ/: mean, skewness, and kurtosis. The relation between individual acoustic measures and intelligibility will be presented. In addition, attempts to predict intelligibility from a linear combination of these acoustic measurements will be discussed. [Work supported by NIDCD.]

11:40–11:50

Discussion

3a WED. AM

## Session 3aSP

## Signal Processing in Acoustics and Structural Acoustics and Vibration: Identification of Structural Systems for Model Validation and Health Monitoring

David McCallen, Chair

*Lawrence Livermore National Laboratory, P.O. Box 808, L-156, Livermore, California 94551*

Chair's Introduction—8:25

### Invited Papers

8:30

**3aSP1. An overview of vibration-based damage detection research at Los Alamos National Laboratory.** Charles R. Farrar and Scott W. Doebling (MS P-946, Los Alamos Natl. Lab., Los Alamos, NM 87545, farrar@lanl.gov)

The ability to monitor the structural health of our aging infrastructure is becoming increasingly important. A wide variety of highly effective local nondestructive evaluation tools are available. However, damage identification based upon changes in vibration characteristics is one of the few methods that monitors changes in the structure on a global basis. The material presented herein will summarize the structural health monitoring research that has been conducted at Los Alamos National Laboratory over the last 8 years. First, the process of vibration-based damage detection will be described as a problem in statistical pattern recognition. This process has three portions: (1) data acquisition and cleansing; (2) feature selection and data compression; and (3) statistical model development. Current research regarding feature selection and statistical model development will be emphasized with the application of this technology to large-scale, *in situ* bridge structures and to bridge columns that were tested in a laboratory. In particular, the ability to quantify uncertainties in measured dynamic properties caused by variable environmental and operational conditions is addressed. The paper will conclude by summarizing some of the underlying technical issues associated with vibration-based global damage detection that continues to make this area of technology a challenging research topic.

9:00

**3aSP2. Seismic testing of full-scale structures through applied ground motion.** Peter A. Mote (The Nevada Testing Inst., 755 E. Flamingo Rd., Las Vegas, NV 89119) and Paul R. Gefken (SRI Intl., Menlo Park, CA 94025)

The Nevada Testing Institute (NeTI) and SRI International (SRI), in partnership with prominent research universities and state and federal government, and industry, have been evolving since 1995 a unique concept of generating earthquake-like ground motion to test large- and full-scale structures and structural systems at the Nevada Test Site. The RESCUE technique is used to produce strong ground motion by simultaneously expanding a planar array of buried vertical sources. The ground motion sources consist of a rubber bladder wrapped around a steel rectangular mandrel. Propellant is burnt in the source to produce low pressures that expand the bladder to move the soil bed. Because pressures are low, the soil is not damaged, allowing for sequential pulses and multiple tests on test structures. Eighty cycles of ground motion over durations of up to 120 s will be available in full scale. Proof-of-concept testing was successfully completed in 1998 in California and Nevada. The tests show that ground motions with accelerations of 1g to potentially 3g, velocities of 100 to 140 cm/s, and displacements of 60 cm to greater than 100 cm can be produced at the full scale test center. Construction of the full-scale testing center beginning in 2000 is planned.

9:30

**3aSP3. Damage detection for enhanced evaluation of structures.** Gregory A. Clark and Gregory C. Burnett (Lawrence Livermore Natl. Lab., P.O. Box 808, L-154, Livermore, CA 94551)

Monumental structures consist of public and private buildings, bridges and other transportation structures, dams, key industrial and scientific buildings and many special structures. Most of these structures are key to our social, commercial and technical infrastructure. All are vulnerable to damage or failure when stressed by abrupt, large-scale extreme events, the most common of which are earthquakes and wind. Forward reaching simulations, structural dynamic analysis techniques, innovative sensing, signal processing and remote communication form the core of this effort. Here we discuss our solutions to the structural damage detection problem. We take a model-based signal processing approach. This means that we exploit prior knowledge from first principle models for the structure to construct signal processing algorithms. Assuming that the models are representative of the actual structure, the model based approach offers significant performance improvement over algorithms based on measured data alone. Here we discuss the processor performance on both simulated and experimental data of a five-story structure.

10:00–10:15 Break

10:15

**3aSP4. Condition assessment of steel stringer highway bridges using experimental modal analysis.** Mike Lenett, Ahmet Turer, and David Brown (Univ. of Cincinnati, Cincinnati, OH 45221)

An experimental modal testing procedure has been developed for assessing the condition of steel stringer highway bridges using an optimized rapid impacting testing procedure. A 30 m to 90 m long two lane bridge can be instrumented and tested in a 5- to 6-h period using a multiple reference impact testing procedure. A special spatial domain multiple reference parameter estimation algorithm is used to obtain a modal model from the measured multiple reference impact data. The modal model is used to predict the static response of the bridge to various loading conditions. This predicted flexibility of the bridge is a good indication of the state of the bridge and can be used to assess the bridge. In the development phase of this procedure, the reliability of the resulting flexibility was verified by comparing measured truckload displacements against simulated truckload displacements, which were computed by virtually loading the modal flexibility matrix with a truckload. Furthermore, induced damage scenarios revealed that modal flexibility was an excellent objective, kernel index for both global and local condition assessment. The modal model is also used in the calibration of a nominal 3-D-FE model which is used in evaluating the rated load capacity of the bridge.

10:45

**3aSP5. Using signal processing techniques to improve finite element modeling.** Gregory C. Burnett, Gregory A. Clark, and James V. Candy (Lawrence Livermore Natl. Lab., P.O. Box 808, Livermore, CA 94551)

The use of signal processing parameter identification techniques to systematically improve finite element (FE) models of extended structures will be presented. FE models have been used with success for many years to simulate the response of structures, and work has recently been undertaken at LLNL to use FE models to model extended structures such as the Oakland/San Francisco Bay Bridge. At present, there is no way to systematically improve the FE models using measured displacements of the structures. Standard signal processing parameter identification techniques adapted for large order models have proven useful in both refining FE models given measured displacement and acceleration data as well as indicating points of failure for structures after damaging events such as simulated earthquakes. These abilities could prove useful in improving current FE models as well as speeding repair time for damaged structures. The presentation will identify the strengths and weaknesses of specific parameter identification algorithms and include illustrative examples of improvements to FE models.

### *Contributed Papers*

11:15

**3aSP6. Application of system identification to ultrasonic waveforms.** Jeffrey A. Daniels, Michael J. Anderson, and Tony J. Anderson (Mech. Eng. Dept., Univ. of Idaho, EPB 324K, Moscow, ID 83844-0902, anderson@uidaho.edu)

In the context of NonDestructive Evaluation (NDE), system identification is a process that can be used to extract sample properties from ultrasonic waveforms. Usually, system identification is applied directly to time series data. We describe a novel application of system identification. The process is applied to ultrasonic data in the frequency domain. Some investigators have pointed out that useful information can be obtained from the cepstral analysis of frequency domain data. Our approach further generalizes this practice, to the point that frequency domain data are identically mapped to a linear system, with frequency playing the role of time. We believe that our approach can be used to extract material properties, provide sensitive measures to flaws, return detailed information about flaws, extrapolate the bandwidth of ultrasonic measurements, and provide avenues for data compression.

11:30

**3aSP7. Pre-differentiated M-sequences for laser Doppler-based identification of vibrational systems.** Ning Xiang and James Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

A laser Doppler vibrometer (LDV) can be used to measure the surface velocity of a vibrating object in a noncontacting manner. One of the basic components in the LDV for sensing surface velocity is frequency demodulation where inevitable phase noise can impair sensing precision of the LDV. Owing to high-noise immunity and computational efficiency the fast M-sequence transform technique has received wide acceptance in various scientific and engineering fields. In the present work, pre-differentiated M-sequences have been employed as excitation signals for system identifications using a laser Doppler vibrometer. Along with inherent cross-correlation mechanism of the fast M-sequence transform, further improvement of phase noise immunity can be achieved. [The work is supported by U. S. Army Communications-Electronics Command.]

3a WED. AM

## Session 3aUW

## Underwater Acoustics: Multiple Volumetric Scattering

David R. Palmer, Chair

NOAA, Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, Florida 33149

Chair's Introduction—8:30

## Invited Papers

8:35

**3aUW1. Atmospheric lidar extinction and multiple scattering in haze, plumes, and clouds as analogues to ocean acoustic sounding.** Wynn L. Eberhard (NOAA Environ. Technol. Lab., R/E/ET2, 325 Broadway, Boulder, CO 80303, weberhard@etl.noaa.gov)

Sonar probing of fish schools, black-smoker plumes, and bubble clouds can experience attenuation (or “shadowing”) and multiple scatter. Similar kinds of effects are typical during atmospheric lidar (light detection and ranging) observations of turbid media, such as haze, smoke plumes, and clouds. Researchers have developed a variety of techniques to account for these effects on atmospheric lidar in different situations. Some of these methods might be adapted by ocean scientists with comparable problems. In order to be clear about the analogues, the physical conditions for the atmospheric problem will be listed. For example, scattering particles are usually assumed to be point targets randomly positioned in the air. Also, pertinent characteristics of the instrument will be specified, including the difference between “photon-bucket” detection (the common type of lidar) and “diffraction-limited” detection (more like sonar). The types of mathematical approaches for solving attenuation and multiple scattering problems in atmospheric lidar will be summarized. Some common atmospheric examples and the corresponding methods to correct the lidar data will be described.

9:00

**3aUW2. Multiple scattering in ocean bubble clouds—When one expansion parameter is large.** Frank S. Henyey (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Foldy constructed a model for collections of scatterers, such as a bubble cloud. He assumed the scatterers were arbitrarily small and placed at random, and that each one scattered isotropically. We are concerned with going beyond his approximate solution of his model. There are two expansion parameters in the multiple-scattering series for Foldy's model,  $4\pi\rho f/k^2$  and  $kf$ . The first of these is often large (for the bubble-cloud application), while the second is small. “Renormalization” can be used to handle the large parameter, but also renormalizes the other parameter. The talk will be concerned with the resulting expansion and a related issue, the transition between the low-frequency coherent backscattering and the high-frequency incoherent backscatter.

9:25

**3aUW3. Multiple scattering in fish and zooplankton acoustics is the exception.** Kenneth G. Foote (Inst. of Marine Res., P.O. Box 1870, Nordnes, N-5817 Bergen, Norway)

Reference is made to the general literature on multiple scattering. This indicates that multiple scattering is negligible if the following condition is fulfilled:  $\rho \ll 2\pi\lambda^{-2}\sigma^{-1/2}$ , where  $\rho$  is the numerical density of scatterers,  $\lambda$  is the acoustic wavelength, and  $\sigma$  is the maximum single-scatterer differential scattering cross section. The condition is examined for three classes of aquatic organisms: fish, euphausiids, and copepods. It is amply satisfied for all plausible organism concentrations and frequencies excepting possibly those of schooling swimbladder-bearing fish at resonance. [Partial support of the EU through RTD-Contract No. MAS3-CT95-0031 is acknowledged.]

9:50

**3aUW4. Low-frequency multiple scattering in fish schools: The self-consistent method for strongly coupled resonators.** C. Feuillade (Naval Res. Lab., Stennis Space Ctr., MS 39529-5004)

Low-frequency acoustic scattering from fish is typically dominated by the swimbladder resonance response. Schools of fish frequently consist of closely spaced individuals of similar size and, at resonance frequencies, this can cause multiple scattering processes between the fish to become significant and complex. The acoustic wavelength at the swimbladder resonance frequency is generally many times the length of the fish, and frequently greater than the dimensions of a small school. Since schooling fish often arrange themselves so that they are about one fish length apart, the scattered wave fields from neighboring fish will interact coherently. This feature must also be incorporated to realistically describe scattering from fish schools. An effective methodology is available through the application of self-consistent multiple scattering techniques, and a school scattering model based on this approach has



recently been developed and used successfully to analyze data [J. Acoust. Soc. Am. **99**, 196–208 (1996)]. Swimbladders are physically similar to bubbles, and the same approach may be used to develop a theory of multiple scattering in bubbly water which applies to strongly resonating scatterers and high-volume fractions. [Work supported by ONR.]

10:15–10:35 Break

### Contributed Papers

10:35

**3aUW5. Effects of interaction between two bubble scatterers.** George Kapodistrias (Dept. of Mech. Eng. and Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, georgek@apl.washington.edu) and Peter H. Dahl (Univ. of Washington, Seattle, WA 98105)

The backscattering of sound from two regularly arranged bubble scatterers is studied theoretically and experimentally. In well-controlled laboratory experiments a bistatic system is used to interrogate the scatterers, which are symmetrically placed on a very fine thread at a distance  $d$  from the combined beam axis of the set of transducers. The radius of each bubble was 585  $\mu\text{m}$ . The frequency range was 80–140 kHz, and by varying  $d$  the variable  $kd$  spanned the range 0.3–21, where  $k$  is the acoustic wave number. Scattering calculations were carried out using a closed form solution derived from the multiple scattering series. Several experiments were performed and the results are in close agreement with the calculations. It was verified that multiple scattering induces an oscillatory behavior about the exact coherent scattering level, with decreasing amplitude for increasing  $kd$ . For  $kd \approx 1$  the backscattered radiation is maximized, while for  $kd < 1$  the radiation is reduced considerably. These and other effects will be presented and discussed. [Work supported by ONR, Code 321 OA.]

10:50

**3aUW6. A theory with application to the scattering of waves from many objects.** M. F. Werby (Naval Res. Lab., Code 7181, Stennis Space Center, MS 39529)

Approximate formulations are available for classical wave scattering from aggregates of objects such as schools of fish, bubbles, rain drops, etc. Exact theories exist for two and three objects while scattering from many objects built upon that scheme has been difficult to develop. Other approximations describe the process in terms of an effective propagating wave number. Here a method is proposed that correctly accounts for scattering from an arbitrary number of targets in terms of a matrix expansion that allows for the isolation of the mean field of the scatters and effects from two, three, on to  $N-1$  correlations for  $N$  objects. This matrix series is convergent and uses single-particle  $t$ -matrices. This leads to terms that correspond to manifest physical processes, each less dominant than the previous term. This allows for truncation based on physical intuition. Further, at any level of truncation the scattering matrix that governs the entire process satisfies symmetry. This method may be reduced to some standard approximations by various limiting procedures. [Work supported by the Office of Naval Research and Naval Research Laboratory.]

11:05

**3aUW7. Backscattering from a large planar array of bubbles.** Redwood W. Nero, Charles H. Thompson, Christopher Feuillade (Naval Res. Lab., Stennis Space Center, MS 39529-5004), and Richard H. Love (Planning Systems, Inc., Slidell, LA 70458)

Backscattering measurements were obtained from a large planar array of bubbles located near the sea surface during preliminary tests of a towed midfrequency volume scattering measurement system. The planar array consisted of sheets of “bubble wrap” packing material suspended be-

tween the sea surface and 3 m depth. The array was originally intended as an approximation of a fish school to provide a strongly reflecting target at midfrequency for a field test of the towed system. Favorable at-sea conditions and successful operation of the towed system provided an opportunity to make additional measurements from the target using cw pulses at frequencies approaching the resonance expected from a single bubble. Weston’s multiple scattering theory for large planar arrays of small bubbles [J. Acoust. Soc. Am. **39**, 316–322 (1965)], and Feuillade *et al.*’s school scattering model [J. Acoust. Soc. Am. **99**, 196–208 (1996)] suggest that the resonance of a single bubble is suppressed and that the planar array should show a generally flat response over the frequency range of 1200 to 5000 Hz. Experimental results are presented and compared with theory and in addition, demonstrate the confounding effects of Lloyd’s mirror. [Work supported by ONR.]

11:20

**3aUW8. Validity of the small perturbation method for volume scattering from marine sediments.** Christopher D. Jones and Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105, cjones@apl.washington.edu, drj@apl.washington.edu)

For high-frequency volume scattering and propagation in marine sediments (at 10–500 kHz), the validity of single scattering methods (e.g., Born approximation) is not well established. Insight into whether single or multiple scattering is significant in a medium can be gained using effective medium theory (bilocal approximation to Dyson’s equation). An alternative analysis can be made using exact numerical Monte Carlo simulations. In this talk, both multiple scattering theory and numerical methods will be used to investigate the applicability of the small perturbation method for volume scattering from a sediment half-space. Parameters of the sediment are limited to marine sand and mud modeled as an acoustic fluid with Gaussian random fluctuations in density and compressibility. Sediment properties (from core sampling) are taken from published data on four experiment sites: Orcas Island; Eckernfoerde Bay; STRESS Midshelf; and Panama City. It is shown that strong scattering may be significant at the Panama City site (assuming the core estimates are accurate). However, even when strong scattering is predicted, the presence of a “transition layer” near the sediment–water interface reduces the effect of multiple scattering on the far-field scattering strength, and the first-order small perturbation approximation may still be applicable.

11:35

**3aUW9. Multiple scattering in marine sediments with volume inhomogeneities.** Anatoliy N. Ivakin (Andreev Acoust. Inst., Shvernika 4, Moscow, 117036 Russia, aniva@glasnet.ru)

Scattering by sediment inhomogeneities is recognized as one of major contributors to sea bed scattering and reverberation. In practical modeling, various single-scattering assumptions are normally used. While these assumptions may be reasonable in many cases, it is usually untested. Here, it is shown that conditions of validity for conventional single-scattering approximations can fail at realistic values of sediment parameters. In particular, the field can be strongly fluctuating and randomized in significant part of insonified sediment volume, which is in essential contradiction

with the main assumption of single-scattering theories. In certain cases, this difficulty can be overcome in the frame of the so-called cumulative forward-scatter single back-scatter (CFSB) approximation developed for the problem of wave scattering in the atmosphere [de Wolf, 1971]. Examples of application of this theory to sea bed scattering are presented and

multiple scattering effects are considered. Also, problems specific for seabed scattering are discussed which arise due to the greater strength of nonforward scattering by strong fluctuations of the density, the proximity of the water-sediment interface, at small grazing angles and near the critical angle. [Work supported by ONR.]

WEDNESDAY MORNING, 3 NOVEMBER 1999

UNION D ROOM, 9:00 A.M.

### Meeting of Accredited Standards Committee (ASC) S2 on Mechanical Vibration and Shock

to be held jointly with the

#### U.S. Technical Advisory Group (TAG) Meeting for ISO/TC 108 Mechanical Vibration and Shock (and Subcommittees ISO/TC 108/SC1, SC2, SC3, SC5, and SC6)

D. J. Evans, Chair S2 and Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108, Mechanical Vibration and Shock

*National Institute of Standards and Technology (NIST), 100 Bureau Drive, Stop 8221, Gaithersburg, Maryland 20899-8221*

**Accredited Standards Committee S2 on Mechanical Vibration and Shock.** Working group chairs will present reports of their recent progress on writing and processing various shock and vibration standards. There will be a report on the interface of S2 activities with those of ISO/TC 108 (the Technical Advisory Group for ISO/TC 108 consists of members of S2, S3, and other persons not necessarily members of those Committees), including plans for future meetings of ISO/TC 108 and/or its Subcommittees.

**Scope of S2:** Standards, specifications, methods of measurement and test, and terminology in the fields of mechanical vibration and shock, and condition monitoring and diagnostics of machines, but excluding those aspects which pertain to biological safety, tolerance and comfort.

WEDNESDAY AFTERNOON, 3 NOVEMBER 1999

MORROW ROOM, 1:00 TO 2:45 P.M.

### Session 3pAA

#### Architectural Acoustics: Measurements

Richard D. Godfrey, Chair

*Owens-Corning, Science and Technology, 2790 Columbus Road, Route 16, Granville, Ohio 43023-1200*

#### Contributed Papers

1:00

**3pAA1. Measurement of directional information in sound fields.** Bradford N. Gover (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada and Dept. of Phys., Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada), James G. Ryan, and Michael R. Stinson (Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

The directional characteristics of sound arriving at a listening point in a room are being studied with a view to improving microphone pickup of speech in teleconference rooms. Knowledge of the directions from which the direct, reflected, and reverberant sounds arrive is expected to be useful in microphone array design. At present, there exist several techniques which attempt to measure the direction of sound arrival, based on comparisons and postprocessing of multiple microphone signals. This postprocessing can involve intensity analysis, deconvolution analysis, or cross-correlation analysis. Each of these approaches has its benefits, drawbacks, and tradeoffs, and may perform better or worse in a given situation. Stud-

ies of several of these directional measurement techniques are underway to ascertain their usefulness and validity in a variety of situations, through both simulations and real room measurements.

1:15

**3pAA2. A boundary element method for pressure and intensity of broadband high-frequency, nonuniform, nondiffuse sound fields in enclosures.** Linda P. Franzoni, Donald B. Bliss, and Jerry W. Rouse (Mech. Eng. & Mater. Sci., Duke Univ., Durham, NC 27708-0300)

A boundary element method for high-frequency sound fields in enclosures is formulated in terms of time-averaged energy and intensity variables. The approach is applicable to high modal density fields, but is not restricted to usual low absorption, diffuse, and quasi-uniform assumptions. The basic building block for the method is a broadband intensity/pressure source. This source has directivity characteristics that are related to the surface correlation distance and the spatial distribution of the incident field

falling on the source location. These sources are distributed over the enclosure surfaces to satisfy boundary conditions on normal intensity. The boundary elements are constructed by integrating intensity/pressure sources over local panels. Sample calculations are performed for rectangular enclosures with different absorption levels and source positions. Results for spatially varying mean square pressures are compared with computationally intensive exact solutions and are found to give reasonable agreement at a small fraction of the computational cost. Some interesting issues in the modeling of spatially varying reverberant fields are also discussed. [This work was supported by NSF.]

1:30

**3pAA3. Anderson localization in coupled reverberation rooms.**

Richard L. Weaver, Oleg Lobkis, and Igor Rozhkov (Theoret. and Appl. Mech., Univ. of Illinois, 104 S. Wright St., Urbana, IL 61801)

It is argued that two undamped reverberation rooms subject to a transient source acting in the first of the rooms, if sufficiently weakly coupled, fail to achieve equipartition, regardless of the amount of time allowed. The critical parameter governing this localization is the ratio of coupling strength (as quantified by the rate at which energy initially leaks from one room to the other) to the modal density. Theoretical estimates for the time evolution of the small amounts of energy that do get transmitted are found to indicate that the energy in room 2 asymptotes at a level proportional to the square root of this ratio. Surprisingly, it is also found that the amount of energy in room 2 overshoots its asymptotic level by 5%–17%. The theoretical predictions are corroborated by numerical simulations of coupled two-dimensional rooms. Laboratory demonstrations, using ultrasonics in coupled elastic bodies, are also presented.

1:45

**3pAA4. A study of an impedance reflection model of sound fields in rooms using the Bergeron method.**

Hidemaro Shimoda (Inst. of Technol., Shimizu Corp., 3-4-17 Etchujima Koto-ku, Tokyo, 135 Japan)

In the Bergeron method [Shimoda *et al.*, Trans. Inst. Electron. Commun. Eng. Jpn., Part A **72**, 1–11 (1989); English transl., 1989 Electron. Commun. Jpn., Vol. 72; *Scripta Technica, Inc.* (Wiley, New York, 1990)], sound reflection surfaces in rooms are modeled by the simple resistances corresponding to the rigid walls. And the low-frequency responses for a middle scale of auditorium are agreed preferably with the real fields [K. Nakagawa and H. Shimoda, ASVA 97]. However, for big absorption acoustic models such as recording studios, or mixing rooms, the surfaces in rooms must be modeled by the impedance reflections. In this paper, the basic treatments for the impedance reflection models in the Bergeron method are investigated, and the wideband time-stretched pulses are used to simulate one-dimensional sound fields. The feasibility of numerical simulations using the Bergeron method are confirmed by supposing a simple surface impedance reflection model.

2:00

**3pAA5. Optimal absorption surface in offices and conference rooms.**

M. A. Picard Lopez, P. E. Solana Quiros, and J. V. Arizo Serrula (Departamento de Fisica Aplicada, Departamento de Ingenieria e Infraestructura de los Transportes, Universidad Politecnica de Valencia)

In the process of acoustic conditioning of enclosures it is necessary to follow an iterative process in order to obtain the suitable reverberation times (RTs) in the different frequencies, as a function of the use of the enclosure. In order to attain accurate results, it is necessary to modify

adequately the absorption characteristics of the internal surfaces of the enclosure and other devices that would have to be included in it. To simplify the calculation process, an inverse process can be used. Noting the desirable reverberation times, optimal absorption surfaces can be obtained. In this way, the most adequate materials and coatings can be selected from a database. From numerous studies and works related to the topic, the optimal reverberation times are determined for the type of enclosures mentioned. From this study, the optimal absorption surfaces can be obtained, which lead to the reverberation times already mentioned. In this work are supplied the graphics and functions in octave bands of the optimal absorption surface as a function of volume; the absorption average coefficient value of enclosure; the value of the constant optimum of the enclosure ( $R$ ); and the optimum reduction coefficient of noise (NRC).

2:15

**3pAA6. Experimentally validated statistical energy analysis models of wood and metal frame walls.**

Richard D. Godfrey (Owens Corning Testing Systems, 2790 Columbus Rd., Granville, OH 43023-1200)

Statistical energy analysis models were developed, using commercially available programs, for insulated and uninsulated wood and metal frame walls with various gypsum configurations. Material properties were taken from the literature. The 1/3 octave band results over the frequency range from 100 to 6300 Hz were compared with measured data from tests conducted in accordance with ASTM E 90. The predicted sound transmission losses were found to compare favorably with the measured data. The predicted power flows gave some insight into which paths were most significant, and the opportunity to alter them with changes in material properties.

2:30

**3pAA7. Uniform error distribution in one-third octave band frequency analysis.**

Alberto Martin, Manuel Recuero (I.N.S.I.A., Universidad Politecnica de Madrid, Crta. Valencia km7, Madrid 28031, Spain, amartin@sec.upm.es), Guillermo de Arcas, and Mariano Ruiz (Universidad Politecnica de Madrid, Madrid 28031, Spain)

In this paper an algorithm for obtaining a uniform error distribution in one-third octave band frequency analysis is presented. The algorithm is a digital signal processing implementation of the ANSI S1.11-1986 standard based on digital filters. First the effects of constant integration time over error distribution are discussed and then a solution is presented based on a statistical signal processing interpretation of the problem. The power distribution measurement is analyzed and studied as a variance estimation problem to establish the requirements of the analysis algorithm that will guarantee a uniform error distribution. The object of the algorithm is to obtain a uniform error distribution maintaining the total number of operations as low as possible in order to be efficient enough to be implemented in real time. The discussion is completed with several simulation results and a performance comparison with other popular alternatives.

**Session 3pAB****Animal Bioacoustics: Noise that Annoys Animals**

Ann E. Bowles, Chair

*Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109***Invited Papers****1:00****3pAB1. Biological perspectives on noise annoyance.** Ann E. Bowles (Hubbs-Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109, annb1@san.rr.com)

For humans, annoying noise is any sound that “disturbs, displeases, troubles, or irritates.” In humans, the experience of annoyance is a metric, obtained by self-reporting. In animals, however, biologists can only measure behavioral reactions or physiological events. Based on such measures, three classes of negative reactions to noise have been identified. The first is aversion caused by relatively automatic physical reactions, such as the startle. Unequivocal measures of aversion include learned avoidance, behavioral suppression, and irritability. The second can best be described as motivation to eliminate interference with function. Noise masking is an example of interference with function, and the response can be quantified by work animals are willing to do to overcome the masking noise. Finally, animals can become more defensive in the presence of noise. This change in psychological state is measured by increased levels of vigilance or arousal, increased activity, species-typical defensive behaviors, or attacks on the noise source. Although humans typically empathize most with the first reaction, aversion, interference with function or perpetual defensiveness have similar, if not greater, impact in the long-term. Predictions of impact on both humans and animals will improve with accurate measurements of these biologically meaningful responses to annoying noise.

**1:15****3pAB2. Neural circuitry and neurotransmitters that mediate the acoustic startle reflex in the rat.** Michael Davis (Dept. of Psychiatry, Emory Univ., 1639 Pierce Dr., Ste. 4000, Atlanta, GA 30322)

Because the acoustic startle reflex has such a short latency (e.g., 5 ms measured electromyographically in the neck, 8 ms in the hindleg), it must be mediated by a relatively simple neural pathway. We believe that the acoustic startle pathway is mediated by three synapses. These include auditory nerve inputs onto cochlear root neurons, which are cells embedded in the auditory nerve that receive the first afferent information in the auditory system. They send exceedingly thick axons (e.g., 7  $\mu\text{m}$  in diameter) to cells in the nucleus reticularis pontis caudalis which in turn project directly to motoneurons in the facial motor nucleus (pinna reflex) or spinal cord (whole body startle). Chemical lesions of cochlear root neurons or of cells in the ventral lateral nucleus reticularis pontis caudalis eliminate startle. Local infusion of either NMDA or non-NMDA antagonists into either the nucleus reticularis pontis caudalis or the spinal cord, via intrathecal infusion, markedly depress acoustic startle, suggesting that both of these glutamate receptor subtypes may mediate the acoustic startle reflex.

**1:35****3pAB3. Physiological responses of harbor seals, California sea lions, and northern elephant seals to simulated sonic booms, air-puffs, and white-noise bursts.** Lawrence F. Wolski (Hubbs-Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109, lwolski@hswri.org)

As part of a study examining auditory effects of simulated sonic booms on pinnipeds, changes in heart rate, orbicularis muscle activity, and respiration rate were measured. Four harbor seals, four sea lions and three elephant seals were exposed to simulated sonic booms (100 or 300 ms, 1, 3, or 6 psf). Two harbor seals and a sea lion were exposed to 50-ms air-puffs, and a harbor seal and sea lion were exposed to 300-ms white-noise bursts for comparison. Harbor seals responded to simulated booms with bradycardia, reduced movement, and brief apneas, consistent with an orienting response. Elephant seals showed similar responses, but of smaller magnitude. Harbor seal and sea lion responses to air-puffs included increased heart rates and muscle contractions (harbor seals), consistent with the startle response. Harbor seal responses to white-noise included muscle contractions and a brief increase, then decrease, in heart rate. Responses of sea lions to sounds were individually variable. While air-puffs stimulated increases in heart rate and muscle contractions, sounds stimulated both increases and decreases in heart rate, movement, and respiration. These results suggest that pinniped species react similarly to air-puffs, but variably to acoustic stimuli, possibly due to inter-species differences in perception. [Work supported by NASA: NAS1-20101.]

**1:55****3pAB4. Noise effects in learning experiments.** Louise M. McDonald (Dept. of Psych., Inst. of Psychiatry, De Crespigny Park, London SE5 8AF, U.K., louise.mcdonald@iop.kcl.ac.uk)

The effects of noise in learning experiments vary considerably, depending on the characteristics and duration of the noise, and the behavior through which learning is being measured. There can also be considerable individual differences in animals' responses to noise, even between members of the same species. Acute, mild noise is typically no more than a distractor. As noise intensity



increases, behavioral effects increase, ranging from increased rates of responding in lever-pressing experiments to freezing in radial-maze tasks. Some paradigms are relatively unaffected by noise, including purely appetitive tasks, such as solving mazes for food reward, and tasks such as the water maze, where the aversive properties of water immersion outweigh the aversive effects of noise. Individual differences in animals' responsiveness to noise can predict learning performance, with stronger responses to noise being associated with more effective learning. Research using rats in our own laboratory shows that chronic exposure to mild but unpredictable noise can attenuate latent inhibition, the capacity to ignore a stimulus that has previously been conditioned to be irrelevant.

2:15

**3pAB5. Behavioral reactions of peregrine falcons (*Falco peregrinus*) to military jet overflights.** Dana L. Nordmeyer (Dept. of Biol. and Wildlife, Univ. of Alaska-Fairbanks, 211 Irving Bldg., Fairbanks, AK 99775, nordmeyd@yahoo.com) and Daniel D. Roby (Oregon State Univ., Corvallis, OR 97330)

Behavioral reactions of peregrine falcons (*Falco peregrinus*) to military jet overflights were observed at 12 nests along the Tanana River, Alaska during each of the 1995–1997 breeding seasons. Low-altitude flights were conducted over a sample of nest sites under observation ("experimental" nests), while other nests were not intentionally overflown ("reference"). Productivity did not differ between experimental and reference nests and was within the normal range for Interior Alaska. Productivity also did not differ between nests exposed to many close overflights ( $\leq 1000$ -m slant distance from the nest) versus those exposed to few, nor between nests exposed to high levels of overflight noise versus those exposed to low levels. Adult peregrine falcons exhibited little or no overt behavioral reaction to the majority (78%) of observed close military jet overflights. Only 5.5% of reactions were classified as intense or flight. There were, however, significant differences among individual adult peregrines in the intensity of reaction to close overflights, after controlling for gender and slant distance. The average reactivity of individuals or pairs was inversely related to productivity. These results suggest that the sensitivity of breeding peregrine falcons to low-altitude jet overflights is a better indicator of subsequent productivity than actual dose of overflights.

### Contributed Papers

2:35

**3pAB6. Military training noise impacts on the endangered Red-Cockaded Woodpecker: Preliminary results.** Larry Pater, David Delaney, Timothy Hayden (US Army CERL, Champaign, IL), Robert Dooling, and Bernard Lohr (Univ. of Maryland, College Park, MD)

A new three-year study has been initiated to determine the degree of impact of military training noise on the endangered Red-Cockaded Woodpecker (RCW). The study is being conducted on an Army installation. The study design and preliminary results from the first nesting season are presented. Noise sources of interest include artillery and tank live fire blast noise, small arms live fire noise, small arms blanks noise and artillery simulator noise that occurs during maneuver training, and helicopter noise. Response criteria include proximate responses such as flushing from a nest and fitness measures such as nesting success. The objective of the study is to determine, for each type of noise, a dose-response relation and/or response thresholds. The study includes a determination of an audiogram for woodpeckers, which will provide the basis for selecting stimulus characterization metrics that include appropriate frequency weighting and will facilitate interpretation of response data. [Work supported by SERDP.]

2:50

**3pAB7. Noisy screams of macaques may function to annoy conspecifics.** Rachel Kitko, Daniel Gesser, and Michael J. Owren (Dept. of Psych., Cornell Univ., Ithaca, NY 14853)

It has been proposed that noisy screams produced by immature rhesus macaques are used as referential calls for help during conflicts with higher-ranking opponents and requisitely are structurally robust communicators of individual identity [Gouzoules *et al.*, 1984, *Anim. Behav.* **32**, 182–193 (1984)]. However, as recent data indicate that these sounds are not individually distinctive [Rendall *et al.*, *J. Acoust. Soc. Am.* **103**, 602–614 (1998)], an alternative possibility is that they are aversive to the agonist and that their function may primarily be to discourage further aggression. Noisy screams were recorded in a five-year study of both rhesus and Japanese macaques to determine contextual specificity, structural properties conducive to communication of individual identity, and acoustic features that may be aversive to other macaques. Analyses revealed that these calls were used in a variety of behavioral contexts and did not consistently elicit responses from possible allies. Calls were also routinely composed of multiple segments, including both broad-spectrum noise and various other acoustic features. Combined with the behavioral data, this diversity of features suggests a function of annoying rather than informing conspecifics.

3p WED. PM

**Session 3pEA****Engineering Acoustics: Acoustics of Cellular Telephones**

Martin Alexander, Chair

*63 Passaic Avenue, Summit, New Jersey 07901***Chair's Introduction—12:55*****Invited Papers*****1:00****3pEA1. Electroacoustic measurements on cellular telephones.** John R. Bareham (Consultant in Electroacoustics, 309 Riverview Dr., Ann Arbor, MI 48104)

The fundamental electroacoustic frequency response measurements on cellular telephones are send, receive, sidetone, and echo. Loudness ratings are calculated from these measurements. Measurements can be made by using a test instrument incorporating a vocoder, a reference base station, a digital audio interface, an actual telephone call, or by testing the phone in loopback mode. In each case, the measurement is calibrated in terms which relate to the public telephone network. Selection of test signals such as sine wave, pink noise, or artificial speech can influence the results. The ear simulator used to measure the receiver can be a sealed complex cavity, sealed but including a defined leak, or a complex cavity plus an anthropomorphic pinna. The choice of ear simulator dramatically influences the results of receiver measurements. Send response is usually measured in close proximity to a standard mouth simulator, but can also be measured in a diffuse sound field to assess pickup of room noise. Hands-free kits pose special measurement challenges which will be briefly reviewed. Key national and international standards associated with electroacoustic measurements on cellular telephones will be cited.

**1:25****3pEA2. Speech quality in wireless systems.** Hans Wilhelm Gierlich, Winfried Krebber, and Frank Kettler (Ebertstrasse 30a, D-52134 Herzogenrath, Germany)

Modern wireless communication systems require a very sophisticated design from the electrical as well as from the acoustical point of view. Since the maximum of functionality needs to be integrated in designs—getting smaller with every new generation—not much space is left for the acoustical components. In addition such sets typically are used in noisy environments. In order to provide a minimum of speech quality a careful design of the acoustical components as well as of the according electrical components is required. Very often there is a strong interaction between those. The presentation will describe the typical types of signal processing and acoustical designs used in such sets and will give methodologies on how to assess the speech quality relevant parameters of such sets. Measurements for optimizing the acoustical design will be described. In addition system evaluation procedures for nonlinear and time-variant signal processing are discussed.

**1:50****3pEA3. Realistic, acoustic telephone evaluation using a HATS (Head and Torso Simulator) based telephone test head.** Andre Matthisson and Soren Jonsson (Briel & Kjaer Sound & Vib. Measurement A/S, Skodsborgvej 307, DK-2850 Naerum, Denmark)

A substantial variety of test heads/simulators for testing the electroacoustic characteristics of telephone terminals exist today. They all intend to replicate essential features of typical human test crews, however, with varying fidelity. Traditionally, standardized measurements are performed using a table top test head (LRGP test head). In this paper we shall present a more anatomically representative "test head" using a head and torso simulator combined with a telephone handset holding device. Using this setup, more realistic measurements can be obtained, still complying with the geometries of the traditional table top test heads.

**2:15****3pEA4. Voice quality in new telephony systems.** Dan Quinlan (Bell Labs., Lucent Technologies, 101 Crawfords Corner Rd., Holmdel, NJ 07733)

During the development of the wired analog network, engineering teams recognized from the start that voice quality was a primary design driver. As such, projects aimed at understanding speech mechanisms and the human auditory system were a key component of the telephony research executed in the first half of this century. The recent migration from analog to digital (and wired to wireless) voice communication systems has given rise to a new set of signal degradations, and these impairments can have severe perceptual effects. While the ideal system would transmit voice at high bit-rates and with no signal degradation, voice-over-data (VODN) and wireless systems require the use of signal compression in the edge equipment, and the transmission paths are far from ideal. With regard to voice transmission, mainstream communications systems can no longer be thought of as linear and time invariant. The

development community is struggling to systematically evaluate the effects of these new impairments, and is beginning to understand that a connection must be made between objective specifications and subjective response. The talk will present a broad overview of the key issues, and will present some thoughts regarding how the communications industry is responding to these interesting problems.

2:40

**3pEA5. Sound solutions for cellular phones.** Morten Kjeldsen (Kirk acoustics A/S, Fuglevangsvej 45, 8700 Horsens, Denmark)

The ongoing miniaturization of cellular phones during the last years has led to the development of smaller and smaller speakers for these products. Today's size of a typical speaker for GSM and CDMA phones is of a diameter of 15 and 13 mm, and we are about to see even smaller speakers on the market. In the design and type approval of the receiving part of the cellular phone, the reference is a standardized measuring ear, which seals toward the ear cap plane of the phone. Under these conditions, it is possible to achieve a flat frequency response representing the whole telephone band, even with speakers in small sizes. However, such a tiny speaker in combination with a very small cellular phone having a poor surface of contact to the human ear leads to heavily reduced speech quality, which cannot be reflected in the measurement result. In other words, there is a very poor correlation between the type real life telephone user situation. This presentation gives the guidelines on how to design smaller cellular phones prepared for type approval using small speakers and at the same time improve sound quality.

### *Contributed Paper*

3:05

**3pEA6. Using experimental design to choose an acoustical model for a hands-free telephone.** Joris Brun-Berthet, Frédéric Laville, and Stéphane Dedieu<sup>a)</sup> (Ecole de Technologie Supérieure, Génie Mécanique, 1100 Notre-Dame Ouest, Montréal, QC H3C 1K3, Canada)

In full duplex speakerphones sound quality is adversely affected by the coupling between the speaker and the microphone. In addition to signal processing, this problem needs to be addressed by a better physical design of the telephone set. This requires the development of a computational model of the telephone acoustics. Due to the complexity of the geometry, the coupling between various systems and the rather large frequency band (150 Hz–7 kHz in wideband telephony), several complementary modeling

approaches such as BEM/FEM, SEA, analytical or empirical methods are required. A determination of the best modeling technique for each phenomenon in each frequency band is needed. Instead of a classical one-factor-at-a-time method, an experimental design approach was used. This method goes beyond the qualitative interpretation of the experimental results and gives a statistical significance to the observed effects of each factor and their interactions. Although this approach is widely used in engineering, no reference was found in the literature of its proposed use to help with acoustical model selection. The method and its application using telephone set prototypes will be presented. [The support of MITEL, Inc. and NSERC is gratefully acknowledged.]<sup>a)</sup> Presently at MITEL, Inc., Kanata, Ontario.

WEDNESDAY AFTERNOON, 3 NOVEMBER 1999 REGENCY BALLROOM SOUTH, 1:25 TO 2:30 P.M.

### **Session 3pID**

### **Interdisciplinary: Hot Topics in Acoustics**

Kenneth E. Gilbert, Chair

*National Center for Physical Acoustics, University of Mississippi, Coliseum Drive, University, Mississippi 38677*

**Chair's Introduction—1:25**

### *Invited Papers*

1:30

**3pID1. New approaches in acoustical signal processing.** James V. Candy (Lawrence Livermore Natl. Lab., P.O. Box 808, L-156, Livermore, CA 94551)

Signal processing is an integral part of most applications throughout the entire spectrum of acoustical applications, whether it is attempting to extract a resonant vibrational frequency of a structure to canceling noise in an active system or concert hall. Simply, signal processing is aimed at extracting the useful information from data, whether measured or simulated, while discarding the extraneous. Such simplistic ideas as time reversal in ultrasound have led to some rather astounding results in nondestructive evaluation. More sophisticated time-frequency representations incorporating wavelets are routinely applied in animal acoustics, while even more sophisticated techniques like matched-field processing incorporate a sophisticated oceanic propagation model as part of the scheme to localize submarines. In this paper we discuss the broad spectrum of approaches available to solve the basic acoustic processing problem with applications from various disciplines to emphasize the result. Processors discussed range from the simple to the sophisticated as dictated by the application. It is demonstrated how signal processing techniques are crucial to extracting the required information for success of the underlying application.

1:50

**3pID2. Hot topics in physical acoustics.** Robert M. Keolian (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804-0030)

A small fraction of the interesting work being performed in physical acoustics will be highlighted in this talk. From among the topics that could be chosen, two stand out in this author's mind, as of this writing. Dramatic improvements, by nearly a factor of 2, have been made in the efficiency of thermoacoustic engines by the thermoacoustics group at Los Alamos National Laboratory. Their new geometry brings the efficiency of thermoacoustic devices in line with conventional technologies, without sacrificing simplicity. Another innovative concept has been developed at the University of Mississippi. James Sabatier has invented a very successful method for the detection of buried land mines using acoustic-seismic coupling. This has the potential for saving a tremendous number of lives worldwide.

2:10

**3pID3. Hot topics in animal bioacoustics.** Sam Ridgway (Navy Marine Mammal Program, D3503, PLBS, San Diego, CA 92152-6266)

Discoveries in the amphibian flipper of the animal bioacoustics community include Alejandro Purgue's discovery that North American Bullfrogs emit 90% of the power of their calls from their eardrums, and only 10% from their vocal sacs has perhaps been the most original and surprising result of this decade. Peter Narins, Ted Lewis, and Blinda McClelland have shown that males of the endemic Madagascar tree frog have hyperextended their vocal repertoire to least 28 different calls. From the bat wing, von Helversen and von Helversen report that a bat-pollinated vine directs echolocating pollinators to its flowers by means of an acoustic guide. Valentine and Moss report on specializations in the bat superior colliculus for acoustic orientation by sonar. Three groups have new findings regarding temporary threshold shift (TTS) in marine mammals. Schusterman and Kastak observed a very small shift after broadband noise of over 20-min duration in a sea lion. Au and Nachtigall observed a 12-dB shift using one-octave noise of 179 dB, that is, 1  $\mu$ Pa in a bottlenose dolphin. Carder, Finneran, Ridgway, Schlundt and co-workers have produced TTS of 6 to 18 dB using 1-s tones of 192 to 202 dB in white whales and dolphins.

WEDNESDAY AFTERNOON, 3 NOVEMBER 1999

FAIRFIELD ROOM, 1:00 TO 3:15 P.M.

### Session 3pUW

## Underwater Acoustics: Propagation

Gerald L. D'Spain, Cochair

*Marine Physical Laboratory, Scripps Institute of Oceanography, La Jolla, California 92093*

Frederick D. Tappert, Cochair

*Applied Marine Physics, University of Miami, RSMAS, 4600 Rickenbacker Causeway, Miami, Florida 33149*

### Contributed Papers

1:00

**3pUW1. An effective modal analysis of an acoustic waveguide in the presence of a fine-scale heterogeneity.** John J. McCoy (School of Eng., The Catholic Univ. of America, Washington, DC 20064, mccoym@cua.edu) and Ben Z. Steinberg (School of Eng., Tel Aviv Univ., Israel)

A fine-scale heterogeneity refers to one that can be observed using a scale that is 50 to 100 times, or more, finer than that for observing the depth of the waveguide. At issue are the effects of this heterogeneity on the lower-order, say the first 10 to 20 modes, and their corresponding modal wave numbers. A multiresolution approach to the homogenization of the Sturm-Liouville eigenfunction/eigenvalue problem that determines these modes/wave numbers is used to obtain a self-consistent formulation. The fine-scale layering is described "effectively" in this formulation, by a spatial structure that is much less detailed than the actual structure. Demonstrated is an equality that applies to the eigenvalues of the "lower-order" modes for the given and the homogenized Sturm-Liouville problems, and a precisely described relationship between the corresponding

eigenfunctions. The demonstration is given both analytically and numerically. The "distortion" in the prediction of large-scale heterogeneity of an ocean waveguide, which would obtain as a result of an unrecognized, localized fine-scale layering is described and demonstrated.

1:15

**3pUW2. Pulse propagation of acoustic waves scattered in a three-dimensional channel.** Alan M. Whitman (Dept. of Mech. Eng., Villanova Univ., Villanova, PA 19085) and Mark J. Beran (Tel Aviv Univ., Ramat Aviv, Israel)

In a previous paper [A. M. Whitman and M. J. Beran, *Waves in Random Media* **11**, 1-11 (1999)] we discussed the scattering of acoustic waves by random sound-speed fluctuations in a two-dimensional channel and presented an asymptotic form for an acoustic pulse propagating in the channel. Here we include the three-dimensional effect of transverse scattering. We find an asymptotic solution in which initially the two-dimensional mode-transfer effect is more important than the transverse



scattering effect. However, for large enough propagation distances the transverse scattering dominates the pulse spread. Here we show the form of the pulse shape in both propagation ranges as well as the transition regime. We begin with a discussion of the physics of the problem and then present a mathematical discussion.

1:30

**3pUW3. A one-way coupled mode solution, Huygens' principle and mode coupling coefficients.** Michael F. Werby (Naval Res. Lab., Stennis Space Center, MS 39529)

The normal mode formulation for waveguide propagation has physical appeal since sound propagates in discrete entities—namely modes—and an understanding of the progression of modes sheds light on arrival times of pulses and features of ocean ducting. In the development of the range-dependent version several researchers employ the conservation of normal pressure and normal particle velocity along the  $z$  axis in the mean as the conserved quantity that allows for an algorithm to couple the modes as sound progresses in range. The usual derivations are based on R. Evan's seminal work on the two-way coupled mode formulation. Here it is shown that Huygens' principle and an approximate form of the Green's function in the waveguide leads to a suitable algorithm as long as reflected or backward going flux is small relative to forward flux. The derivation, moreover, leads to expressions for mode coupling amplitudes that allow one to follow the distribution of flux along specific modal tracts. [Work supported by the Office of Naval Research and Naval Research Laboratory.]

1:45

**3pUW4. Attenuation and dispersion in pulse propagation over an elastic sediment.** Wayne Jerzak, Rahul S. Kulkarni (Rensselaer Polytechnic Inst., Troy, NY 12180-3590), and Michael D. Collins (Naval Res. Lab., Washington, DC 20375-5320)

The parabolic equation method is used to model pulse propagation in an ocean overlying a lossy and dispersive elastic sediment. As for a fluid sediment [Kulkarni *et al.*, *J. Acoust. Soc. Am.* **104**, 1356–1362], loss and dispersion are included in the frequency domain propagation, and Fourier synthesis is used to recover the broadband signal. Loss is known to depend linearly on frequency in many materials. This is evidently because there are no relevant attenuation length scales in such materials. Lossy materials must also be dispersive in order to keep the waves causal, and the dispersive correction is a Hilbert transform of the attenuation as a consequence of the Kramers–Krönig relationship. In particular, the wave speed depends logarithmically on frequency when loss depends linearly on frequency. Examples are presented to illustrate the effects of dispersion on signals in shallow water environments. [Work supported by ONR.]

2:00

**3pUW5. Explosive beam spreading due to ray chaos.** Frederick Tappert (Appl. Marine Phys., Rosenstiel School, Univ. of Miami, Miami, FL 33149)

In an ocean acoustic situation where rays are known to be chaotic, a theory based on a Gaussian beam approximation is used to show that the width of a narrow-angle beam tends to increase exponentially (explosively) with range at the mean rate that is given by the mean Lyapunov exponent that is calculated in the geometric limit. Numerical simulations have been performed with the PE/SSF model that support the theoretical predictions. This finite frequency result shows that the phenomenon of “wave chaos” is real, and might be observable experimentally. [Work supported by ONR, Code 3210A.]

**3pUW6. Focal regions in spectrograms of long range propagation.**

William A. Kuperman, Gerald L. D'Spain (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0701), and Kevin D. Heaney (Sci. Applications Intl. Corp., 888 Prospect St., La Jolla, CA 92037)

A source near the deep sound channel axis excites mode groups (or paths) that involve both deep sound channel and boundary interacting propagation. Modal group speeds have a functional transition when passing through purely refractive to boundary reflecting phase speed regions. The result is that arrivals in this transition region line up in time across frequency. The combination of this alignment together with a similar, though broader coincidence of the last deep sound channel arrivals provide two time markers on a single phone spectrogram (intensity as a function of arrival time and frequency). Indeed, data from the Acoustic Thermometry of the Ocean (ATOC) program [ATOC Consortium, *Science* **281**, 1327 (1998)] show this effect. These time markers, depending on auxiliary information provide source localization or inversion capability. [Work supported by ONR and DTRA.]

2:30

**3pUW7. The measurements of the effect of storm-produced bubbles on high-frequency sound propagation.** Ralph R. Goodman (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804), Stephen J. Stanic, Jerald W. Caruthers, and Marcia A. Wilson (Naval Res. Lab., Stennis Space Center, MS 39529)

A high-frequency acoustics experiment to measure the spatial and temporal coherence of direct path signals was performed in shallow water, off the coast of Panama City, Florida. Transmission was along acoustic paths between fixed towers at middepth in 10 m of water. During the experiment there were two stormy days with strong winds, breaking waves and, nearby, high surf. It was observed that: (a) average time of arrivals increased significantly over those measured on calm days; (b) pulse to pulse variations in arrival times (over seconds) increased, showing an increase of variance of about a factor of 6, and (c) dispersion was observed. The range between the source and receivers was approximately 60 m. The only known cause for these variations is the presence of bubbles. Although no direct bubble density or distribution was measured, it is possible to construct a plausible bubble model to estimate the total average void fraction along the acoustic path and the variation necessary to cause the increase in variance. Void fractions on the order of  $2.6 \times 10^{-6}$  and fluctuations of  $0.9 \times 10^{-6}$  about this value can explain the observed data. [This work was supported by the NRL.]

2:45

**3pUW8. Changes of the acoustic modal phase velocity, group velocity, and interference distance in an eddy.** Gao Tianfu and Chen Yaoming (Inst. of Acoust., Chinese Acad. of Sci., 17 Zhonggauncun St., Beijing 100080, PROC, cym@oceana.ioa.ca.cn),

In this paper, by using the modal dispersion equation with the phase-integral approximation, and considering an eddy (or water mass) as a sound channel disturbance, the effects of the undisturbed channel, cold-core eddy, and warm-core eddy on the acoustic propagation characteristics are discussed. According to the solutions of the dispersion equation, the relation is obtained between the modal phase velocity ( $C_n$ ), group velocity ( $U_n$ ), and interference distance ( $L_n$ ), and the eddy intensity. When the plane wave (with an incident angle  $\alpha$ ) travels toward the center of a warm-core eddy (disturbed intensity  $B_M$ ), “double channel phenomenon” will take place in case of  $\sin^2 \alpha < B_M < 2(1 - \cos \alpha)$ , and then the modal phase velocity and interference distance will have anomalous changes. The interference distance of normal mode has a minimum. The larger the number of modes and the higher the source frequency, then the more obvious the “hollow case” for interference distance. However, for any cold-core eddy, the modal phase velocity or the interference distance will be monotonically changed.

3:00

**3pUW9. Experimental research on adaptive match of ocean sound channel.** Geng Chen (Inst. of Acoust., Chinese Acad. of Sci., Beijing 100080, PROC, chg@canna.ioa.ac.cn)

Ocean sound channel is a time-space-variant channel. Due to weighting of the channel's transfer function, acoustic signals are often distorted after they propagated, and with distorted results the receiver using correlation processing can not get its theoretical gain. In this paper based on the study of time-variant characteristics of the channel, the channel response function is found by adaptive filtering. A transversal filter with adaptive weights is used to simulate a time-variant multipath channel. The key

problem is how to increase the convergent speed of adaptive match algorithm. For this purpose, first the equation of adaptive filtering is solved by using the method of modified adaptive equation with adaptive step size factor. Second, by implementing adaptive filtering in the frequency domain, the calculation time is reduced. Third, by rational selecting the weight number and deterring the time relation among the weights, convergent speed increased greatly. Finally, convergent weights of the transversal filter are obtained as a response function of the channel and it can be used for channel match. The result of experiment shows that under the condition of rational selection of weight parameters a complicated response function of the channel can be expressed by very limited weight.

WEDNESDAY AFTERNOON, 3 NOVEMBER 1999 REGENCY BALLROOM SOUTH, 3:30 TO 5:00 P.M.

### **Plenary Session, Business Meeting, and Awards Ceremony**

Patricia K. Kuhl, President  
*Acoustical Society of America*

Business Meeting

Presentation of Certificates to New Fellows

Presentation of Awards

Science Writing Award for Professionals in Acoustics to Ilene J. Busch-Vishniac

Silver Medal in Biomedical Ultrasound/Bioresponse to Vibration to Ronald T. Verrillo

Silver Medal in Noise to Larry H. Royster

Trent-Crede Medal to David Feit

WEDNESDAY EVENING, 3 NOVEMBER 1999

FAIRFIELD ROOM, 8:00 TO 10:00 P.M.

### **Session 3eAO**

#### **Acoustical Oceanography: Workshop in Oceanography—Internal Waves in Deep and Shallow Water**

James F. Lynch, Chair  
*Woods Hole Oceanographic Institution, 203 Bigelow Building, Woods Hole, Massachusetts 02543*

#### **Chair's Introduction—8:00**

In this first "Oceanography Workshop" Robert Pintel (Scripps Inst. of Oceanography, La Jolla, CA 92093) and Lev Ostrovsky (Univ. of Colorado/NOAA Environmental Technology Lab., Boulder, CO 80303) will each present one-hour talks on deep and shallow water internal waves, respectively, and also lead discussions.

**Session 4aAA****Architectural Acoustics and Musical Acoustics: Archeological Acoustics I**

David Lubman, Chair

*David Lubman & Associates, 14301 Middletown Lane, Westminster, California 92683***Chair's Introduction—8:55*****Invited Papers*****9:00****4aAA1. Sense and nonsense concerning design and engineering of the theaters of classical antiquity and the writings of M. Vitruvius Pollio on the subject.** George C. Izenour (16 Flying Point Rd., Stony Creek, CT 06405)

This paper is a concentrated encapsulation of the theater architecture of classical antiquity utilizing paired slides of theater restorations in plan and perspective section accompanied by selected photographs of archeological sites. All of this is occasioned by four decades of the author's travels, researches, writings, ruminations, and conclusions concerning the theaters of ancient Greece and Rome. Considered in the abstract and irrespective of historical epoch, the functional purpose underlying all theater architecture is to provide the physical means by which an audience, preferably comfortably seated, is enabled to both see and hear a staged performance. It is the purpose of this paper to equate the archeologically revealed "SENSE" with the romantically derived fictitious "NONSENSE" of both the outdoor and the roofed theaters of classical antiquity. A contemporary perspective concerning the "SENSE" and the "NONSENSE" of the writings of M. Vitruvius Pollio on theater architecture is also provided. O tempora O mores.

**9:30****4aAA2. Acoustics of ancient theatrical buildings in China.** Ji-qing Wang (Inst. of Acoust., Tongji Univ., Shanghai 200092, PROC)

The performing arts in China could be traced back to a long history and had well developed during the Song and Yuan Dynasties, 11th–14th centuries. A unique form of pavilion stage, opened on three sides and thrust into the audience area, was then the most popular for the open-air theatres, the courtyard theatres, and the indoor theatres. As the traditional Chinese opera, Beijing opera and a variety of local opera, was performed in an abstract sense, no stage sets were used. Therefore, it was possible that the low stage ceiling and back wall of a pavilion stage could function as the reflective shell, which increased the early reflections and also the intensity of the sound to the audience, and meanwhile provided sufficient support to the singer. The measured reverberation times of these halls were around 1.0 s at major frequency bands that were optimal for occupied halls of moderate size. Such a stage house provided the performers a desirable sense that the auditorium was responsive to their effort. Illustrations of acoustical features in different types of stage house will be provided during presentation.

**10:00****4aAA3. The Rani Gumpha: A 2nd century B. C. acoustical theatre in India.** C. Thomas Ault (Theater Dept., Indiana Univ. of Pennsylvania, Indiana, PA 15705) and Umashankar Manthravadi (New Delhi, India)

The Rani Gumpha is a bi-level theatre, built in ancient Khalinga (now 5 km from Bhubaneswar, Orissa) by King Kharavela circa the 2nd century B.C. It is the first or Vikirsta type of theatre discussed by Bharat Muni in the *Natyashastra*, and is remarkably similar to the Hellenistic theaters of Greece, except that it has a square, not round, ground-plan. It is unique for its resonating chambers on the upper and lower levels which surround the performance space. We shall present a brief history of the site and, using visual and acoustical materials created from our on-site testing, demonstrate the effects of gain, resonance, and duration in the performance and spectator spaces. We shall also address vocal and instrumental range, harmonics, function of specific chambers and their interaction with the other chambers. Please note that this is a work in progress, not a complete or definitive assessment of the site. We anticipate that our colleagues will enhance our research through their shared expertise.

**10:30****4aAA4. Legends of echoes linked through acoustics to prehistoric art.** Steven Waller (American Rock Art Res. Assn., Deer Valley Rock Art Ctr., P.O. Box 41998, Phoenix, AZ 85080-1998)

A new area for acoustic research has arisen from a previously unsuspected relationship between ancient legends of echoes, and prehistoric rock art. This art includes Ice Age deep cave paintings, and Native American petroglyphs typically found high on canyon walls. This acoustic connection is important because there has been no satisfactory explanation of the motivation for the production of rock art (neither the unusual locations nor the subject matter the artists chose). Yet producing rock art clearly was a major preoccupation for early Homo sapiens sapiens over a span of tens of thousands of years. The concept of sound wave propagation and

reflection is a modern paradigm, and is based on the abstraction of invisible pressure waves being diverted by boundaries between media of different densities. In ancient times, however, the causes for many natural phenomena were explained by personification or animism, including attributing echoes to be the responses of spirits. Could echoes have motivated the ancient artists? Acoustic studies at rock art sites may be starting to answer this important question.

11:00

**4aAA5. Mayan acoustics: Of rainbows and resplendent quetzals.** David Lubman (David Lubman & Assoc., 14301 Middletown Ln., Westminster, CA 92683)

Progress is reported in understanding physical and mythological elements of the recently reported discovery that handclaps at the Mayan Temple of Kukulkan at Chichen Itza, Mexico produce birdlike descending chirped echoes. Echoes bear strong cognitive resemblance to the sound of the resplendent quetzal, the Mayan sacred bird associated with the temple. "Picket fence" (periodic reflections) from temple staircases explain the presence, frequency, and trajectory of descending chirps. But the rich harmonics are better explained by modeling staircases as acoustical analogues of inclined optical Bragg diffraction gratings. The remarkable conversion of handclap into chirped echo is explained through the time-dispersive, pulse-stretching properties of these gratings. Chirped echoes can thus be termed "acoustical rainbows" because acoustical energy is selectively dispersed over time, much as optical gratings selectively disperse colors over space. New evidence strengthens mythological arguments for intentional Mayan use of acoustical features. Echoes may have been exploited at solstice ceremonies. Preliminary evidence further suggests substantial sound reinforcement from the top of the temple to the plaza below. This could explain how Mayan kings addressed large crowds. Impressive flutter echoes and whispering galleries found at the nearby Great Ballcourt suggest possible widespread exploitation of acoustical architecture at this Mayan ceremonial site.

THURSDAY MORNING, 4 NOVEMBER 1999

KNOX ROOM, 8:15 TO 11:55 A.M.

### Session 4aBB

## Biomedical Ultrasound/Bioresponse to Vibration: Therapeutic Applications of Ultrasound

Inder Raj S. Makin, Chair

*Ethicon Endo-Surgery, 4545 Creek Road, Cincinnati, Ohio 45242*

### Invited Papers

8:15

**4aBB1. Acoustic hemostasis.** Lawrence A. Crum, Michael Bailey, Kirk Beach, Stephen Carter, Wayne Chandler, Peter Kaczkowski, Roy Martin, Pierre Mourad, Sandra Poliachik, Shahram Vaezy (Univ. of Washington, Seattle, WA 98105), George Keilman (Sonic Concepts, Woodinville, WA 98072), Thomas L Anderson, Lee Weng, and David M. Perozek (Therus Corp., Seattle, WA 98121)

When high-intensity focused ultrasound (HIFU) at megahertz frequencies and kilowatt intensities is applied to animal tissue, absorption can result in focal temperatures that exceed 70 °C within a few seconds. These high temperatures result in protein denaturation and coagulative necrosis, with the result that bleeding in the capillary bed and within small vessels is rapidly terminated. Additionally, if HIFU is applied to tears or cuts in large vessels, a combination of acoustic streaming and sound absorption can lead to acoustic hemostasis also in a few seconds. The role of acoustic cavitation in this entire process is not clear. We have examined the application of this phenomenon to trauma care and general intraoperative surgery and find that it has many promising attributes. We shall present a general review of our work in this area as well as our most recent results and their clinical implications. [Work supported in part by DARPA.]

8:35

**4aBB2. Influence of beam geometry and tissue boundaries on thermal ultrasonic lesions.** Frederic L. Lizzi (Riverside Res. Inst., New York, NY 10036, lizzi@rrinyc.org)

The size and shape of thermal lesions induced by high-intensity focused ultrasound (HIFU) are key factors in treating disease. Studies are being conducted to determine how HIFU beams and exposures can be designed to produce useful lesion attributes in specific target tissues. The studies include radially asymmetric beams and tissues with different acoustic absorption coefficients; sharp thermal gradients at such boundaries can significantly affect lesion geometry. Ocular tumor therapy uses asymmetric beams that produce lesions with ellipsoidal cross sections to facilitate treatments with closely packed lesion matrices. Three-dimensional simulations have analyzed heat patterns and lesions induced by these beams in irregular tumors whose geometry is derived from clinical serial-plane scans using 50-MHz diagnostic ultrasound. The simulations documented difficulties in treating boundary regions adjacent to the waterlike vitreous humor. Lesion results were consistent with *in vitro* and *in vivo* animal experiments. A related study involves an annular-array therapy system with a central rectangular aperture, housing a diagnostic array. The system (Spectrasonics, Inc.) is intended for treatments in various organs, including liver. Since the therapy beams are not radially symmetric, 3-D simulations and initial animal experiments have examined the asymmetry in induced temperature patterns and lesions. These investigations also addressed spurious grating peaks in axial intensity.



**4aBB3. Experimental studies of the use of a split-beam transducer for prostate cancer therapy in comparison to a single-beam transducer.** Jimmy S. J. Wu, Narendra T. Sanghvi, Michael Phillips, and M. Kuznetsov (Focus Surgery, Inc., 3940 Pendelton Way, Indianapolis, IN 46226, nsanghvi@focus-surgery.com )

The therapy of prostate cancer by high-intensity focused ultrasound (HIFU) requires the treatment of whole prostate tissue. Single-beam HIFU results in treatment time of about 3 to 4 h. However, the treatment time can be reduced by the use of multiple HIFU beams that are generated simultaneously. We have approached this problem by splitting the main ultrasound beam into one main beam that is surrounded by multiple sidelobes (coined "split-beam transducer"). In our split-beam transducer by means of geometrical configuration of the transducer, the main beam is surrounded by four sidelobes, producing a focal area about three times larger compared to that of a single beam. The step size between two adjacent ablative exposures can be increased using SBT. Experiments were designed to understand the use of "split-beam transducer (SBT)" on the Sonablate 200 device (Focus Surgery Inc.). Test objects, (1) Mylar strip, (2) Plexiglass block, and (3) turkey breast tissue were used. Temperature measurements using multiple thermocouples were made in the tissue near the beam entrance and about 5 mm deep into the samples. Results show that the split beam creates a larger volume of lesion than the single beam in turkey breast tissue. For the single beam used in BPH treatment step size was set at 1.8 mm to create connected necrosis. In the current study a step size of 2.5 mm can be used to create a connected lesion in turkey breast tissue, leading to a 30% reduction in treatment time for the same volume of tissue.

**4aBB4. Self-focusing ultrasound phased arrays for noninvasive surgery.** Emad S. Ebbini (Dept. of ECE, Univ. of Minnesota, Minneapolis, MN 55455)

Modern transducer technology is providing phased arrays with high-power, relatively large fractional bandwidth, and low cross coupling between array elements. As a result, it is possible to use these arrays in a dual imaging/therapy mode. We have recently demonstrated this using a 64-element 1-MHz array with  $1.5 \times 50$  mm<sup>2</sup> elements on a spherical shell with a 100-mm radius of curvature. In the therapy mode, the array was shown to be capable of producing focal intensities in excess of 10 000 W/cm<sup>2</sup> and cause well-defined lesions in tissue at depth. In the imaging mode, the array was shown to be capable of forming images of acceptable quality in a region covering  $\pm 35$  mm axially and  $\pm 25$  mm laterally around its geometric center. In this paper, we describe new algorithms for using this array in self-focusing mode based on acoustic feedback from the intended focal region. Two modes of acoustic feedback will be described and compared, specifically, direct hydrophone measurements at the desired focus and backscatter from the focal region (collected in imaging mode). Experimental data are given to demonstrate optimal focusing of phased array systems in the presence of strongly scattering objects, e.g., the ribcage.

**4aBB5. Ultrasonic tissue cutting with silicon surgical tools.** Amit Lal (Dept. of Elec. and Computer Eng., Univ. of Wisconsin, Madison, WI 53706)

Ultrasonic cutting is a well accepted way to cut tissues. However the exact nature of the cutting is complicated and not well understood. Cavitation, direct cutting, inertia stiffening, and thermal damage are just some of the effects believed to occur. Furthermore, we have been developing silicon-based ultrasonic surgical tools that hold promise for more functional surgical tools. In this paper we present results on the measurement of forces during ultrasonic cutting. Temperature sensors embedded in the tissue are used to measure any ultrasound-induced heating. We will present force and temperature data measured with actuators with different frequencies and intensities. Tradeoffs between the cutting tip sharpness and ultrasonic tip velocity effecting tissue damage will be presented in the context of Hertzian contact theory. We will use this data to formulate a phenomenological theory of ultrasonic cutting useful to design transducers for ultrasonic cutting. [Work supported by the Whitaker Foundation.]

**4aBB6. *In vivo* ultrasound-induced occlusion of arteries and veins.** Kullervo Hynynen and Vincent Colucci (Div. of MRI, Dept. of Radiol., Brigham and Women's Hospital, Harvard Med. School, 221 Longwood Ave., Boston, MA 02115)

In this study rabbit ear vessels were sonicated *in vivo* with focused spherically curved ultrasound transducer (diameter 100 mm, radius of curvature 80 mm, frequency 2 MHz). The sonication duration was varied from 0.2 to 5 s. The acoustic power was between about 50 and 200 acoustic watts. The vessels were visualized under microscope during and after the sonication to follow the blood flow and the vessel diameter. A video camera on the microscope was used to record the observations. The sonication caused blood vessels to contract temporarily at low powers and at high powers the vessels occluded and remained closed over the 2-h follow-up. However, the x-ray angiograms performed by injecting x-ray contrast agent in the ear vessels after the animal was sacrificed showed often that complete occlusion was not achieved. A further increase in the power caused the skin to coagulate and these effects had less effect on the blood flow. The vessels were completely sealed by performing a follow-up sonication at about 34 W for 10 s. This coagulated the tissue around the blood vessels and stopped the flow.

10:25

**4aBB7. A hemisphere transducer array for transskull ultrasound therapy and surgery.** Jie Sun, Greg Clement, and Kullervo Hynynen (Div. of MRI, Dept. of Radiol., Brigham and Women's Hospital, Harvard Med. School, 221 Longwood Ave., Boston, MA 02115)

Recently, it has been shown that it is feasible to generate a focused ultrasound field through an intact human skull for therapeutic purposes. Also it is clear that, in order to minimize the skull heating, a large 2-D array should be adopted to maximize the penetration area on the skull. With this in mind, an unconventional hemisphere transducer array is studied and developed. The radius of the transducer is chosen to be 15 cm in order to accommodate most head sizes and mechanical shifting while the transducer can still be fit into an NMR magnet coil to monitor treatments. The transducer's operating frequency, total number of elements (and the associated element size), and element geometry are determined based on numerical studies. The numerical results are presented for the transskull ultrasound field when the transducer is divided into 228 elements. The transducer is constructed and tested in water bath. The preliminary experimental results using this transducer are also presented.

10:40

**4aBB8. On the role of acoustic cavitation in enhancing hyperthermia from high-intensity focused ultrasound.** Patrick Edson, R. Glynn Holt, and Ronald A. Roy (Dept. of Aersp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

There are several physical mechanisms that lead to localized heating of tissue and tissue-like media with high-intensity focused ultrasound (HIFU). Experimental results obtained *in vivo* [Hynynen, *Ultrasound Med. Biol.* **17**, 157 (1991)] and *in vitro* [Edson *et al.*, *J. Acoust. Soc. Am.* **104**, 1844 (1998)] clearly indicate the existence of an insonation pressure threshold above which cavitation activity can profoundly enhance heating rates at MHz insonation frequencies. The dominant mechanisms through which bubbles facilitate heating have yet to be determined. Candidate mechanisms include viscous dissipation in the flow near the bubble surface, absorption of radiated acoustic waves, attenuation of multiply scattered waves, heat transfer across the bubble wall, etc. Results of numerical simulations designed to investigate the relative impact of these mechanisms on heat deposition from single bubbles and bubbly assemblages are reported. The critical roles of equilibrium bubble size and rectified diffusion are addressed. Our goal is to understand the physics of bubble-mediated heating in order to affect cavitation-enhanced hyperthermia in a reproducible and controllable manner. [Work supported by DARPA.]

10:55

**4aBB9. Sonoporation of monolayer cells by diagnostic ultrasound activation of contrast-agent gas bodies.** Douglas L. Miller and Jawaid Quddus (Dept. of Radiol., Univ. of Michigan, Ann Arbor, MI 48109-0553, douglm@umich.edu)

Fluorescent dextran (500 kD) uptake by sonoporation was observed after 1-min ultrasound exposure of human (A431 epidermoid carcinoma) and mouse (strain L connective tissue) cell monolayers in the presence of 1% Optison ultrasound contrast agent (Mallinckrodt, Inc.). Ultrasound exposure was provided by a 3.5-MHz array (Acoustic Imaging model 5200B, Dornier Medical Systems) operated in the spectral Doppler mode (5-s pulses, 4.4-kHz PRF). The array was mounted in a 37 °C water bath and aimed upward 7 cm away at a 1-mm-thick chamber with the monolayer grown on the upper 5-m-thick mylar window. For a water path, up to about 10% of the cells in a rectangular 0.37-mm<sup>2</sup> observation area (smaller than the -3 dB beam width) exhibited fluorescent dextran uptake for the high-power setting. For exposure with a 6-cm tissue-mimicking phantom, which approximated a -0.3 dB/cm/MHz derating factor, some fluorescent cells were noted even for low power (0.25 MPa

p-, 0.13 MI) increasing to about 4%–8% at high power. These results indicate that activation of contrast agent gas bodies by diagnostic ultrasound can be biologically effective at the cellular level. [Work supported by NIH CA42947.]

11:10

**4aBB10. Ultrasound-induced cell lysis and sonoporation enhanced by contrast agents.** Mark Ward, Jr. and Junru Wu (Dept. of Phys., Univ. of Vermont, Burlington, VT 05405)

The enhancement of ultrasound-induced cell destruction, lysis, and sonoporation in low cell concentration suspensions ( $2 \times 10^5$ /mL) by the presence of contrast agents (gas bubble to cell ratio = 3D 230) was demonstrated using cervical cancer cells (HeLa S3) suspensions containing micron-size denatured albumin microspheres filled with air (Albunex = AE) or octafluoropropane (OptisonTM). The suspensions were insonified by 2-MHz continuous or toneburst ultrasound in the near field. The spatial peak pressure amplitude was 0.2 MPa. The enhancement of cell destruction due to Optison was shown to be much higher than that due to Albunex for similar bubble concentration and ultrasound conditions. For toneburst exposures, significant lysis and sonoporation only occurred in the presence of a contrast agent. The majority of the bioeffects observed occurred in the first 5 min of exposure. The relationship between the enhancement of bioeffects and duty cycle of toneburst ultrasound appears to indicate that both stable gas spheres of contrast agents and cavitation nuclei created by the disruption of the gas spheres play a significant role in causing the bioeffects.

11:25

**4aBB11. Ultrasound-induced enhancement of skin permeability to Octa-L-Lysine.** Ludwig J. Weimann, Jr. and Junru Wu (Dept. of Phys., Univ. of Vermont, Burlington, VT 05405)

The efficiency of ultrasound in transdermal delivery of high molecular weight protein drugs has been demonstrated recently [S. Mitragotri *et al.*, *Science* **269**, 850–853 (1995)]. It has been shown in our previous paper [J. Wu *et al.*, *Ultrasound Med. Biol.* **24**, 1–6 (1998)] that ultrasound-induced structural disorder in human stratum corneum may be responsible for the enhancement of skin permeability. In this presentation, the effect of 20-kHz ultrasound on the permeation of Octa-L-Lysine tagged with Fluorescein (Octa-L-Lysine-4-FITC, effective molecular weight 2.5 kDa) through the skin was shown by optical reflection images and confocal microscopic cross-sectional images of skin. It has been indicated that both sonophoresis and sonoporation may be involved in the procedure. The possible mechanisms and bioeffects introduced by the procedure will also be discussed.

11:40

**4aBB12. A study of the effect of ultrasound on mandibular osteodistraction.** Tarek H. A. El-Bialy, Thomas J. Royston, Richard L. Magin, and Carlotta A. Evans (Univ. of Illinois at Chicago, Chicago, IL 60607, troyston@uic.edu)

Previous studies have shown that low-frequency ultrasound can enhance bone fracture healing. Other craniofacial research studies have indicated that distraction osteogenesis (bone lengthening) can be performed with some degree of success on the mandible. However, a complication typically reported is the bending of the anterior portions of the osteotomized mandibles which is believed to be caused by the stretching of the circumoral muscles on the newly formed bone callus. It is hypothesized that application of a low-frequency ultrasound treatment to osteotomized mandibles may enhance bone formation and minimize the healing period, thus minimizing the stretching of newly formed callus. Preliminary results are reported here which test this hypothesis on rabbit models. The ultrasound treatment regimen and the acoustic techniques used for assessing healing are discussed.

## Session 4aEA

## Engineering Acoustics: Acoustic Properties and Characterization of Materials

Stephen E. Forsythe, Chair

*Naval Undersea Warfare Center, 1176 Howell Road, Newport, Rhode Island 02841*

## Contributed Papers

8:15

**4aEA1. Analysis of ultrasonic nondestructive evaluation data using singular value decomposition of the Hankel data matrix.** Timothy C. Hanshaw, Chin S. Hsu (School of Elec. Eng. and Computer Sci., Washington State Univ., Pullman, WA 99164-2752), and Michael J. Anderson (Univ. of Idaho, Moscow, ID 83844-2752)

A method is presented for processing data from ultrasonic nondestructive evaluation (NDE) tests of material specimens by performing a singular value decomposition of the Hankel data matrix. The singular vectors and singular values of the Hankel data matrix correspond roughly to modal shapes and amplitudes in the ultrasonic waveform. A great deal of information about the specimen being tested can thus be conveyed with relatively little data; one approach which has been explored is to generate a colormap in which the intensity of the RGB components are determined by selected singular values. This approach was applied to simulated test data for both nominal and flawed specimens and the resulting colormap has been compared to a gray scale based on the rms value of the response data. The color map provides more information about the flaw than the gray scale in two ways: the intensity of the color components increases monotonically with increasing flaw severity, and the color map provides information about the flaw location.

8:30

**4aEA2. Recursive application of cross-correlation for extraction of sparse impulse trains from ultrasonic nondestructive evaluation data.** Timothy C. Hanshaw, Chin S. Hsu (School of Elec. Eng. and Computer Sci., Washington State Univ., Pullman, WA 99164-2752), Jeffrey A. Daniels, and Michael J. Anderson (Univ. of Idaho, Moscow, ID 83844-2752)

The goal of Quantitative Nondestructive Evaluation is to determine numerical values for a specimen's properties. In many cases, especially when the specimen is composed of thin layers, the material properties manifest themselves in a sparse train of impulses. This is the case for many state of the art hybrid materials. Signal processing techniques can be used to extract the desired impulse train from the ultrasonic data. In this paper, a cancellation method is used to recursively subtract a reference wavelet from the acoustic response of a test specimen in order to determine the desired impulse train of the specimen. The method was applied to acoustic data for a thin aluminum sheet. The resulting impulse train displayed enhanced fidelity as compared to the original wave form. Wave travel times and transmission coefficients thus obtained agree with expected values to within 6%.

8:45

**4aEA3. Ultrasonic spectroscopy inversion method for the determination of thicknesses and acoustical properties in a thin-layered medium.** D. Lévesque, M. Choquet, M. Massabki, C. Néron, and J-P. Monchalain (Industrial Mater. Inst., Natl. Res. Council of Canada, 75 de Mortagne Blvd., Boucherville, QC J4B 6Y4, Canada, daniel.levésque@nrc.ca)

Ultrasonic waves reflected from the front and back surfaces of a thin-layered medium ( $<2$  mm) overlap and interfere in the time domain. A spectroscopy approach has been proposed in the past, collecting and ana-

lyzing a large number of echoes together in the frequency domain to measure resonance frequencies and determine properties of the layered medium. With the presence of a fluid between layers, the resonance modes are mainly those of individual layers in the vacuum and an inversion algorithm has been successfully applied to recover some properties. For adhesively bonded layers, the resonance modes of the different layers are strongly coupled and important frequency shifts are observed. In this paper, a characteristic equation which includes an adhesion parameter has been derived and a numerical inversion technique has been used to simultaneously obtain thickness maps of the individual layers from measured resonance frequencies. One application is the laser-ultrasonic detection of hidden corrosion in aircraft lap joints where the paint layer has a significant impact on the resonance frequencies of the paint/metal-skin structure. Results will be presented for samples simulating aircraft lap joints with various levels of metal loss. [Work funded by the U.S. Department of Defense under Contract F33615-98-C-5200.]

9:00

**4aEA4. Ultrasonics for process monitoring and control.** Leonard Bond, Margaret Greenwood, and Judith Bamberger (Pacific Northwest Natl. Lab., P.O. Box 999, M.S. K5-25, Richland, WA 99352)

Ultrasonic waves are well suited for use in measurements that characterize multiphase fluids and flows. The waves can interrogate fluids and dense optically opaque suspensions. Transducers can give signals that penetrate vessels. Such transducers, together with supporting instrumentation, can provide on-line real-time measurements for use in process monitoring and control. Staff at the Pacific Northwest National Laboratory have developed a family of devices that measure fluid and slurry density, viscosity, flow rheology, particle size distribution, concentration, and velocity profiles. The transducers have been deployed on model systems to provide data that can be integrated to provide a more complete characterization of a process streams, including to detect time-dependent changing interfaces caused by fouling or phase changes, to track process conditions during mixing, sedimentation, stratification, and slurry transport. A summary of the fundamental physical acoustics for wave-process stream interactions will be provided. Examples that illustrate the capabilities of each type of measurement applied to slurry systems at Hanford site and for wood pulp process streams will be presented. The potential for enhanced process monitoring and control will be illustrated using wood pulp data, where it is required to measure the pulp consistency and degree of refining.

9:15

**4aEA5. Inexpensive technique for accurate determination of elastic properties by free decay in circular plates.** Thomas B. Gabrielson (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, tbg3@psu.edu)

Rapid determinations of elastic modulus are often made in the lab by measuring the resonances of regular bars of the material. When this is done, the most accurate results are obtained by suspending the specimen with as close to free conditions as possible. A circular plate supported by a taut wire through a central hole behaves accurately as an annular plate with both edges free. For low-loss materials, 10–20 modes are easily detected with an inexpensive microphone when such a plate is struck. Furthermore, the analytical solution for a thick, annular plate can be pro-

grammed readily in computational tools such as MATHCAD, MATLAB, or MATHEMATICA. (Even for plates traditionally considered thin—thickness-to-diameter ratios less than 0.1—thin-plate theory can produce significant errors for higher-order modes.) The first two resonance frequencies can be used to determine the elastic modulus and Poisson's ratio and the higher modes can be used to assess the uncertainty or to uncover anisotropy. The resonance frequencies can be measured with a fast Fourier transform spectrum analyzer and an understanding of the response of such an analyzer to a slowly decaying sinusoid. [Work supported by the Naval Sea Systems Command.]

9:30

**4aEA6. Estimation of model-parameter errors.** M. Roman Serbyn (Phys. Dept., Morgan State Univ., 1700 E. Cold Spring Ln., Baltimore, MD 21251)

The linear time-invariant two-port, characterized by four complex parameters, has established itself as a reliable model for a variety of physical systems. It has been studied in two aspects: as the "forward" problem (analysis) and the "inverse" problem (synthesis). The focus of the research here reported is on the latter, that is, on inferring the values of the model parameters from the results of measurements at the input and output ports. In particular, the goal of this investigation has been to quantify the relationship between uncertainties in the measured values and the resulting errors in the parameter estimates. Two physical systems have been used in the present study: an electrical circuit with known component values and an uncalibrated electromechanical transducer. Mathematically, each device was modeled as a Moebius transformation,  $w = (az + b)/(cz + d)$ , whose parameters,  $(a, b, c, d)$ , were estimated from least-squares fits to measured values of the variables  $(z, w)$ . The *a priori* known electrical circuit provided a base line for checking the validity of the computation algorithms. MAPLE, MATLAB, and SPICE programs were utilized. [Work supported by HUD Special Projects Grant No. B98SPMD0074.]

9:45

**4aEA7. Electromechanical material properties for new high-power sonar transduction materials.** Elizabeth McLaughlin and James Powers (Code 2132, NAVSEA Newport, 1176 Howell St., Newport, RI 02841)

New high-power electroactive materials such as PMN-PT ceramic and PZN-PT single crystal are being considered for Naval sonar applications. But unlike conventional PZT ceramic, these new materials often have a large mechanical stress and temperature dependence. In this presentation we will describe a method which characterizes the stress and temperature dependence of these new materials under high-drive conditions and present data on several recently measured materials. The measurement generates the quasistatic polarization and strain versus field curves for different values of compressive stress and temperature, and the stress versus strain curve for different values of dc bias field. From these data, the large signal piezoelectric  $d$  constant, dielectric constant, short-circuit Young's modulus,  $k_{33}$  coupling factor, and energy density can be determined. As an example, the large-signal piezoelectric constant for an 80/20/2 (PMN/PT/Lanthanum) material dropped 30% for an increase in temperature of 23 °C and dropped 14% for an increase of 34-MPa prestress. For each of the same conditions the coupling factor dropped 13% and the energy density dropped 22% and 13%, respectively. [Work supported by ONR, SPAWAR, and PEO-USW.]

10:00–10:15 Break

10:15

**4aEA8. Sonar dome deflection measurements on the USS Radford.** Joel F. Covey, Robert D. Corsaro (Naval Res. Lab., Washington, DC 20375), Diane B. Weaver, and Jonathan B. Walker (SFA, Inc., Largo, MD 20774)

Sonar dome deflection measurements were performed on the new monolithic sonar dome of the USS Radford in Oct.–Dec. 1997. Over 700 Mbytes of raw data from 22 environmental sensors and 128 displacement

sensors were obtained. The at-sea data acquisition was highly successful with data collected over sea states 1 through 4, and for numerous ship speeds up to 31 knots. Analysis of this data has provided a more complete characterization of previously measured sonar dome deformations. Many new conclusions and observations on sonar dome behavior have been observed and correlated with ship motions.

10:30

**4aEA9. A calculation of the flow resistivity in intermediate flow regime.** M. A. Picard Lopez, P. E. Solana Quiros, and J. V. Arizo Serrulla (Dept. de Fisica Aplicada, Univ. Politecnica de Valencia, Camino de Vera, 14, 46022 Valencia, Spain)

Many of the acoustic properties of fibrous materials can be modeled and predicted with the aid of appropriate empirical formulas. Among the input parameters that are required by such types of macroscopic models, the flow resistivity is a very well-established quantity in the literature, as usually this parameter is measured in steady Poiseuille flow conditions, i.e., at comparatively low flow speeds. One major concern with respect to such a parameter is the question of its possible frequency dependence. From a theoretical point of view, we have examined to which extent the capilar pore approximation can be utilized in intermediate flow regimes. The tendency towards lower values of flow resistivity for increasing acoustic Reynolds number also appears in all samples studied. And only the greater density samples, for the tested frequencies, were found in the flow regime of Poiseuille. However, if we want to obtain the value of the flow resistivity, for any sample, the employed experimental procedure makes it possible by a relationship between the experimental values and those of the Biot function. The value of the flow resistivity in different dynamical flow regimes to that of Poiseuille can be obtained.

10:45

**4aEA10. Normal incidence sound absorption coefficient measurement using sound intensity technique: An accuracy investigation.** Mohamad N. Dimon, Ahmad K. Said (Elec. Eng. Faculty, UTM, 81310, Skudai, Johor, Malaysia), Md. N. Ibrahim (UTM, 81310, Skudai, Johor, Malaysia), and Md. Y. Jaafar (MARDI, Serdang, Selangor, Malaysia)

Sound absorption data are required in specifying and proposing the correct materials for room acoustic treatment. However, there are instances where noncommon materials such as direct piercing carved wood panels (dpcwps) are used without sound absorption data. Sound absorption measurements obtained using an impedance tube and reverberation chamber are not accurate for dpcwp. This is primarily because the wall backing the dpcwp during the measurements constitutes the carved surface, which does not exist in actual installation. Therefore, an alternative measurement technique using a sound intensity measurement technique is more appropriate for dpcwp. This paper focuses on experimental methodology where various aspects need to be taken into consideration to achieve accurate results. Various criteria affecting measurement instruments' capabilities and other factors affecting sound intensity measurement accuracy will be discussed. Sound absorption measurement was conducted on three different perforation panels which resemble dpcwp and it was found that the sound absorption measured with a deviation from an average value of  $-5.59\%$  to  $5.95\%$  can be achieved, and is particularly accurate.

11:00

**4aEA11. Rayleigh waves in anisotropic porous media.** Alexander Kaptsov and Sergey Kuznetsov (Inst. for Problems in Mech., Prospect Vernadskogo, 101a, Moscow, 117526 Russia)

Rayleigh waves represent waves propagating in a half-space and attenuating exponentially with depth. During the last 30 years considerable progress has been achieved in developing both analytical and numerical methods for analysis of Rayleigh waves propagating in homogeneous media with arbitrary elastic anisotropy. The propagation of Rayleigh waves in heterogeneous media has not been studied nearly as often. Very few works are devoted to the analysis of speed and attenuation of Rayleigh waves in media (mainly isotropic) with uniformly distributed pores and/or



microcracks. The advantage of applying Rayleigh waves to the nondestructive determination of concentration and the preferred orientation of pores lies in its high sensitivity to variation of the material elastic parameters due to preexisting pores. The developed approach for analysis of Rayleigh waves in porous or cracky anisotropic media is based on a com-

ination of six- and three-dimensional complex formalism and the two-scale asymptotic analysis. In its turn, the latter utilizes a newly developed spatially periodic boundary integral equation method. This is used for determination of the effective characteristics of heterogeneous media containing isolated uniformly distributed pores. Numerical data are discussed.

THURSDAY MORNING, 4 NOVEMBER 1999

HAYES ROOM, 10:00 A.M. TO 12:30 P.M.

### Session 4aED

#### Education in Acoustics: Hands-On Demonstrations

Uwe J. Hansen, Chair

*Department of Physics, Indiana State University, Terre Haute, Indiana 47809*

#### Chair's Introduction—10:00

Approximately 20 experiments designed to introduce various acoustics principles will be set up in the room. After a brief introduction of each demonstration they will be available for experimentation by visiting high school students. Space permitting, conference participants are also invited to view the demonstrations and perform the hands-on experiments, provided such activity does not interfere with the purpose of the session, to introduce acoustics experiments to high school students. Experiments performed by the students will include many, ranging in sophistication from simple wave studies to mapping of normal modes in two-dimensional structures.

THURSDAY MORNING, 4 NOVEMBER 1999 REGENCY BALLROOM SOUTH, 9:25 TO 10:30 A.M.

### Session 4aID

#### Interdisciplinary: Distinguished Lecture on Smart Structures and Microelectromechanical Systems (MEMS)

Jerry H. Ginsberg, Chair

*School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332*

#### Chair's Introduction—9:25

#### Invited Papers

9:30

**4aID1. Smart structures and microelectromechanical systems (MEMS).** B. T. Khuri-Yakub (E. L. Ginzton Lab., Rm. 11, Stanford Univ., Stanford, CA 94305-4085)

Capacitor transducers have been around for as long as piezoelectric transducers. They have not presented much competition for piezoelectric transducers as transmitters and receivers because the method of their manufacture did not optimize and highlight their performance. With the advent of silicon micromachining, it is now possible to make capacitors with very thin gaps that sustain electric fields of the order of 109 V/m or more. At these levels of electric field, the transformer coupling between the electrical and mechanical parts of the capacitor transducer, and thus its performance, become comparable to that of piezoelectric transducers. Advantages such as practically infinite bandwidth, ease of manufacture, and the ability to integrate electronic circuitry make this type of transducer a very important candidate for many ultrasonic applications. We will review the design and performance of capacitor transducers for both immersion and airborne ultrasound applications. Single-element, 1-D, and 2-D arrays of transducers will be presented along with imaging results. The overall dynamic range of systems with these transducers will be shown to be over 140 dB/V/vHz. We will also present a design for making surface-wave and Lamb-wave transducers using the same capacitor concept. [This work was sponsored by the Office of Naval Research, the Defense Advanced Research Projects Agency, and the Air Force Office of Scientific Research.]

4a THU. AM

**Session 4aMU**

**Musical Acoustics: Music, Rhythm and Development I**

Caroline Palmer, Cochair

*Department of Psychology, The Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210*

Mari Riess Jones, Cochair

*Department of Psychology, The Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210*

***Invited Papers***

**9:00**

**4aMU1. Music, rhythm, and development: Chairs' introduction.** Caroline Palmer and Mari Riess Jones (Dept. of Psych., Ohio State Univ., 1885 Neil Ave., Columbus, OH 43210, palmer.1@osu.edu)

This special session is sponsored by the Caroline B. Monahan Fund and by the Center for Cognitive Science at Ohio State University. Caroline Monahan's contributions to the field of music cognition focus on auditory perception and memory, including topics of pitch and temporal parallels in music and speech perception, recognition memory for rhythmic and pitch patterns, and the role of timbral and loudness variables in discrimination of auditory patterns. Dr. Monahan's work has been concerned with applications of these topics in speech as well as in music, including language acquisition, hearing impairment, and second-language learning. Dr. Monahan received her Ph.D. at UCLA in 1984 in Psychology, and served as a Research Scientist at the Central Institute for the Deaf. She also served as a Research Associate in the Department of Communication Disorders at the University of Oklahoma. We acknowledge her contributions to and generous support of the field of music cognition.

**9:05**

**4aMU2. Memory and the experience of time.** W. Jay Dowling (Prog. in Cognit. Sci., Univ. of Texas at Dallas, Richardson, TX 75083-0688)

Researchers reflecting on the perception of music and speech, from William James a century ago to Eric Clarke very recently (and many in between), have been driven to the conclusion that we do not perceive auditory events instant by instant as they are received by the ear. Rather, we perceive the contents of our working memory where "sound bites" of the order of 5 s in length can be heard. This picture is complicated by the consideration that the contents of working memory are themselves in a state of flux over time. Recent research is reported showing ways in which memory for novel melodies changes systematically over periods of 5 min following presentation. Implications for the understanding of music and the experience of time are drawn.

**9:30**

**4aMU3. Issues in music, rhythm, and development.** Mari Riess Jones (Dept. of Psych., Townshend Hall, Ohio State Univ., Columbus, OH 43210, Jones.80@osu.edu)

Some general issues pertaining to the response of infants and children to auditory patterns will be discussed. One issue concerns the tempo of simple auditory sequences used as stimuli in perception and production experiments with subjects of different ages. Is there any evidence that children exhibit preferences for certain tempi? If so, do these preferences change with age? What implications can we draw from such findings? A second issue concerns children's sensitivity to relative time structure, including hierarchical time relationships. To what extent are children of different ages responsive to higher-order time relationships, e.g., as in metrical time relationships? A third issue also concerns relative timing. Are young children and/or infants preferentially sensitive to certain kinds of rhythmic relationships? If so, what might these be and what does this imply for understanding the development of rhythmic competencies? This talk will selectively review research addressed to these topics. The aim is to suggest a few general themes that are emerging in the field.

**9:55**

**4aMU4. Children tap faster and within a smaller tempo window than adults.** Carolyn Drake (Laboratoire de Psych. Exp., 28 rue Serpente, 75006, Paris, France)

Dynamic Attending Theory [Jones (1976)] proposes that, when listening to auditory sequences, listeners adapt their internal rhythms to those in the sequence (attunement), spontaneously focusing on events occurring at intermediate rates (referent level), and shifting attention to events occurring at faster or slower hierarchical levels (focal attending). The development of these abilities with age and musical training is examined by comparing motor tapping in 4-, 6-, 8-, and 10-year-old children and adults, with or without musical training. Seven motor tapping tasks (including spontaneous motor tempo and synchronization with simple sequences and music) revealed three changes with increased age and musical training: (1) a slowing of mean tapping rate (a reflection of referent period) and mean synchronization rate (a reflection of referent level); (2) an increase in range of tapping rates toward slower rates

(improved focal attending); and (3) a greater ability to adapt the taps to sequence structure, in particular incorporating a greater use of the hierarchical structure (improved attunement and focal attending). Interpreted in terms of attentional oscillators, these results suggest that increased interaction and experience with a particular type of sequence facilitate the passage from the initial use of a single oscillator toward the coupling of multiple oscillators.

**10:20–10:30 Break**

**10:30**

**4aMU5. Mothers' speech and song for infants.** Sandra E. Trehub and Tonya R. Bergeson (Univ. of Toronto at Mississauga, Mississauga, ON L5L 1C6, Canada)

The speech and singing interactions of mothers with their 4- to 7-month-old infants were recorded on two occasions separated by approximately one week. Repetitions of the same songs across sessions had similar tempo (mean difference of 3.73 beats per min) and pitch level (mean difference of 1.07 semitones). By contrast, repetitions of stereotyped verbal phrases had considerably greater differences in tempo (22.82 beats per min) and pitch level (4.96 semitones). Nevertheless, utterance repetitions were perceived as similar to the original because they retained its rhythmic structure. The characteristic focus of researchers on the pitch patterning of maternal speech may be obscuring important consistencies in the spoken rhythms of caregivers. Moreover, the demonstrable sensitivity of young infants to the rhythmic structure of auditory sequences makes it essential to document temporal regularities in the spoken and sung messages that are typically directed to infants. Rhythmic aspects of maternal speech may well emerge as critical components of the infant-directed speech register. [Work supported by the Social Sciences and Humanities Research Council of Canada.]

**10:55**

**4aMU6. Recognition of prosodic cues in music performance.** Caroline Palmer, Melissa K. Blakeslee (Psych. Dept., Ohio State Univ., 1885 Neil Ave., Columbus, OH 43210, palmer.1@osu.edu), and Peter W. Jusczyk (Johns Hopkins Univ., Baltimore, MD 21218)

As early as infancy, humans are sensitive to prosodic cues that can aid recognition of words embedded in sentence contexts. We addressed the question of whether similar cues formed by expressive nuances in music performances aid listeners in recognition of musical phrases embedded in melodic contexts. Utilizing a task similar to infant-research habituation paradigms, we report experiments with musically experienced and inexperienced listeners who are familiarized with performances of musical phrases that were identical in pitch/duration contents but differed in their intensity and articulation cues. Listeners then completed a recognition task for performances of the same musical excerpt whose cues either matched or did not match the performances at familiarization, and whose cues were either consistent or inconsistent with the rhythmic context in which the excerpts were embedded. Findings show that listeners can distinguish musical phrases that differ only in expressive nuances, and a mismatch of expressive nuances to the rhythmic context can facilitate recognition. These findings suggest that the prosodic cues that differentiate human performances are part of listeners' memory for melodies, and similar acoustic features may enable the recognition of auditory events in music as in speech. [Work supported by NIMH.]

**11:20**

**4aMU7. Music cognition and Williams Syndrome.** Daniel Levitin (Stanford Univ., Stanford, CA 94305) and Ursula Bellugi (Salk Inst. for Biological Studies, La Jolla, CA)

New studies of the musical abilities of individuals with Williams Syndrome (WMS) are presented. WMS is associated with poor spatial, quantitative, and reasoning abilities, coupled with excellent face processing and relatively preserved language abilities. WMS individuals tend to have richer (larger and more colorful) vocabularies than would be expected for their mental age and tend to be more emotionally expressive than normals. There have long been anecdotal reports of increased musicality in WMS (many of them are competent musicians who play from memory) and these are supported by the finding that the planum temporale (PT, a region adjacent to A1) is proportionately enlarged in WMS; in addition, the PT is unusually asymmetric in WMS, with the left side enlarged in a similar fashion to that found in professional musicians. The authors report new studies of rhythm production and memory for music in WMS. The results suggest that this domain of cognitive ability may also be relatively preserved. These findings provide additional evidence that music constitutes a domain-specific higher-level system of expertise in humans, perhaps independent of other mental abilities.

**11:45–12:00**

**Panel Discussion**

**Session 4aNS****Noise: Progress Report and Discussion on the Continuing Activity of ASA's Role in Noise and Its Control**

Louis C. Sutherland, Cochair

*27803 Longhill Drive, Rancho Palos Verdes, California 90275-3908*

David Lubman, Cochair

*David Lubman & Associates, 14301 Middletown Lane, Westminster, California 92683*

The Technical Committees on Noise and Architectural Acoustics are holding a meeting to review current progress and invite further discussion on activity to increase the role of the ASA in noise, noise control and related architectural acoustics issues. This outreach effort has included activity on increasing public awareness about noise and noise control, on public hearing screening testing and development of self-testing techniques, on seminars on industrial noise, on meeting room acoustic environments and, currently of special interest, on classroom acoustics. Discussions about these and/or new related activity, including encouraging joint activity in these areas with other professional organizations, will be encouraged. Attendees will be particularly encouraged to discuss elements of a draft standard on classroom acoustics that is currently being developed.

**Session 4aPA****Physical Acoustics and Committee on Archives and History: History of Physical Acoustics**

Henry E. Bass, Chair

*National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677***Chair's Introduction—10:45*****Invited Papers***

**4aPA1. The history of physical acoustics: The worldwide scene.** Robert T. Beyer (Dept. of Phys., Brown Univ., Providence, RI 02912)

The term physical acoustics, used by von Helmholtz in his book in 1862 to distinguish the subject from physiological acoustics and music, did not become popular until after World War II. The term covers the production, transmission, and reception of sound, as well as the interaction of sound with matter. With such a broad definition, we shall be able here to delineate only a few major threads. The production of sound moved in the early nineteenth century from musical instruments, explosions, and the human voice, to the ingenious devices of Helmholtz, into the field of ultrasonics through the work of the Curie brothers, and into underwater sound from the researches of Boyle and Langevin. The attenuation of sound includes the theoretical work of Stokes and Kirchhoff, the experiments of Neklepaev, the further theory of Einstein, and the experiments of Kneser. Nonlinear acoustics began with the theories of Poisson and Riemann, was pushed along by Rayleigh, and developed in this century by Fubini, Lighthill, and Khoklov. The optical effect, discovered by Debye and, independently, by Lucas and Biquard, received its theoretical basis in the work of Raman and Nath, and was confirmed in detail experimentally by Nomoto.

**4aPA2. Selected topics in the history of atmospheric acoustics in North America, 1865–1940.** David T. Blackstock (Mech. Eng. Dept. and Appl. Res. Labs., Univ. of Texas, Austin, TX 78712-1063)

Considered here are three topics in atmospheric acoustics, studied by North American scientists during the 75-year span between the Civil War and World War II. First, in 1865 Joseph Henry (1799–1878), as a member of the Light-House Board, began to investigate fog signaling. His experiments, carried on in friendly competition with John Tyndall in England, added a great deal to our practical knowledge of atmospheric refraction. Second, the first-ever experiments to measure the absorption of sound in the atmosphere were carried out by A. Wilbur Duff (1864–1951) in New Brunswick, Canada, in 1898 and in 1900. The failure of viscosity and heat conduction to account for the measured absorption prompted Lord Rayleigh to postulate that a relaxation mechanism in the air



might be responsible for the difference. Finally, George Washington Pierce (1872–1956), who is perhaps best known for his work on piezoelectric and magnetostrictive devices, used an acoustic interferometer to measure dispersion in gases, particularly carbon dioxide. He also measured sound transmission over reflective surfaces. His results confirm the at-first surprising result that at grazing incidence even a very hard boundary seems to act as a pressure release surface.

THURSDAY MORNING, 4 NOVEMBER 1999

MARION ROOM, 9:00 TO 11:15 A.M.

### Session 4aPP

## Psychological and Physiological Acoustics: Binaural and Sound Field

Raymond H. Dye, Chair

*Parmly Hearing Institute, Loyola University, 6525 North Sheridan Road, Chicago, Illinois 60626*

### Contributed Papers

9:00

**4aPP1. Identification and localization of sound sources in the median sagittal plane.** Brad Rakerd, William M. Hartmann, and Timothy L. McCaskey (Michigan State Univ., East Lansing, MI 48824)

The ability of human listeners to identify broadband noises having different spectral structures was studied for multiple sound-source locations in the median sagittal plane. The purpose of the study was to understand how sound identification is affected by spectral variations caused by directionally dependent head-related transfer functions. It was found that listeners could accurately identify noises with different spectral peaks and valleys when the source location was fixed. Listeners could also identify noises when the source location was roved in the median sagittal plane when the relevant spectral features were at low frequency. Listeners failed to identify noises with roved location when the spectral structure was at high frequency, presumably because the spectral structure was confused with the spectral variations caused by different locations. Parallel experiments on source localization showed that listeners can localize noises that they cannot identify. The combination of identification and localization experiments leads to the conclusion that listeners cannot compensate for directionally dependent filtering by their own heads when they try to identify sounds. [Work supported by the NIDCD.]

9:15

**4aPP2. Contribution of spectra of input signals to two ears to sound localization in the sagittal plane.** Kazuhiro Iida (AVC Res. Lab., Matsushita Communication Ind. Co. Ltd., 600 Saedo, Tsuzuki, Yokohama, 224-8539 Japan, kiida@adl.mci.mei.co.jp), Eigo Rin, Yasuko Kuroki, and Masayuki Morimoto (Kobe Univ., Nada, Kobe, 657-8501 Japan)

The previous studies show that the amplitude spectra of input signals to two ears contribute to sound localization in the sagittal plane [e.g., M. B. Gardner, *J. Acoust. Soc. Am.* **54**, 1489–1495 (1973)]. Furthermore, it is known that the ear on the source side has more influence than the ear on the opposite side on localization [M. Morimoto, Dissertation, Univ. of Tokyo (1982); R. A. Humanski and R. A. Butler, *J. Acoust. Soc. Am.* **83**, 2300–2310 (1988)]. It is, however, not clarified how each ear spectrum contributes to localization. The present paper builds up some hypotheses on this issue and examines them for their validity by some psychoacoustical experiments.

9:30

**4aPP3. The influence of later arriving sounds on the ability of listeners to judge the lateral position of a source.** Raymond H. Dye, Jr. (Parmly Hearing Inst. and Dept. of Psych., Loyola Univ., 6525 N. Sheridan Rd., Chicago, IL 60626)

This investigation focuses on the effect that later sounds have on the ability of humans to report the spatial location of earlier sounds. Two dichotic pulses were presented (via headphones), separated by an echo delay between 4 and 64 ms. Listeners were asked to judge whether the first

click appeared to the left or right of the intracranial midline. The interaural delay of each pulse was independently selected from a Gaussian distribution ( $\mu, \sigma=0, 100 \mu\text{s}$ ). The level of the echo relative to the source was 0, -6, -12, -18, -24, -30, or -36 dB. The effect of the echo was determined by measuring proportion correct and by deriving a normalized source weight. For echo delays of 16, 32, and 64 ms, the source and echo click were weighted equally when presented at the same level, and weights barely changed until the echo was attenuated by 18 dB. As the level of the second click was further reduced, the source weight approached 1.0 and the percentage correct approached that obtained in the “no echo” condition. For shorter echo delays, source weight and proportion correct increased more quickly as the echo was attenuated. [Work supported by NIDCD & AFOSR.]

9:45

**4aPP4. Low-frequency ILD elevation cues.** V. Ralph Algazi, Carlos Avendano (CIPIC, UC Davis, Davis, CA 95616, algazi@ece.ucdavis.edu), and Richard O. Duda (San Jose State Univ., San Jose, CA 95192)

It is well known that the binaural ITD (interaural time difference) and ILD (interaural level difference) are the primary cues for azimuth, while monaural spectral features due to pinna diffraction are the primary cues for elevation. Pinna cues appear above 3 kHz, where the wavelength becomes comparable to pinna size. However, it is shown that there are also important low-frequency ILD elevation cues primarily due to torso diffraction. In the experiments reported, random noise bursts were filtered by individualized head-related transfer functions, and four subjects were asked to report the elevation angle. Eight conditions were tested, depending on whether the source was in front or in back, in the median plane or on a 45-deg cone of confusion, and had wide bandwidth or was band limited to 3 kHz. For the band-limited signal, localization accuracy was at chance level in the median plane, and was poor in front. However, at 45-deg azimuth in the back, the accuracy was close to that for a wideband source, the average correlation coefficient being approximately 0.75 for the narrow-band source and 0.85 for the wideband source. A physical explanation for the cues is presented. [Work supported by NSF under Grant No. IRI-9619339.]

10:00

**4aPP5. The effect of frequency modulation on the ability to judge dynamic changes in interaural level differences.** William M. Whitmer and Raymond H. Dye, Jr. (Parmly Hearing Inst., Loyola Univ., Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, wwhitme@luc.edu)

The cues for apparent auditory motion include dynamic interaural temporal differences, level changes, and frequency modulation (FM) [Rosenblum, *Perception* **16**, 175–186 (1987)]. The manner in which these cues interact was examined in a task in which listeners judged the point at

which a stimulus having dynamic changes in interaural level difference (ILD) passed through the intracranial midline. In base-line conditions, stimuli were 2.0–2.5-s, 0.5- or 2-kHz sinusoids presented through headphones. In experimental conditions, a frequency-modulated tone was added. Frequency was modulated by a linear ramp or according to a Doppler-based algorithm. The midpoint of the frequency-sweep occurred either at the instant that the ILD passed through 0 dB, acting as a task cue, or at some other point. FM rate and magnitude, temporal position of frequency-sweep midpoint, and direction of apparent movement were randomized across trials. Results indicated that the presence of FM had a significant effect on the perception of intracranial midline. This effect was obtained across FM parameter values. Results were considered in terms of how dynamic cues for perceived motion perceptually interact. [Work supported by NIH and AFOSR.]

10:15

**4aPP6. Does precedence prevail through sudden changes in selected partials of a complex natural spectrum?** Miriam N. Valenzuela and Ervin R. Hafter (Dept. of Psychol., Univ. of California, Berkeley, CA 94720, miriam@ear.berkeley.edu)

Localization of a source followed by a delayed version of it from another direction is dominated by the direction of the first wavefront. It has been suggested that this “precedence” effect and its implication of echo suppression is a dynamic process, subject to listeners’ expectations about what a “plausible” echo should be. The present study investigates what makes an echo “plausible” [Rakerd and Hartmann, *J. Acoust. Soc. Am.* **78**, 524–533 (1985); Clifton *et al.*, *J. Acoust. Soc. Am.* **95**, 1525–1533 (1994)]. Synthetic piano tones presented through pairs of speakers in an anechoic room were designed to simulate natural sounds heard in echoic environments. Trials consisted of conditioning stimuli followed by a test stimulus. Each stimulus was made up of a hypothetical source and echo. The spectra of the test echoes could be changed to reflect abrupt changes in the absorption qualities of simulated reflections. As in precedence, conditioning stimuli were localized in the direction of their sources. Attenuation of selected partials in the test-echo did not break down precedence, but when the echo-spectra were modified in less plausible ways by amplification of selected partials, precedence broke down.

10:30

**4aPP7. The relative contribution of onset asynchrony, harmonic ratios and angular separation of sound sources to the cross-spectral grouping of complex tones in a free field.** Martine Turgeon and Albert S. Bregman (Dept. of Psych., McGill Univ., Montreal, QC H3A 1B1, Canada, martine@hebb.psych.mcgill.ca)

The Rhythmic Masking Release paradigm (RMR) was used to study how onset asynchrony, harmonic ratios and speakers separation interact toward the cross-spectral grouping of concurrent complex tones. In RMR, a regular sequence is perceptually masked by some irregularly spaced masking sounds, which are acoustically identical to those of the rhythm. The identification of the rhythm becomes contingent upon the perceptual grouping of the masking sounds with some flanking sounds of different frequencies. The rhythm-detection accuracy in a 2-AFC procedure was used to estimate the degree of perceptual grouping of 48-ms long masking and flanking tones, each composed of four harmonics. For each of 18 listeners, synchrony of onset and offset between the masking and flanking tones yielded an almost perfect rhythm-detection performance, independent of whether or not they shared a common fundamental frequency and of the spatial separation of their sources. Rhythm-detection accuracy

strongly diminished with onset asynchrony and weakly diminished with inharmonicity and speakers separation. Our results suggest that onset asynchrony is the determining factor for the perception of concurrent sounds of a short duration as separate events; inharmonicity and the spatial separation of sources are weak factors, although they appear to reinforce each other toward perceptual segregation.

10:45

**4aPP8. Computational auditory scene analysis-constrained array processing for sound source separation.** Laura A. Drake (Elec. & Computer Eng. Dept., Northwestern Univ., Evanston, IL 60208-3118), Janet C. Rutledge (Univ. of Maryland at Baltimore, Baltimore, MD 21201), and Aggelos Katsaggelos (Northwestern Univ., Evanston, IL 60208-3118)

In this work, techniques are developed and studied for the extraction of single-source acoustic signals out of multi-source signals. Such extracted signals can be used in a variety of applications including: automatic speech recognition, teleconferencing, and robot auditory systems. Most previous approaches fall into two categories: computational auditory scene analysis (CASA) and array signal processing. The approach taken here is to combine these complementary techniques into an integrated one: CASA-constrained array processing. In principle, this integrated approach should provide a performance gain since the information used by array processing (direction of propagation through a sound-field) is independent of other CASA features (fundamental frequency, on/offset, etc.). One difficulty encountered by CASA that can be overcome by array processing is the sequential grouping of spectrally dissimilar phonemes in a speech signal, such as a fricative followed by a vowel. The method presented here differs from standard array processing by the addition of CASA features for the signal separation decision. Compared to other CASA systems that use binaural cues, it: (1) is not limited to two microphones (since the goal is not auditory system modeling); and (2) makes complete use of source location and other CASA features—for simultaneous and sequential grouping.

11:00

**4aPP9. Reducing hearing protector test time with a minimum audible pressure to field transfer function.** William J. Murphy and John R. Franks (Bioacoust. and Occupational Vib. Section, NIOSH, MS C-27, 4676 Columbia Pkwy., Cincinnati, OH 45226-1998)

Effective hearing-loss prevention programs should train workers to properly fit hearing protection devices. In order to minimize the training time for protector fitting, unoccluded minimum audible field threshold (MAF) for narrow-band noise could be estimated from the worker’s annual minimum audible pressure (MAP) audiogram. Murphy *et al.* previously reported a transfer function for 5- and 1-dB audiometry for 75 and 10 subjects, respectively [W. J. Murphy, J. R. Franks, S. L. Hall, and E. F. Krieg, *J. Acoust. Soc. Am.* **101**, 3126 (1997)]. Significant differences at several frequencies between the 1- and 5-dB groups prompted further study of the transfer function. This paper reports MAF and MAP results for 44 subjects tested with both 1- and 5-dB audiometry and 1-dB sound field. Pure-tone thresholds were significantly different, for 1- and 5-dB step sizes. The binaural reference equivalent threshold sound pressure level (RETSPL) sound-field thresholds were compared with the RETSPL thresholds for monaural listening in a sound-field listed in ANSI S3.6-1996, Table 9. The binaural thresholds were found to be significantly higher ( $p < 0.05$ ) than the standard values at all frequencies except 8000 Hz.

## Session 4aSCa

## Speech Communication: Physiology and Modeling of Voice

Jody E. Kreiman, Chair

*Bureau of Glottal Affairs, Head and Neck Surgery, UCLA School of Medicine, Los Angeles, California 90095*

## Contributed Papers

8:00

**4aSCa1. Measurement of vocal fold depth and thickness in frozen and thawed canine larynges.** Niro Tayama (Natl. Ctr. for Voice and Speech, Dept. of Speech Pathol. and Audiol., The Univ. of Iowa and Dept. of Otolaryngol., Univ. of Tokyo, Tokyo, Japan, ntayama@dolce.shc.uiowa.edu), Roger Chan (The Univ. of Iowa), Kimitaka Kaga (Univ. of Tokyo, Tokyo, Japan), and Ingo Titze (The Univ. of Iowa, Iowa City, IA 52242)

Anatomical data on vocal fold dimensions are necessary for defining the vocal fold boundaries in biomechanical modeling of vocal fold vibration. In the mid-membranous coronal section, vocal fold depth can be defined as the distance from the vocal fold medial surface to the thyroid cartilage, whereas thickness can be defined as the distance from the inferior border of the thyroarytenoid muscle to the vocal fold superior surface. Unfortunately, reliable geometric data from histological sections can be obtained only if the effects of sample preparation are quantified. For instance, tissue deformations are often induced by fixation and dehydration, sometimes producing shrinkages around 30%. In this study, reliable geometric data of the canine vocal fold were obtained by comparing frozen and thawed larynges. Coronal sections of frozen larynges were thawed gradually in saline solution. Images of the mid-membranous coronal sections at various thawing stages were captured by a digital camera. Measurements of vocal fold depth and thickness were made using a graphic software package (NIH image). Results showed that geometric changes of the vocal fold induced by freezing were likely reversed by thawing, such that the vocal fold depth and thickness measured on thawed larynges were representative of the pre-freezing state. [Work supported by NIH Grant No. P60 DC00976.]

8:15

**4aSCa2. Vocal rise time and perception of a hard glottal attack.** Rahul Shrivastav (Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

Three modes of voice onset are commonly recognized—breathy, normal and hard [Moore (1938)]. These differ in terms of the pulmonary airflow and the degree of medial compression force associated with vocal fold adduction [Koike *et al.* (1967)]. Perceptually, this is characterized by a “transient rise and fall in pitch and loudness at the onset of voicing” [Orlikoff and Kahane (1996)]. Perceptual identification of hard glottal attacks (HGA) is an important part of assessment and management of voice disorders. The relation between frequency and intensity rise times and perception of HGA was studied using a continuum of synthesized monosyllabic words varying in the frequency rise times. These tokens were recorded from a normal talker, modified and re-synthesized using STRAIGHT [Kawahara (1997)]. Five listeners were asked to rate the perceived strength of the HGA on a five-point scale. Preliminary results with tokens varying in frequency rise time suggest an increase in the perceived strength of HGA when the frequency rise time increases from 20 ms to 50 ms. The perceived strength of HGA appears to plateau beyond rise times of 50 ms.

8:30

**4aSCa3. Pressure profiles on the walls of the glottis for oblique glottal ducts.** Ronald C. Scherer (Dept. of Commun. Disord., Bowling Green State Univ., Bowling Green, OH 43403) and Daoud Shinwari (Univ. of Toledo, Toledo, OH 43607)

Oblique angles of the glottis occur during both normal and abnormal phonation. This study examined pressures on the glottal walls for a variety of oblique angle conditions. A Plexiglas model of the larynx was used to obtain wall pressures at 14 taps along the glottal surfaces. The model is 7.5 times larger than real life. A minimal diameter of 0.04 cm and duct axial length of 0.3 cm were used for all cases. Results indicated, for example, that for an included angle of 10 deg (divergence) and a transglottal pressure of 5-cm H<sub>2</sub>O, an oblique glottal angle of 15 deg compared to the symmetric glottis: (1) displaced the minimal wall pressure location on the divergent side a short distance downstream and on the convergent side to mid-glottis; (2) increased the pressure drop on the divergent side by about 15% and decreased the pressure drop on the convergent side by about 13%; and (3) broadened the minimum pressure dip on both the divergent and convergent sides. These results suggest paradoxical (opposite) driving pressures on the inferior glottal walls, which would enhance the nontypical phase differences between the two vocal folds. [Research support: NIH grant 1R01DC03577.]

8:45

**4aSCa4. Estimation of minimum glottal flow using optimal, low-pass filtering.** Yingyong Qi (P.O. Box 210071, Tucson, AZ 85721, yqi@u.arizona.edu) and Robert E. Hillman (243 Charles St., Boston, MA 02114)

The amount of minimum glottal flow in each period of a sustained phonation is an important parameter in voice research and clinic. In many cases, it is highly desirable to be able to measure the minimum glottal flow automatically. Here, we present a method for estimating the minimum glottal flow using an optimal, low-pass filter. The cutoff frequency of the low-pass filter is determined so that the sum of the variance within each “closed” phase of a recorded flow signal and the difference between the recorded and filtered flow signals is minimal. The minimum glottal flow is derived from this optimally, low-pass-filtered signal. This simple optimization procedure results in a complete automatic estimation of minimum glottal flow. Experiments using synthetic and real flow signals indicated that the method is highly accurate and robust.

9:00

**4aSCa5. Modeling tremulous voices.** Jody Kreiman, Brian Gabelman, and Bruce R. Gerratt (Bureau of Glottal Affairs, Div. of Head/Neck Surgery, UCLA School of Medicine, Los Angeles, CA 90095)

Vocal tremors prominently characterize many pathological voices, but acoustic-perceptual aspects of tremor are poorly understood. To investigate this relationship, two tremor models were implemented in a custom voice synthesizer. The first modulated  $F_0$  with a sine wave. The second provided “random” modulation based on white-noise low-pass filtered with cutoff frequency equal to the nominal modulation rate (usually 1–10 Hz). Control parameters in both models were the rate and extent of  $F_0$  modulation. (Amplitude modulation was not separately modeled.) Thirty

randomly selected 1-s samples of /a/ were modeled. Two synthetic versions of each vowel were created: one with tremor parameters derived from plots of  $F_0$  versus time, and one with parameters chosen to match the original stimulus auditorily, regardless of measured tremor values. Listeners judged the similarity of each synthetic voice to the original stimulus. Sine wave and random tremor models both provided excellent matches to subsets of the voices. Preliminary results suggest that tremor rates could be satisfactorily derived from  $F_0$  plots, but tremor deviations derived from these plots sounded too extreme. These results also demonstrate the advantage of this perceptual analysis-by-synthesis method in integrating acoustic and perceptual approaches to measuring voice quality.

9:15

**4aSCa6. Vocal fundamental frequency drop during VCV productions in Japanese speakers.** Makoto Kariyasu and John Michel (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, Lawrence, KS 66045-2181, kariyasu@kuhub.cc.ukans.edu)

Vocal fundamental frequency ( $f_0$ ) drop in relation to changes in intraoral pressure ( $P_o$ ) was investigated for adult male and female speakers of Japanese producing /b/ and /v/ in VCV contexts. Three experiments were conducted to examine the following effects on  $f_0$  drop: (1) vocal tract constriction, (2) vowel context, and (3) speech rate. The results showed that the  $f_0$  drop from vowels to a consonant was greater for /aba/ than for /ava/; it was greater for high vowel contexts, /ibi/, /abi/, /iba/, than for a low vowel context, /aba/. Speech rate changes with metronome and self-paced had variable effects on the  $f_0$  drop. Overall, the  $f_0$  drop for /aba/ ranged from 4% to 13% or from 0.75 to 2.17 semitones. Comparisons with English speakers will be reported. Effects of gender, vocal pitch, and intensity levels, and accent may be discussed.

9:30

**4aSCa7. Finite element analysis of vocal fold posturing.** Eric J. Hunter (Dept. of Speech Pathol. and Audiol., Univ. of Iowa, Iowa City, IA 52246, eric-hunter@uiowa.edu)

Steady-state vocal fold posturing was simulated with finite-element analysis using three-dimensional elastic elements. The thyroarytenoid muscle was simulated by applying nodal forces along the fiber direction as measured in our laboratory. Other intrinsic laryngeal muscles and cartilage were defined as boundary conditions or applied loads. Because the arytenoid cartilage forms a complex boundary geometry, the arytenoid was modeled as part of the soft tissue, but with much stiffer, nearly rigid body, material constants. The finite-element model is based on a small displacement vibrational model of the vocal folds [Titze and Scherer, *Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control*, pp. 183–190] and a large displacement model of the tongue [Wilhelms-Tricarico, *J. Acoust. Soc. Am.* **97**, 3085–3098 (1994)]. The medial shape of the modeled vocal folds was compared to published histological data.

9:45

**4aSCa8. On the mechanism of lowering the voice fundamental frequency in the production of tones of Standard Chinese.** Hiroya Fujisaki (Dept. of Appl. Electron., Sci. Univ. of Tokyo, 2641 Yamazaki, Noda, 278-8510 Japan, fujisaki@te.noda.sut.ac.jp)

While it is well known that the cricothyroid (CT) muscle is mainly responsible for lowering the voice fundamental frequency ( $F_0$ ) in many languages, the mechanism for  $F_0$  lowering in languages such as Standard Chinese (SC) has not been elucidated. Although several studies have shown that the sternohyoid (SH) muscle activity is strongly correlated with  $F_0$  lowering in SC, the mechanism itself is not clear since SH is not

directly attached to the thyroid cartilage, whose movement is essential in changing the length and tension of the vocal chord. On the basis of an earlier finding on the production of Thai tones [D. Erickson, *Annual Bulletin No. 27, RILP, Univ. Tokyo*, pp. 135–149 (1993)], the present author has suggested the active role of the thyrohyoid (TH) muscle in  $F_0$  lowering in languages such as SC and Swedish [H. Fujisaki, *Proc. XXIII World Congress of the International Association of Logopedics and Phoniatrics*, pp. 156–159 (1995)]. The present study shows the detailed mechanism for  $F_0$  lowering involving both TH and SH in the production of the second, third, and fourth tones of SC based on new electromyographic observations.

10:00

**4aSCa9. Using time-temperature superposition to estimate the elastic shear modulus and dynamic viscosity of human vocal fold tissues at frequencies of phonation.** Roger Chan and Ingo Titze (Natl. Ctr. for Voice and Speech, Dept. of Speech Pathol. and Audiol., The Univ. of Iowa, Iowa City, IA 52242, roger-chan@uiowa.edu)

The principle of time-temperature superposition [J. D. Ferry, *Viscoelastic Properties of Polymers* (Wiley, New York, 1980), pp. 264–320] has been widely used by rheologists to estimate the viscoelastic properties of polymeric materials at time or frequency scales not readily accessible experimentally. This principle is based on the observation that for many polymeric systems molecular configurational changes that occur in a given time scale at a low temperature correspond to those that occur in a shorter time scale at a higher temperature. Thus the viscoelastic properties of these systems empirically measured at various temperatures at a certain frequency are indicative of measurements at different frequencies at a single reference temperature. Using a rotational rheometer, the elastic shear modulus and dynamic viscosity of human vocal fold mucosal tissues were measured at relatively low temperatures (5–25 °C) at 0.01–15 Hz. Data were empirically shifted based on this principle to yield a composite “master curve” which gives a prediction of the shear properties at higher frequencies at 37 °C. Results showed that the time-temperature superposition principle may be used to estimate the viscoelastic shear properties of vocal fold tissues at frequencies of vocal fold vibration, on the order of 100 Hz. [Work supported by NIH Grant No. P60 DC00976.]

10:15

**4aSCa10. Effects of vocal tract on aerodynamics of hemilarynx.** Fariborz Alipour, Douglas Montequin, and Niro Tayama (Dept. of Speech Pathol. and Audiol., The Univ. of Iowa, Iowa City, IA 52242)

Pressure-flow relationship was examined in the excised canine and human larynges with and without vocal tract. Canine and human larynges were prepared and cut in the midsagittal plane from the top to about 10 mm below the vocal folds. The right half was removed and replaced with a Plexiglas plate with imbedded pressure taps along the medial surface. The thyroid cartilage was glued to the plate and the arytenoid was pressed against the plate with a two-pronged probe for adduction control. The vocal tract was simulated with a 15-cm plastic tube of 25-mm diameter. Simultaneous recordings were made of the glottal pressure, mean subglottal pressure, and average airflow at various levels of adduction. Glottal adduction was controlled mechanically by inserting shims of various sizes. Oscillation was generated by the flow of heated and humidified air through the glottis. Preliminary data indicate that the pressure-flow relationships are similar to those of full larynx and are almost linear. The addition of the vocal tract increased the glottal resistance by moving these pressure-flow lines to the lower flow and higher-pressure region. The human larynx appears to phonate easier on the bench and has lower phonation threshold pressure. [Work supported by NIDCD grant DC03566.]



## Session 4aSCb

## Speech Communication: Potpourri (Poster Session)

Mary E. Beckman, Chair

Department of Linguistics, The Ohio State University, Columbus, Ohio 43210

## Contributed Papers

To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon. To allow for extended viewing time, posters will be on display from 9:00 a.m. to 10:00 p.m.

**4aSCb1. On estimating length and shape of the vocal tract.** Edward P. Neuburg (IDA-CCRP, Thanet Rd., Princeton, NJ 08540, epn@idaccr.org)

Automatic vowel recognition is sometimes done by trying to estimate vocal-tract areas, usually derived using linear predictive coding. In the derivation there is a free parameter, namely, the length of the vocal tract; every choice of length leads to a different set of areas. This paper describes a revisit to experiments done in the early 1970s, in which the chosen length is the one that makes the vocal-tract shape most "human." Using some improved criteria for optimality of shape, involving both smoothness and small variance, it is found that: (a) estimates of vocal-tract length are consistent, or at least highly correlated, with actual length; and (b) the tube shape (area function) derived using the optimal length seems to give a reliable characterization of the vowel, independent of the talker (and, in particular, independent of the talker's gender). Classification results are about the same as those using the currently popular cepstral coefficients, but areas have the possible advantage that they are physiological features that the talker can actually control.

**4aSCb2. Principal components analysis of x-ray microbeam pellet positions.** Robert E. Beaudoin and Richard S. McGowan (Sensimetrics Corp., 48 Grove St., Somerville, MA 02144)

Sets of pellet coordinates from the X-ray Microbeam Speech Production Database, each corresponding to a static articulatory configuration, are submitted to a principal components analysis. Several talkers are analyzed separately, and various criteria are used to select the times in the continuous speech data from which to extract pellet data. In some instances, all of a talker's speech data are submitted to an automatic procedure intended to select syllable nuclei, based on Mermelstein's algorithm [P. Mermelstein, *J. Acoust. Soc. Am.* **58**, 880–883 (1975)]. In other instances specific data are chosen by hand according to utterance type (e.g., vowel, glide, and consonant–vowel transition). These analyses are intended for use in a project to recover articulation from speech acoustics. In particular, we examine the relationship between pellet positions from data, the tongue shape reconstructed from the principal components analysis, and the parameters of a simple acoustic tube model, which is derived from the Stevens and House model [K. N. Stevens and A. S. House, *J. Acoust. Soc. Am.* **27**, 484–493 (1955)]. [Work supported by Grant NIDCD-01247 to Sensimetrics Corporation.]

**4aSCb3. Articulatory and acoustic characteristics of emphasized and unemphasized vowels.** Donna Erickson (Natl. Inst. of Multimedia Education, 2-12 Wakaba, Mihama, Chiba, 261-0014 Japan), Osamu Fujimura (The Ohio State Univ., Columbus, OH 43221), and Jianwu Dang (ATR Human Information Processing Res. Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan)

This study examines the relation between formants, jaw  $x-y$  position and tongue dorsum  $x-y$  position for emphasized and unemphasized vowels (/ay/ as in "high," /iy/ as in "he," and /eh/ as in "head"). Acoustic and articulatory measurements were made for target syllables at the moment of jaw height minimum, comparing emphasized and unemphasized conditions of the same vowel. The results can be interpreted as a hyper-articulation of both jaw and tongue movement for emphasis. When emphasized, mandibular position shows increased lowering with forward movement along the front–back axis of the occlusal plane, for all three vowels. The tongue dorsum shows increased raising and fronting for /iy/ and /eh/ but increased lowering and backing for /ay/. The associated  $F1-F2$  show a corresponding spreading apart of  $F1-F2$  for the higher vowels, and a bunching of  $F1-F2$  for the low vowel consistent with the biomechanics of tongue/jaw articulation. These results are interpreted in light of the C/D model which assumes that prosody affects syllable magnitude, which in turn extrapolates the deviation from the neutral gesture along with more opening of the jaw. The C/D model also has the potential to explain timing discrepancies among the extrema as observed in different articulators.

**4aSCb4. Respiratory system changes in relation to prosodic cues at the beginning of speech.** Janet Slifka (50 Vassar St., Rm. 36-549 RLE, Cambridge, MA 02139, slifka@mit.edu)

The research presented here examines regional changes in respiratory drive and respiratory-influenced laryngeal adjustments in relation to acoustic signal changes associated with prosody. The respiratory system acts to create subglottal pressure ( $P_{SG}$ ). When  $P_{SG}$  is changing rapidly, such as at the initiation of an utterance at the start of an exhalation, timing and amplitude constraints may be present. The actions of the articulators and the chest wall must be coordinated to rapidly reverse a negative  $P_{SG}$  during inhalation to a  $P_{SG}$  in the speech range. The current study has involved collection of simultaneous recordings of the acoustic signal and several physiologically related signals including subglottal pressure as estimated from esophageal pressure. Test utterances either begin with a stressed syllable or have one unstressed syllable prior to the first stressed syllable. Present results indicate that: (1)  $P_{SG}$  continues to rise from the initiation of an exhale until the first stressed syllable; (2)  $P_{SG}$  falls following the initial peak; and (3) fundamental frequency rise is not directly related to the rise in  $P_{SG}$ . Applications such as articulatory-related speech

synthesis may benefit from such information. [This research was supported in part by NIH grants #5T32DC00038 and #5R01-DC00266-14.]

**4aSCb5. Kinematic evidence for the existence of gradient speech errors.** Marianne Pouplier, Larissa Chen (Dept. of Linguist., Yale Univ., New Haven, CT 06511), Louis Goldstein (Haskins Labs., New Haven, CT 06511-6695), and Dani Byrd (Univ. of Southern California, Los Angeles, CA 90088-1693)

Speech errors have long been appealed to as evidence for segmental units in speech production. However, if speech errors can be shown to be potentially gradient, as opposed to always categorical, in nature, the view of segments such as “swapping places” must be called into question. Severe limitation of a transcription approach to evaluating speech errors is that gradient errors may go unrepresented in the transcriptional record if they are obscured by other articulatory events due to the coproduction of speech gestures. In a pioneering production study of errors, Mowrey and MacKay [J. Acoust. Soc. Am. **88**, 1299–1312 (1990)] present EMG data that suggest that gradient errors do exist, although they are not always audible. However, the anomalous activity they observe in single-motor-unit recordings does not preclude the possibility that a segmental unit was either still produced correctly via compensation by other muscle activity or omitted in its entirety in the articulatory kinematics. Using magnetometry, the present study for the first time is able to provide articulatory movement tracking data that exhibit gradient errors in phrases like “cop top” and “Bligh Bay” repeated rapidly. For example, during the [t] in “cop top” some tokens show a small (sometimes inaudible) raising of the tongue dorsum. [Work supported by NIH.]

**4aSCb6. Articulatory and acoustic data for vowels in isolation and in an /sVd/ context: An x-ray microbeam study.** Angela Slama, Gary Weismer, and Kate Bunton (Dept. of Communicative Disord. and Waisman Ctr., Goodnight Hall, Univ. of Wisconsin-Madison, Madison, WI 53706, weismer@waisman.wisc.edu)

Interspeaker variability in articulatory processes has been of great interest in the past decade, but studies addressing this issue have typically been limited to a relatively small group of speakers. In the present study we report on articulatory coordinates and formant frequencies for corner vowels spoken by 50 young adults. The vowels were produced in isolation, and in an /sVd/ context. Articulatory coordinates for the upper and lower lips and four tongue pellets were obtained at temporal midpoint of the vowels using the x-ray microbeam database. Formant frequencies, measured at the same temporal location, were derived from a well-known speech analysis package. Representations of both kinematic and acoustic interspeaker variability will be described and discussed, and relations between the kinematic and acoustic working space for vowels will be examined. A goal of this project is to develop some standards for kinematic/acoustic representations for vowels that can be applied to speakers with dysarthria. [Work supported by NIH DC00820.]

**4aSCb7. Kinematics of compensatory vowel shortening: Intra and interarticulatory timing.** Susan Shaiman (Dept. of Commun. Sci. and Disord., Univ. of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260)

The acoustic shortening of vowels has been demonstrated to occur across a variety of contextual variations. The current study examined the kinematic adjustments involved in vowel shortening, as a function of speaking rate and coda composition (singleton consonants versus consonant clusters). Five normal speakers repeated the syllables /paep/, /paeps/, and /paepst/, embedded in a carrier phrase, across three distinct speaking rates (slow, normal, and fast). Changes in the timing of the jaw-closing gesture for post-vocalic bilabial production, and in the upper lip–jaw timing relationship (measured via phase angles), were examined. The onset of the jaw-closing gesture typically shifted earlier in the cycle of jaw movement for consonant cluster productions, across all speaking rates. Subject-

specific modifications of interarticulatory timing were observed, with some speakers adjusting upper lip movement in order to maintain constant timing with the jaw, while others tended to dissociate the upper lip and jaw. Both coda composition and speaking-rate manipulations resulted in substantial intersubject variability in the lip–jaw timing relationship. These findings suggest that, in order to achieve the intended acoustic-perceptual goals, articulatory coordination may not be absolutely invariant, but rather, systematically and individually organized across task manipulations. [Work supported by CRDF-University of Pittsburgh and NSERC.]

**4aSCb8. Using tone similarity judgments in tests of intertranscriber reliability.** Julia McGory (Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210), Rebecca Herman (Indiana Univ.), and Ann Syrdal (AT&T Labs.–Res.)

An important, albeit often neglected, aspect of intonational labeling is testing the consistency between labelers. When reliability among different labelers has been tested, two aspects in which transcribers differ from each other is in placement and choice of tone labels. Only identical labels count as “agreement” between transcriptions. This can be problematic given that some tone categories are perceived to be similar, making the choice between labels at times difficult. This work suggests a refinement of existing tests of intertranscriber reliability. Previous intertranscriber reliability tests did not take into account each labeler’s perceived similarity between pitch accent categories. We examined a set of ToBI transcriptions of a single speaker corpus labeled independently by four experienced intonation labelers. The proposed method incorporates each transcriber’s tone similarity judgments, which were elicited for the purposes of this experiment. The labelers judged the similarity of pitch accent pairs using a rating scale of 1–7. Multidimensional scaling (MDS) of similarity judgments resulted in quantifiable measures of degree of similarity among tones. Euclidean distances between pairs of tones were calculated from the MDS results. We used these values as the weights in calculating mean weighted agreements. [This work is supported by AT&T Labs.–Research.]

**4aSCb9. Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate.** J. Sean Allen and Joanne L. Miller (Dept. of Psych., 125 NI, Northeastern Univ., Boston, MA 02115)

A series of experiments examined the effects of an acoustic-phonetic contextual factor, speaking rate, and a higher-order linguistic contextual factor, lexical status, on the internal structure of a voicing category, specified by voice-onset-time (VOT). In keeping with previous results, speaking rate fundamentally altered the structure of the voiceless category, not only affecting the perception of stimuli in the voiced–voiceless category boundary region but also altering which tokens were rated as the best exemplars of the voiceless category. In contrast, the effect of lexical status was more limited. Although (as expected) lexical status also affected the perception of stimuli in the category boundary region, this effect disappeared in the region of the best-rated exemplars. This distinction between the effects of speaking rate and lexical status on the internal structure of the voiceless category mirrors the effects of these factors in speech production: It is well known that speaking rate alters the VOT values of voiceless consonants, whereas we confirmed in a speech production experiment that lexical status has no such effect. Based on these findings, we argue that higher-order contextual factors such as lexical status operate differentially in processing from production-based, acoustic-phonetic factors such as speaking rate. [Work supported by NIH/NIDCD.]

**4aSCb10. Acoustic and articulatory differences across word position for American English /r/.** Suzanne E. Boyce (Univ. of Cincinnati, Mail Location 379, Cincinnati, OH 45221) and Carol Y. Espy-Wilson (Dept. of Elec. and Computer Eng., Boston Univ., Boston, MA 02216)

Typically, American English /r/ is manifested with different formant frequency patterns in word-initial, word-final and word-medial position. In particular, it has been reported (Lehiste, 1962; Espy-Wilson, 1992) that initial /r/'s show lower third formants. This may be due to differences in position or configuration of the primary articulator for /r/—the tongue—or it may reflect an increased degree of lip-rounding. Differences in tongue shape across word position have been reported (Zawadzki and Kuehn, 1982), but it is not clear how reported differences may relate to formant lowering. In this study, eight native speakers of American English recorded words with initial, final and medial /r/'s. Articulatory movements of the tongue tip, tongue blade and tongue dorsum, lips and jaw were tracked simultaneously by Electro-Magnetic Midsagittal Articulometer (EMMA). For each speaker, we compare spatial position, time course, and trajectory for articulators across positional variants. Preliminary data suggest that tongue articulators show similar time course in all three positions. Results are discussed with regard to acoustical modeling studies, theories of articulatory strengthening and interarticulator timing. [Research supported by NIH and NSF.]

**4aSCb11. Proportional and linear constancy in a repetitive speech production task.** Kenneth de Jong (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405)

Speech production studies have shown local temporal stabilities in intergestural timing and in its acoustic consequences [e.g., Kent and Moll, JSHR (1975)], such that the absolute duration of certain speech events is remarkably constant over differences in speech rate. Other production studies examining longer stretches of speech have found event durations to be proportionally constant with respect to a larger frame. Proportional constancy is especially apparent in repetitive production tasks [Cummins and Port, J. Phon. (1998)]. This paper examines how absolute constancy, such as would be specified by segmental contrasts, and proportional constancy interact in a repetitive speech task. Speakers repeated linguistically specified syllables in time to a metronome which specified repetition rate. Rates were varied by a factor of 2.7. Durations of speech events which indicate segmental contrasts, such as voice onset time in onset stops, remain fairly constant over these large changes in speech rate. Other durations, such as vowel durations in open syllables, often show proportional constancy. Vowel durations before coda stops, where duration acts as a secondary cue to consonant voicing, often exhibit linear constancy, suggesting a difference in the degree to which linear linguistic factors restrict how speakers perform a repetitive task.

**4aSCb12. Commonality in the dorsal articulations of English vowels and liquids.** Bryan Gick, A. Min Kang, and D. H. Whalen (Haskins Labs., 270 Crown St., New Haven, CT 06511)

Phonological studies have predicted that the dorsal articulations of English /r/ and /l/ correspond with those of schwa and open o, respectively [Gick, Phonology (in press)]. Specifically, /r/ and schwa are hypothesized to share pharyngeal configuration, while /l/ and open o share upper pharyngeal/uvular configuration. To test this prediction, midsagittal MRI images of the vocal tract of a male speaker of American English were collected and midsagittal distance (of airspace above the tongue surface) measured at 44 3-mm intervals along the vocal tract length. Regions of the vocal tract were defined as pharyngeal, uvular and oral, as in Whalen *et al.* [JSLHR (in press)], with the pharyngeal region divided into upper and lower halves. Midsagittal distances were collected for eleven sustained vowels plus /r/ and /l/. Distances for average vowels were subtracted point by point from /r/ and /l/ and a single rms calculated within each region of the vocal tract. As predicted, in the upper pharyngeal and upper pharyngeal/uvular regions, /l/ showed the greatest correspondence with

open o, while /r/ was most similar to schwa throughout the pharynx. These results support the phonological interpretation of the dorsal gestures of English liquids as vocalic. [Work supported by NIH grant DC-02717.]

**4aSCb13. Acoustic modification in English place assimilation.** David W. Gow, Jr. (Neuropsychol. Lab., VBK 821, Massachusetts General Hospital, 55 Fruit St., Boston, MA 02114, gow@helix.mgh.harvard.edu) and Peter Hussami (MIT, Cambridge, MA 02139)

Recent behavioral research suggests that listeners hearing words containing segments that have undergone place assimilation are able to recover the underlying form of the modified segment and anticipate the place of the segment that triggers assimilation. The present study contrasts acoustic place cues in unmodified coronals, assimilated underlying coronals, and underlying noncoronals in connected read speech in an attempt to characterize the nature of acoustic modification produced by place assimilation, and to understand how a single segment might encode the places of articulation of two segments. Adult male and female speakers produced triplets of sentences showing this three-way contrast across a variety of consonants and vowel contexts. Acoustic measures examined formant transitions and relative amplitudes. The implications of the results of these analyses for the structure of acoustic categories for place information and the abstract representation of place are discussed.

**4aSCb14. An interactive atlas of English vowels: Design considerations.** Robert Hagiwara, Sharon Hargus, Richard Wright (Dept. of Linguist., Univ. of Washington, Box 354340, Seattle, WA 98195), and Isaac Sterling (Univ. of Washington, Seattle, WA 98195)

Discussions of vowel quality differences in dialects of English are often based on imprecise, subjective descriptors (more open, slightly less back) and transcriptions rather than instrumental measurements. Available instrumental data focus on individual dialects rather than a broader picture of dialectal variation. In part to address these concerns, a database of vowel formant frequencies is currently being developed at the University of Washington, beginning with a sampling of static formant frequencies in the American West. The ultimate intent is to produce a web-based, expandable database, or atlas, of reference English vowel formant frequencies and audio examples. It is intended primarily as a pedagogical tool to accurately demonstrate vowel quality variation, and as a "jumping off point" for more intensive studies of acoustic variation. This poster discusses long- and short-term goals of the project, current data, the standardized methodology being used, web tools being developed, and the possibilities for future collaboration in data collection and analysis.

**4aSCb15. Variations in temporal patterns of speech production among speakers of English.** Bruce L. Smith (Dept. of Commun. Sci. and Disord., Northwestern Univ., Evanston, IL 60208-3570)

Results from most studies concerning temporal properties of English are typically presented as group averages. Although various trends emerge from such analyses, relatively little is known about the extent to which individual speakers actually exhibit these tendencies. While some amount of variation among subjects is to be expected, it is also important to determine how regularly individual speakers manifest such temporal patterns. The present study examined several hundred productions by each of 10 native, adult speakers of English to assess the extent to which they showed various temporal characteristics. While a majority of subjects demonstrated, in varying degrees, most of the temporal patterns considered, some did not. For example, most speakers exhibited phrase-final vowel lengthening, averaging approximately a 20%–25% increase for the group across several different words; however, several speakers either did not show phrase-final vowel lengthening for any stimuli, or they showed it for certain words but not others. In addition, the range of performance among speakers who did manifest phrase-final vowel lengthening varied



from as little as 5% increases for some stimuli to more than double for others. Implications of these findings for development, second language acquisition, etc., will be discussed.

**4aSCb16. Place of articulation for consonants: Comparing nasals and stops, and syllable position.** Kenneth N. Stevens, Sharon Y. Manuel, and Melanie L. Matthies (Res. Lab. of Electron. and Dept. EECS, MIT, Cambridge, MA 02139, [stevens@speech.mit.edu](mailto:stevens@speech.mit.edu))

This study reports on measurements of transitions of the second formant ( $F_2$ ) for syllable-initial and syllable-final nasal and stop consonants in English, as cues to place of articulation.  $F_2$  values and the amount and direction of  $F_2$  changes in a 20 ms interval were determined in vowels adjacent to consonant “implosion” and release. The corpus included consonants in sentences and isolated nonsense syllables. The  $F_2$  transitions for a given place of articulation are roughly similar for nasals and stops for different syllable positions, as expected, but some interesting differences exist. Since the stop-consonant bursts influence the spectrum sampling point relative to release, there are some shifts in  $F_2$  onset frequencies for stops, as compared to nasals. Syllable-initial alveolars show additional manner differences, which can be attributed to differences in tongue configurations for alveolar nasals relative to stops in back-vowel contexts. These latter shifts are tentatively ascribed to differential coarticulatory effects for /n/ versus /d/ and /t/ differences which might be related to the lack of a distinctive syllable-initial velar nasal in English. There are also systematic shifts in  $F_2$  transitions as a function of syllable position, for both nasals and stops. [Supported in part by NIH Grants DC00075 and DC02525.]

**4aSCb17. Effects of speech rhythm on timing accuracy of stressed syllable beats.** Robert F. Port, Mafuyu Kitahara, Kenneth de Jong, David R. Collins, and Deborah Felkins Burleson (Linguistics Dept., Indiana Univ., Bloomington, IN 47405)

It has been shown that English speakers have a strong tendency to locate stressed syllable onsets (that is, “beats” or “P centers”) near harmonic fractions (like 1/2, 1/3, 2/3) of the repetition cycle of a repeated phrase [Cummins and Port, J. Phonetics (1999)]. These results were interpreted as evidence of a role for frequencies at multiples of the repetition cycle creating temporal attractors for stressed syllable onsets at harmonic fractions of that cycle. If harmonic oscillators account for temporal attractors, then, having an attractor at 2/3 should necessarily imply the existence of an attractor at 1/3, whereas an attractor at 1/2 would not similarly imply an attractor at 1/4 (since a third oscillator would be required). Thus, if a phrase like *take a pack of cards* is repeated with *card* onset at phase 1/2, there should be a weaker attractor controlling the timing accuracy for *pack* onset than if *card* onset is at 2/3. Four speakers were tested using various text materials and speaking rates. Preliminary results support the hypothesis predicting greater temporal variability for *pack* with *card* beginning at 1/2 than at 2/3.

**4aSCb18. Evidence from /k/ versus /t/ burst spectra for variable lingual contact precision in normal versus atypical phonological development.** Jan Edwards, Marios Fourakis (Speech and Hearing Sci., Ohio State Univ., Columbus, OH 43210-1002, [edwards.212@osu.edu](mailto:edwards.212@osu.edu)), Mary E. Beckman, Pauline Welby, and Satoko Katagiri (Ohio State Univ., Columbus, OH 43210-1298)

Almost universally, phoneme inventories include at least one pair of lingual consonants contrasting an apical “front” contact to a dorsal “back” contact. Studies of infant productions across languages suggest that control of this contrast is acquired later than that between lingual and labial contact, and “velar fronting” (apparent substitution of /t/, /d/ for /k/, /g/) is a commonly observed pattern of functional misarticulation in English-acquiring children. This study uses spectral moments analysis of the burst to examine the patterns of emerging differentiation of alveolar

and velar voiceless stops in eight pairs of preschool children with phonological disorder and normal age peers. Productions of four adults showed consistently higher skewness values and lower center of gravity for /k/ as compared to /t/. The children showed considerable intersubject variability, with some exaggerating the adult distinction and others showing substantial overlap. There was especially extreme overlap in the productions by children with “velar fronting.” However, the spectral moments values for these children did not suggest a categorical substitution. Instead, values were in the middle of the distribution on both dimensions, as if the child were making an “undifferentiated lingual gesture” [F. E. Gibbon, J. Speech. Hear. Language Res. 42, 382–397 (1999)]. [Work funded by NIDCD No. DC02932.]

**4aSCb19. Attention and laterality preferences in children who clutter.** Gordon W. Blood, Ingrid M. Blood, and Glen W. Tellis (Dept. of Commun. Disord., Penn State Univ., University Park, PA 16802, [F2X@psu.edu](mailto:F2X@psu.edu))

Children who clutter were compared to children who stutter and children who did not stutter on a synthetically generated dichotic consonant-vowel listening task and a timed response assessment task of attention. The Dichotic Listening Test consisted of 120 pairs of synthetically generated stop consonants. Subjects used a pointing response and tasks involved three listening conditions (free recall, directed left, directed right). The Conners’ Continuous Performance Test was used to obtain an overall attentional index score, number of correct responses, percentage of omissions, percentage of commissions, hit reaction time, and beta and sensitivity scores. Results revealed that during the directed right and directed left listening tasks children who cluttered and control subjects showed right and left ear advantages. Children who stuttered demonstrated mixed laterality on the DLT. It appeared that children who cluttered benefitted directly from a task which directed their attention to a specific ear. Results of the Continuous Performance Test supported this hypothesis. Children who cluttered showed poorer performance on the overall attentional index as well as a number of the subscores than children who stuttered and children who did not stutter. Implications for future research will be discussed.

**4aSCb20. Final consonant recognition in young children.** Jan Edwards, Robert A. Fox, and Catherine L. Rogers (Dept. of Speech and Hearing Sci., The Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210, [edwards.212@osu.edu](mailto:edwards.212@osu.edu))

Research has shown that typically developing children evidence gradual improvement with age on speech perception tasks. Studies have also shown that children with phonological disorder and specific language impairment perform more poorly than typically developing age peers on these same tasks. Previous research has mostly used synthetic speech stimuli, although it is known that additional attentional resources are required to process synthetic speech. Furthermore, most research has focused on identification or discrimination of word-initial consonants, rather than on consonants in other word positions. In our research, experimental paradigms were developed that use digitized natural speech, rather than synthetic speech, and focus on final rather than initial consonants, since final consonants tend to be acquired later in production. Results of a study with 15 typically developing 3- and 5-year olds will be presented. Familiar word pairs (cap versus cat and tap versus tack) were recorded, digitized, and presented at three different gates (up to 60 ms removed from final consonant closure). Values of  $d$ -prime were calculated for each gate, and an improvement in identification with age was found for the longer gates. Results will also be presented comparing 10 children with phonological disorder and age peers. [Research supported by NIDCD No. DC02932.]



**4aSCb21. Children's and adults' perception of speech in noise.** Marianne Fallon, Sandra E. Trehub, and Bruce A. Schneider (Dept. of Psych., Univ. of Toronto, 3359 Mississauga Rd. North, Mississauga, ON L5L 1C6, Canada, marianne@psych.utoronto.ca)

Age-related differences in identifying speech in noise were investigated by means of a computerized four-alternative forced-choice task. In experiment 1, children 5 to 11 and adults heard prerecorded instructions (e.g., "Touch the X") mixed with multitalker babble. Selections were made on a touch-sensitive screen and corrective feedback was provided. To control for age-related differences in sensitivity, signal-to-noise ratios yielding 85% correct were determined for each age group. The S/N difference between the youngest and oldest listeners was 5 dB, consistent with previous research. Equivalent additions of background noise had comparable effects for all age groups, which implies that previously reported age differences in noise were attributable to sensitivity differences rather than noise effects. Experiment 2 investigated the degree to which sentential context affected speech identification in similar noise. Children 5 and 9 years of age and adults identified the final word in high-context (e.g., "I carried the water in a pail") and low-context sentences (e.g., "He looked at the pail"), with differences in auditory sensitivity controlled. Adults benefited more from high sentential context than did younger children, indicating that limitations in cognitive processing adversely affect children's perception of speech in noise. [Work supported by the Medical Research Council of Canada.]

**4aSCb22. Possible perceptual cues for accent distinction in whispered speech—A case of an aphonic child.** Sawako Hirai, Noriko Kobayashi, and Hajime Hirose (School of Allied Health Sci., Kitasato Univ., 1-15-1, Kitasato, Sagamihara, Kanagawa, 228-8555 Japan)

Japanese word accent is associated with pitch change and  $F_0$  change is its major perceptual cue. In whispered speech, however,  $F_0$  is absent. We experienced a 10-year-old boy who had been using only whispered speech since birth, although he had no apparent physiological or psychological problems. After voice therapy, he started using voiced speech, but accent distinctions were poorly perceived in his voiced speech, while his whispered speech appeared to have better accent pattern variations. The purpose of this study was to find out the acoustic characteristics and perceptual cues produced by the child to make word accent distinctions in his voiced and whispered speech. The subject produced a vowel sequence /aa/ with different accent patterns. The results showed that accent distinctions were correctly perceived by listeners in both conditions, and that  $F_1$  change was significantly larger in whispered utterances than in voiced ones. The present result appeared to support the hypothesis of Higashikawa *et al.* (1997) that  $F_1$  change played a principal role in accent distinction in whispered speech. It was assumed that when the major cue ( $F_0$  change) was not available, another possible cue,  $F_1$  change, was exaggerated.

**4aSCb23. Which dyslexic children have speech perception difficulties?** Patricia A. Keating (Phonet. Lab, Linguist. Dept., UCLA, Los Angeles, CA 90095-1543, keating@humnet.ucla.edu), Marc F. Joanisse, Franklin A. Manis, and Mark S. Seidenberg (Univ. of Southern California)

The incidence of speech perception difficulties, and their co-incidence with other impairments, were examined in third-grade poor (bottom-quartile) readers. Several subgroups of dyslexics, plus age and reading-level controls, were compared in their identification of /d/ vs /t/ in a "dug-tug" continuum, and /p/ vs /k/ in a "spy-sky" continuum. Only those dyslexics who also did poorly on two standardized language tests of inflection and vocabulary—9 of 63 dyslexics—differed from controls in performance on these speech sound categorization tasks, with shallower categorizations. In contrast, dyslexics who did poorly on a test of phoneme deletion, but normally on the language tests, performed like controls in speech perception. Thus an association was found between speech perception and morphological/lexical knowledge, but not between speech per-

ception and phonological knowledge—an unexpected result. A follow-up study tested AX discrimination using the "spy-sky" stimuli. The same dyslexic children who had done poorly on the identification tasks differed from controls in discrimination; the other dyslexics again performed like controls. However, the deficit of the poorer perceivers is not auditory—their cross-category discrimination was normal, and their within-category discrimination better than normal. Thus their difficulty is specifically with speech sound categorization. [Work supported by NICHD, NIMH, and NSERC.]

**4aSCb24. Building and exploring (speech) categories in early infancy.** Francisco Lacerda (Dept. of Linguist., Stockholm Univ., S-106 91 Stockholm, Sweden, frasse@ling.su.se)

This paper will report the results from a speech perception study assessing the impact of correlated audio and visual information on category formation in infancy. The amount of speech and of visual information as well as the degree of correlation between the two modalities is systematically varied across subjects and will be subsequently related to the infants' success in picking up the categories implicit in the multisensory stimuli. Also the infants' active versus passive role during the conditioning phase is experimentally assessed in an attempt to study how attentional effects may influence the early stages of speech category formation. The study is designed to provide experimental evidence for the notion of emergent phonology and the exemplar-based approach presented by Lacerda (1998) at the 135th ASA meeting, Seattle.

**4aSCb25. The relation between infant speech perception and early communicative development: A longitudinal study.** Hwei-Mei Liu, Feng-Ming Tsao, Erica B. Stevens, and Patricia K. Kuhl (CHDD, Box 357920, Univ. of Washington, Seattle, WA 98195)

This study investigated the correlation between speech discrimination in young infants and early communicative development. The MacArthur Communicative Development Inventory, a parent-report measure of words and gestures, was administered to 20 13-month-old infants who had participated in a study of speech discrimination when they were 6 months old. In the earlier study, the conditioned Head-Turn (HT) technique was used to assess infants' discrimination of the Finnish vowels /i/ and /y/. Infants were required to meet a criterion of seven out of eight correct responses prior to completing a 30-trial test phase. Results showed that the number of trials infants needed to meet the criterion at 6 months of age was significantly negatively correlated at 13 months with the number of phrases understood ( $-0.662$ ) and words understood ( $-0.700$ ). In a regression analysis, the number of criterion trials was shown to be moderately predictive of later communicative development, accounting for approximately 40% and 45% of the variance in phrases understood and words understood. The development of these children will be followed with a second administration of The MacArthur Communicative Development Inventory when they are 16 months of age, to further map the relationship between infant speech perception and early communicative development.

**4aSCb26. Relationships between expressive vocabulary size and spoken word recognition in children.** Benjamin Munson (Dept. of Speech and Hearing Sci., Ohio State Univ., 110 Pressey Hall, Columbus, OH 43212)

Studies have shown that children require more acoustic information than adults to accurately identify spoken words [Munson *et al.*, *J. Acoust. Soc. Am.* **99**, 2591 (1996)]. In this study, 61 children aged 3.0 to 7.11 were tested to assess whether spoken word recognition accuracy was predicted by scores on standardized tests of expressive and receptive vocabulary, articulation ability, preliteracy skills, and phonological awareness. Two word recognition tasks were used: a gated series [Elliot *et al.*, *Percept. Psychophys.* **42**, 150–157 (1987)] in which subjects identified CVC words

with missing final consonants, and a noise-center series [Fox *et al.*, *J. Speech Hear. Res.* **35**, 892–902 (1992)], in which subjects identified CVC words with the medial vowel replaced by noise. Each task had four experimental conditions. Subjects' age predicted a significant proportion of variance in word recognition scores on the five conditions for which the majority of subjects performed above chance. Additionally, scores on a test of expressive vocabulary predicted a significant proportion of variance in two of these five conditions. No other test predicted relationships between expressive vocabulary growth and developmental changes in spoken word recognition. [Work supported by NIH.]

**4aSCb27. Developmental changes in perceptual weighting strategies for speech are contrast specific.** Susan Nittrouer and Marnie E. Miller (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

In the labeling of /s/-/ʃ/, compelling evidence exists that children pay more attention to formant transitions and less attention to noise spectra than adults do. The purpose of this work was to examine whether such age-related differences extend to other contrasts, specifically /f/-/θ/. This contrast was chosen because compelling evidence also exists that adults pay more attention to formant transitions and less attention to noise spectra for /f/-/θ/ decisions than they do for /s/-/ʃ/ decisions, undoubtedly because /f/ and /θ/ noise spectra provide little information. Thus adults' weighting strategies for this contrast should resemble those of children. To test that hypothesis, children (4, 6, and 8 years) and adults labeled fricative-vowel syllables consisting of natural /f/ and /θ/ noises combined with synthetic /a/ and /u/ portions in which formant transitions varied from ones appropriate for /θ/ to ones appropriate for /f/. Results showed similar weighting strategies for adults and children for /a/ syllables, but adults and 8-year-olds used the sparse information in the noise spectra for /u/ syllables to a surprising extent. Conclusions were that weighting strategies for some contrasts require no developmental adjustments, but that flexibility in perceptual weighting strategies is one general developmental change.

**4aSCb28. The influence of pitch contour on vowel discrimination in infants.** Laurel J. Trainor and Renée N. Desjardins (Dept. of Psych., McMaster Univ., Hamilton, ON L8S 4K1, Canada, ljt@mcmaster.ca)

Because the harmonics are farther apart in high- than in low-pitched vowels, it was predicted that it should be difficult to extract the frequencies of the formant resonances, and therefore, to discriminate high-pitched vowels. Infant-directed speech is very high in pitch, yet studies showing that young infants can discriminate vowels have used low-pitched male voices. The present study showed that 6-month olds were able to discriminate /i/ and /I/ when the pitch was 240 Hz, but were unable to do so when the pitch was 340 Hz, demonstrating that the high pitch of infant-directed speech indeed impedes vowel discrimination. How, then, do infants learn vowel discrimination? The answer may lie with the large pitch contours that are also characteristic of infant-directed speech. Because the harmonic structure follows the frequency of the fundamental, but the formant structure remains relatively constant across frequency shifts, a second prediction was made: discrimination should be better for vowels containing pitch contours than for those with steady pitch. This is exactly what was found. The addition of a downward pitch contour at both high (440–340 Hz) and low (240–140 Hz) pitches significantly improved infants discrimination of /i/ and /I/. [Research supported by NSERC.]

**4aSCb29. Differences in child and adult loudness judgments of rock music.** Donald Fucci (Lindley 219, School of Hearing and Speech Sci., Ohio Univ., Athens, OH 45701), Heather Kabler, Debbie Webster, and Doug McColl (Ohio Univ., Athens, OH 45701)

The present study was concerned with the perceptual processing of complex auditory stimuli in ten children ( $M$  age=8.1) as compared to ten young adults ( $M$  age=19.3) and ten older adult subjects ( $M$  age=54.2).

The auditory stimulus used was 10 s of rock music [Led Zeppelin, CD Recording No. 19127–2 (Atlantic Recording Group, New York 1969)]. All three groups provided numerical responses to nine intensity levels of the rock music stimulus (10, 20, 30, 40, 50, 60, 70, 80, and 90 dB above threshold). Results showed that the children demonstrated a wider range of numerical responses than both adult groups. The range of mean numerical responses for the children was 0.54 to 54.24. For the young adults the range was 0.76 to 11.37, and for the older subjects it was 1.6 to 23.31. Results suggest that the children were not bound by the same set of rules as the adults in regard to magnitude estimation scaling of the loudness of the rock music stimulus. Their internal scaling mechanisms appeared to be more flexible and broader based than those of the adults who participated in this study.

**4aSCb30. Speech intelligibility and vowel production in persons with dysphagia.** Fredericka Bell-Berti, Diane M. Scott, Rita Kachmarchyk, and Nancy M. Colodny (Dept. of Speech, Commun. Sci. and Theatre, St. John's Univ., Jamaica, NY 11439, bellf@mail.stjohns.edu)

Neurologically based dysphagias are often accompanied by dysarthria or apraxia. Dysarthrias have traditionally been identified perceptually, rather than with more objective instrumental measures, and vowel articulation is examined only rarely in dysarthria evaluations. It is possible that some speakers with neurologically based dysphagia have sub-clinical disturbances of speech that affect vowel as well as consonant production. This preliminary report examines the speech intelligibility and acoustical measures of vowel productions in a group of dysphagic subjects. Our goal is a better understanding of the relations between the dysphagia and the dysarthria. [Work supported by St. John's University.]

**4aSCb31. Intonation and speech timing: Association or dissociation?** Samuel A. Seddoh (Dept. of Speech Pathol. & Audiol., Southern Univ., P.O. Box 9227, Baton Rouge, LA 70813, seddoh@cluster.engr.subr.edu)

A number of prosodic studies on brain damaged patients [e.g., Gandour *et al.*, *Brain & Lang.* **43**, 275–307 (1992)] have concluded that the production of fundamental frequency ( $F_0$ ) is tied to temporal control, and that abnormalities in  $F_0$  associated with intonation exhibited by aphasic patients is due to underlying disturbance in speech timing. The present study examined  $F_0$  and durational measures in syntactically simple English statements and interrogatives produced by fluent and nonfluent aphasic patients as well as a group of normal subjects. Measures conducted included initial  $F_0$  peak (P1), the initial frequency at the terminal region of the  $F_0$  contour (TIF), amount of terminal  $F_0$  fall/rise, utterance length, duration of terminal  $F_0$  fall/rise, and distance between P1 and TIF. Results showed that the nonfluent aphasic patients performed at comparable levels with the normal subjects on all  $F_0$  measures, but they were severely impaired across the board on temporal measures. The fluent aphasic patients' performance on both sets of measures were generally comparable to normal performance, although in absolute terms their durations were also longer than normal. These findings contradict the view that  $F_0$  production depends on speech timing. Rather they suggest that the two parameters may be dissociated.

**4aSCb32. The effect of speaking rate on vowel production in Parkinson's disease.** Kris Tjaden (Dept. of Communicative Disord. and Sci., The State Univ. of New York at Buffalo, 3435 Main St., Buffalo, NY 14214-3005)

Kinematic and acoustic descriptions of articulatory characteristics in Parkinson's disease (PD) generally are consistent with reports of hypokinesia and bradykinesia in the limbs. That is, articulatory movements tend to be underscaled and produced at slower than normal rates in comparison

to healthy speakers, although movement duration may be preserved. Slowed speaking rate is one therapy technique used for improving the adequacy of speech production in PD. The assumption is that the slowed rate allows for more extensive articulatory displacements, thereby increasing the acoustic-perceptual distinctiveness of phonetic events. Few studies have quantified articulatory changes associated with speaking rate variation in PD, however. The present study examined the effect of speaking rate on vowel production for individuals with PD and healthy controls. Participants read a passage at habitual, slow, and fast rates. Segment durations,  $F1$ , and  $F2$  midpoint frequencies for the vowels /i/, /a/, /u/, and /æ/ were measured. Vowel space area was calculated for each speaker's habitual, fast, and slow reading rate. The change in vowel space area was related to the magnitude of speaking rate change across rate conditions for both groups. Vowel space areas for PD and healthy speakers also were compared for each rate condition. [Work supported by NIH.]

**4aSCb33. Effects of a computer-based voice modification device for a speaker with recurrent nerve paralysis.** Noriko Kobayashi, Hajime Hirose (School of Allied Health Sci., KITASATO Univ., 1-15-1, Kitasato, Sagami-hara, Kanagawa, 228-8555 Japan), Kenji Matsui, and Noriyo Hara (Matsushita Electric Co., Kyoto, 619-0237 Japan)

Patients with unilateral recurrent laryngeal nerve (RNL) paralysis may exhibit severe dysphonia depending on the resting position of the affected vocal fold. Major vocal symptoms of these patients are breathy voice, decreased vocal intensity, and physical fatigue due to great effort to approximate the vocal folds and to increase subglottal pressure. In some cases, phonosurgery is not applicable due to other medical problems, while a good prognosis is not expected by voice therapy. Technological assistance may be needed for such a patient. In this study, a real-time formant analysis-synthesis method was used to enhance speech intelligibility of a severely dysphonic patient with unilateral RNL paralysis. The voice source was modified by using inverse-filtered signals extracted from a normal speaker. A special hardware unit was designed to perform an analysis-synthesis process with temporal delay of 40 ms for signal processing. Vowels and sentences produced by the subject both with and without the device were analyzed. An approximately 20 dB increase of overall loudness level and increased harmonic components were observed for the processed speech. Perceptual judgment revealed improved speech intelligibility and less breathy voice quality for the processed speech. The subject reported less fatigue with the device as less effort was required for phonation.

**4aSCb34. Structure of phonetic categories produced by general learning mechanisms.** Andrew J. Lotto (Dept. of Psych. and Parmlly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626), Lori L. Holt (Carnegie Mellon Univ., Pittsburgh, PA 15213), and Keith R. Kluender (Univ. of Wisconsin, Madison, WI 53706)

The development of categories for complex auditory stimuli is an interest for both studies of general category learning and language acquisition. Previous work [Kluender *et al.*, *J. Acoust. Soc. Am.* **104**, 3568–3582 (1998)] demonstrated that avian species can learn to respond differentially to sounds from two vowel categories and the structure of their responses correlate well with human adult ratings of the vowels. In the current study, Japanese quail (*Coturnix japonica*) were trained to respond to either members of an /i/ or /ε/ distribution and to refrain, in both cases, from responding to members of an /I/ and /æ/ distribution. Birds responding to /ε/ (surrounded by /I/ and /æ/ in the vowel space) showed a prominent peak or “prototype” in their responses. Birds responding to /i/ (extreme in the vowel space) showed a weak or no “prototype,” but showed a strong gradient with response rate increasing for tokens further away from the other vowel distributions in the  $F1$ – $F2$  space. These data demonstrate that internal structure of (phonetic) categories is strongly influenced by relations to the competing stimulus set (vowel space). This is particularly important for theories of categorization or language acquisition that rely heavily on the existence of a “prototype.”

**4aSCb35. Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?** Lori L. Holt (Dept. of Psych., Carnegie Mellon Univ., Pittsburgh, PA 15213), Andrew J. Lotto (Loyola Univ. Chicago, Chicago, IL 60626), and Keith R. Kluender (Univ. of Wisconsin-Madison, Madison, WI 53706)

Listeners labeling members of an acoustic series modeling VOT (e.g., /ba/-/pa/) are more likely to identify tokens with higher  $f0$  as voiceless than they are for otherwise-identical tokens with lower  $f0$ s. This pattern of results may arise because a high  $f0$  enhances perception of voicelessness, in line with auditory enhancement accounts of speech perception. Alternatively, because  $f0$  and VOT covary in English production, it is possible that listeners respond in this manner due to experience with VOT/ $f0$  covariation in the speech signal. The present investigation was designed to tease apart the relative contributions of these two potential mechanisms. Japanese quail (*Coturnix coturnix japonica*) were trained to “label” stimuli drawn from VOT series by pecking a key. During training, each quail experienced one of three styles of VOT/ $f0$  covariation. For one group of quail, VOT and  $f0$  covaried naturally with voiceless series members having higher  $f0$ s than voiced members. Another group of quail heard the inverse, “unnatural” covariation. A final group experienced stimuli for which there was no covariation between VOT and  $f0$ . Results indicate that experience with VOT/ $f0$  covariation is the predominant force in shaping perception. Thus, general learning mechanisms may account for this symmetry between perception and production.

**Session 4aSP****Signal Processing in Acoustics: Wavelets in Acoustics**

Leon H. Sibul, Chair

*Pennsylvania State University, P.O. Box 30, University Park, Pennsylvania 16804***Chair's Introduction—8:15*****Invited Papers*****8:20****4aSP1. Wavelet domain system/channel characterization.** Randy Young (Appl. Res. Lab, Penn State Univ., P.O. Box 30, State College, PA 16804)

The primary operators in implementing wavelet transforms are the delay, scale, and correlation operators. The physical motion of sources and sensors causes signals to be compressed or dilated (scaled) depending upon their relative direction of motion. For many systems that use a signal modulation with a carrier frequency, this scaling operator leads to a Doppler shift of the carrier. For many acoustic signals (voice, machinery vibrations, structural vibrations, biomedical signals, etc.), they are not carrier-frequency modulated and they possess a large relative bandwidth. Additionally, the relatively slow propagation speeds compared to source/sensor motion leads to much higher compression/dilation ratios than in electromagnetics. Since scaling applies a different Doppler to each frequency, large bandwidth signals experiencing scaling are not well approximated by a single Doppler shift. Thus, the scaling operation, and, therefore, wavelet transforms are more applicable in acoustic analysis/system modeling than electromagnetic analysis. This discussion concentrates on the wavelet transform as an operator utilizing scaling, rather than a signal decomposition technique. Although wavelet transforms as signal decomposition tools have been extensively researched, their contribution as a system/channel analyzer for estimating motion of sources/sensors/environments is less developed. [Nancy Harned, ONR321US, is acknowledged for supporting this work (Contract N00039-97-D-0042).]

**8:50****4aSP2. Wavelet processing for wideband spreading function estimation.** Lora G. Weiss (Appl. Res. Lab, Penn State Univ., State College, PA 16804)

The statistic used to quantify the amount of environmental spreading a signal undergoes as it traverses through a channel is called a spreading function, and it includes the effects of moving, distributed scattering objects, multipath, boundary effects, etc., on the signal. Traditionally, narrow-band signals have been transmitted, and the spreading functions were estimated by calculating the outputs of narrow-band-matched filters. Now that wideband processing has become more accessible, we need a solid concept of estimating wideband spreading functions. This paper shows that a wideband spreading function can be estimated by computing the wavelet transform of the received signal while using the transmitted signal as the mother wavelet. The paper then computes the second-order statistic, called the wideband scattering function, associated with the wideband spreading function. To assess the total scattering, several scattering functions are convolved to yield an overall representation of the environment. This representation can then be incorporated into a detection processor. The payoff is that if any portion of the scattering environment is known *a priori*, this information can be exploited in the detector. As knowledge of the scattering process is acquired, it is combined via the cascaded scattering function formulation.

**9:20****4aSP3. Application of wavelets to scattering for wave-packet synthesis.** Charles F. Gaumont and David M. Drumheller (Code 7142, Acoust. Div. Naval Res. Lab., Washington, DC 20375-5320)

The wavelet transform can be used to analyze the time-scale domain structure of echoes derived from broadband, monostatic, acoustic scattering from ribbed shells. In the time-scale domain, the echoes can be seen to be composed primarily of a few individual components. Each component corresponds to a separate physical scattering mechanism in the shell. Yen has called these separated components wave packets, each having distinctive time-scale dimensions that derive from the physical process that generates it. Because each wave packet has distinctive time-scale domain properties and the wavelet transform is invertible, the wavelet transform can be used to isolate and synthesize a set of wave packets. The resulting wave packets are functions of aspect and time. Greater efficiency of representation can be achieved by representing these wave packets with a bilinear expansion. Wave packets generated with the wavelet transform can be used for matched filter detection. [Work supported by the Office of Naval Research.]

**9:50–10:05 Break**



10:05

**4aSP4. A new wavelet basis for extracting periodic acoustic signals in noise.** Antonio J. Miller and Karl M. Reichard (Grad. Prog. in Acoust., P.O. Box 30, State College, PA 16804-0030)

A new wavelet basis is constructed to extract periodic acoustic signals in noise. A wavelet based comb filter is constructed from the linear superposition of narrow-band, harmonically related wavelet components. The best choice of wavelet components has proven to be a hybrid design incorporating the flat passband of the Shannon wavelet with the transition band of the Morlet wavelet. There are several benefits to taking a wavelet approach to comb filtering when dealing with periodic acoustic signals. Good noise rejection using a finite impulse response filter requires a precise measurement of the signal period and an adaptive sample rate to capture an integral number of samples per period. The wavelet approach relaxes the perfect frequency localization requirement and gives the comb filter "teeth" a small, but finite passband. This passband allows for measurement error in the signal period and for the signal to be sampled at an arbitrary rate. The new approach has proved to be an attractive alternative to the time domain averaging techniques used in accessing mechanical fault conditions from vibration measurements of rotating machines.

10:20

**4aSP5. Wavelet parameters for speech synthesis.** Brian C. Tuttle and Claus P. Janota (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

A standard method of analyzing human speech is to divide it into its constituent phonemes. Synthesizing speech from these phonemes without including the effects of coarticulation, however, can result in discontinuous and sometimes unintelligible sounds. A recently proposed model for the analysis of speech coarticulation uses wavelet system characterization where wavelet transforms describe the time-frequency behavior of a signal's transmission channel. The objective of this research is to verify the proposed model by synthesizing the original speech from its analysis. The coarticulated speech in question is the consonant-vowel-consonant combination that occurs in words such as "deed" or "bib" where the two consonants are the same. These words along with their vowels spoken in isolation are recorded digitally and processed using the wavelet system model. The result is an analysis of the coarticulated word with respect to its vowel. The analysis shows in time and frequency how the vowel changes when spoken in context. Synthesizing the speech using the inverse process serves to verify the model. Comparing the result to the original speech reveals that the accuracy is determined mainly by the scale resolution of the wavelet transforms.

10:35

**4aSP6. Application of the discrete wavelet transform (DWT) to the enhancement of pulmonary breath sounds.** J. F. Forren and G. L. Gibian (Planning Systems, Inc., 7923 Jones Branch Dr., McLean, VA 22102, ggibian@plansys.com)

Aural evaluation of pulmonary breath sounds, or auscultation, is a critical part of diagnosis and management in Pulmonary and Critical Care Medicine. However, reliability between examiners within an institution or across institutions is low. A study by Wilkins *et al.* of 277 North American physicians at the American College of Chest Physicians 1988 annual meeting revealed great inconsistencies in interpretation of breath sounds. Normal breath sounds were identified as normal by only 80% of these physicians. In addition, variable interpretations were provided for patients with inspiratory and expiratory crackles as well as those with stridor. Clinicians had difficulty defining rhonchi and often misclassified them as wheezes. The algorithm, based on nonlinear operations on the DWT coefficients, makes adventitious features of lung sounds easier to hear. The algorithm makes both coarse and fine crackles much more clearly audible by suppressing the breath and other interfering sounds. While the processed stridor sounds less like stridor than before, the processing makes it easier to hear the difference between stridor and wheezes and thus aids in distinguishing vocal cord dysfunction from asthma. [Work supported by N.I.H.]

10:50

**4aSP7. Application of adaptive impulsive noise separation to automotive squeak and rattle detection/quantification.** Vy Tran, Sheau-Fang Lei, and Keng Hsueh (Ford Motor Co., vtran@ford.com)

Abnormal noises such as road-induced squeaks and rattles (S&R) in automobiles are very undesired and perceived as quality risks to the customers. Subjective evaluation by trained experts is the typical S&R assessment procedure for product development and quality control. S&R noises usually have impulsive characteristics (time varying, and transient), and are very difficult to be detected and quantified objectively using traditional signal processing techniques. Emerging wavelet signal processing technique and kurtosis impulsive criteria were used to develop an objective impulsive noise detection and separation technique (US Patent Filed). The method has been shown to work successfully with S&R samples in various frequency bands. The technique has been implemented in a portable PC to acquire or operate on any pre-recorded acoustic signal. The system then automatically detects the presence of any S&R, and separates them from the original composite signal. The results can be played back for listening or processed to characterize, and quantify the S&R noises. These new impulsive noise detection and separation tools have been applied to S&R assessment at the vehicle level and subsystem durability tests. Applications to vibration problems have been envisioned. Conceptually, this technique may also have potential signal enhancement applications in the telecommunication field.

11:05–11:35

#### Panel Discussion

#### Why Wavelet Transform Domain Processing?

## Session 4aUW

**Underwater Acoustics, Acoustical Oceanography and Animal Bioacoustics: The Effect of Man-Made Sound on Marine Mammals I**

James F. Lynch, Chair

*Woods Hole Oceanographic Institution, 203 Bigelow Building, Woods Hole, Massachusetts 02543***Chair's Introduction—8:00***Invited Papers***8:05**

**4aUW1. An acoustic integration model (AIM) for assessing the impact of underwater noise on marine wildlife.** William T. Ellison (Marine Acoust., Inc., P.O. Box 340, Litchfield, CT 06759), Karen Weixel (Marine Acoust., Inc., Middletown, RI 02842), and Christopher W. Clark (Cornell Univ., Ithaca, NY 14850)

In recent years there has been a heightened awareness of the environmental impact of noise, especially man-made noise, on marine wildlife. The National Environmental Policy Act (NEPA), Executive Order 12114, The Endangered Species Act, The Marine Mammal Protection Act, and the Coastal Zone Management Act each provide for varying levels of regulation and control in protection of the environment and marine wildlife. In order to assess the environmental impact of a sound source, one must predict the sound levels that any given species will be exposed to over time in the locale of the source's radiated field. This is a three-part process involving (1) the ability to measure or predict an animal's location in space in time, (2) the ability to measure or predict the sound field at these times and locations, and finally, (3) integration of these two data sets so as to determine the net acoustic impact of the sound source on any specific animal. This paper describes a modeling methodology for accomplishing this task. Model inputs required to specify the acoustic environment, animal distribution and behavior, and sound source characteristics are discussed in detail. The AIM model output capabilities are described together with topical examples.

**8:35**

**4aUW2. Masked temporary threshold shift (MTTS): A relevant measure for hearing shifts in open waters.** Sam H. Ridgway, Donald A. Carder, James J. Finneran (SPAWARSYSCEN SAN DIEGO, Div. D35, 49620 Beluga Rd., San Diego, CA 92152-6266; ridgway@spawar.navy.mil), and Carolyn E. Schlundt (Sci. Applications Int'l. Corp., 3990 Old Town Ave., Ste. 105A, San Diego, CA 92110)

Cetaceans use sound in foraging, communication, and navigation. There is growing concern that human-made sound is a potentially serious auditory problem for cetaceans. Available data were inadequate to allow confident predictions regarding the levels of sound that should be of concern. An accepted method of determining the level of sound that produces a temporary reduction in the ears ability to respond fully is called temporary threshold shift (TTS). In the open waters of San Diego Bay, masked hearing thresholds were estimated for dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, before and after exposure to louder 1-s tones at five sonar frequencies. Masking noise was used to create a floor effect, thereby eliminating threshold variability in the ever-changing ambient noise environment in the natural waters where tests were conducted. The criterion for masked TTS (MTTS) was a 6-dB increase in threshold over baseline. Subjects exhibited alterations in their behavior starting at levels around 180 dB, and experienced MTTS at levels between 192-202 dB. Maximum levels of sound exposure without risk of hearing damage can be determined with the MTTS approach. More research employing different sound sources is needed. [Supported by Office of Naval Research.]

*Contributed Paper***9:05**

**4aUW3. Characteristics of a set of noises used to test effects of underwater explosions on sea mammal hearing.** Joseph A. Clark, Paul M. Moore (CDNSWC, Code 734, Bethesda, MD 20084), Jane A. Young (CDNSWC, Bethesda, MD 20084), and Joel B. Gaspin (IHNSWC, Indian Head, MD 20640)

As part of a study of temporary threshold shifts (TTS) expected to occur in sea mammals near underwater explosions, a set of ten explosion-like noises were synthesized [J. A. Clark, J. A. Young, and J. B. Gaspin,

J. Acoust. Soc. Am. **105**, 1048 (1999)]. The ten transient noises differed principally in maximum peak-to-peak pressure and were used in the TTS study as an ascending series of levels presented to two beluga whales and two bottlenose dolphins. The set of levels were selected from a larger set of predicted signatures of underwater explosions corresponding to a variety of charge weights and ranges in an underwater environment of interest which was characterized by charge depth, receiver depth, water depth, and sound-speed profile. This talk describes the ten transients used for the TTS studies. Time histories, frequency spectra, time-frequency distributions, and a variety of characteristics derived from these graphical descriptions will be presented.

## Invited Paper

9:20

**4aUW4. Temporary threshold shift in pinnipeds induced by octave-band noise in water.** David Kastak, Brandon L. Southall, Ronald J. Schusterman, and Colleen J. Reichmuth (Long Marine Lab., Univ. of Calif., 100 Shaffer Rd., Santa Cruz, CA 95060)

Low-frequency noise from manmade sources represents an increasing portion of the total noise in the ocean. Such noise may adversely impact diving pinnipeds, which hear relatively well at low frequencies. The effects of octave bands of moderately intense noise (under 4 kHz, 60–76 dB SL, 20–22 min duration) were examined in three species of pinniped, the California sea lion (*Zalophus californianus*), harbor seal (*Phoca vitulina*), and northern elephant seal (*Mirounga angustirostris*). Auditory sensitivity data were obtained behaviorally, using a go/no-go psychophysical procedure. Thresholds were determined before noise exposure, immediately after noise exposure, and again after 24 h. Mean threshold shifts ranged from 2.9–4.9 dB at the center frequency of the noise band. Recovery was complete after 24 h. These results show that noise levels as low as 60 dB SL can induce TTS. However, between 60 and 76 dB SL, there was no relationship between noise level and degree of TTS. Additionally, maximum hearing loss occurred at center frequency rather than one-half octave above the center frequency. Further testing at longer exposure durations and higher noise levels is needed to more completely assess the degree to which pinniped underwater hearing is affected by continuous noise exposure.

9:50–10:05 Break

## Contributed Paper

10:05

**4aUW5. A software package calculating zones of impact on marine mammals around industrial noise sources.** Christine Erbe and David M. Farmer (IOS-Ocean Acoust., 9860 W. Saanich Rd., Sidney, BC V8L 4B2, Canada, erbec@dfo-mpo.gc.ca)

A software package is presented which estimates zones of interference around underwater noise sources affecting marine mammals. An ocean sound propagation model based on ray theory computes the spreading of complex underwater sound such as broadband animal vocalizations and manmade noise. On a grid of receiver locations (representing the affected marine mammal), the received signal and noise sound spectra are compared. Given a species-specific audiogram, the software package plots

zones of audibility around the noise source. Given species-specific vocalizations, zones of masking are plotted based on results obtained during an earlier study which measured masked hearing thresholds of a beluga whale. Tools developed during this study (such as an artificial neural network model to predict the amount of masking) can be linked to the software package. Zones of behavioral disturbance are plotted based on received sound levels reported in the literature. Zones of discomfort, injury, and hearing loss could be plotted if thresholds were known or using current estimates. The software package is applicable to a variety of ocean environments requiring location-specific oceanographic input data. The case of icebreakers affecting beluga whales in the Beaufort Sea is demonstrated.

## Invited Paper

10:20

**4aUW6. Temporary threshold shift in hearing induced by an octave band of continuous noise in the bottlenose dolphin.** Whitlow W. L. Au, Paul E. Nachtigall, and Jeffery L. Pawloski (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

Temporary threshold shift in hearing at 7.5 kHz was studied with an Atlantic bottlenose dolphin. Immediately following a threshold measurement, the animal was required to station in a hoop and be exposed to an octave band of continuous noise from 5 to 10 kHz. Noise exposure sessions lasted about 50 min, with the requirement that the animal spend a total of 30 min in the hoop. The dolphin also had two preferred locations both about a meter to the side of the hoop, one at the surface, and the other at the hoop depth. The noise levels at the hoop and to the side were about the same but with different spectra. The noise at the surface was about 3-dB lower. After exposure to the fatiguing stimulus, the animal's hearing sensitivity was immediately measured. The animal's hearing was not affected when the noise was 171 dB at 1  $\mu\text{Pa}$  with a total energy flux density of 205 dB at 1  $\mu\text{Pa}^2$  s. Temporary threshold shifts of 12–18 dB were obtained when the noise increased to 179 dB with an energy flux density of 213 dB or 1330  $\text{J/m}^2$ . The fatiguing stimulus was about 96 dB above the animal's pure tone threshold of 84 dB. [Work supported by ONR.]

10:50

**4aUW7. Masked temporary threshold shift for impulsive sounds in dolphins and white whales.** Donald A. Carder, James J. Finneran, Sam H. Ridgway (SPAWARSYSCEN San Diego, Div. D35, 49620 Beluga Rd., San Diego, CA 92152-6266), and Carolyn E. Schlundt (Sci. Applications Intl. Corp., San Diego, CA 92110)

In an effort to develop acoustic safety criteria for marine mammals that may be exposed to impulsive sounds a masked temporary threshold shift (MTTS) experimental approach was used. Two dolphins, *Tursiops truncatus*, and two white whales, *Delphinapterus leucas*, were systematically exposed to increasing levels of impulsive sound (S-1) under water, then tested to threshold at three pure-tone frequencies (S-2) under water in the presence of controlled masking noise. One hearing frequency tested was at the pulse peak frequency, with the other two S-2 test tones at higher frequencies. Behavioral responses were monitored and hearing thresholds were ascertained to determine if criterion (6-dB increase over baseline) was reached. Thresholds were estimated using the free response paradigm method and an up-down staircase psychophysical procedure for stimulus presentation in open water in San Diego Bay. These data suggest that the rapid rise time of the leading edge in 1-s tones used in a previous study is not responsible for producing MTTS. Results of the present study add to

the existing knowledge of MTTS for marine mammals and provide the first data on MTTS in cetaceans for impulsive sounds. [Work supported by the Office of Naval Research.]

11:05

**4aUW8. The ASA standards program and underwater sound, are they incompatible?** Daniel L. Johnson (Brüel Bertrand & Johnson Acoustics, 4719 Mile High Dr., Provo, UT 84604)

In the ASA National Standards Catalog, there are over 70 standards that concern the measurement and evaluation of sound in air. There are only a few standards, such as ANSI S1.11-1986 (filters) or ANSI S1.22-1992 (underwater transducer calibration) that concern underwater sound. This latter standard is the only standard of the 111 current standards that specifically relates to underwater acoustics. Standards such as ANSI S1.26-1995 (absorption of sound by the atmosphere) or ANSI S12.9-1996 (measurement of occupational noise exposure) just do not have a counterpart for water. Is this because there is no need? Is there no value in standardizing some common practices for underwater sound measurement and evaluation? Is it indeed true that underwater sound needs few standards? Comments from the audience are expected.

11:20–11:50

Panel Discussion



**Meeting of the Standards Committee Plenary Group**

**ORGANIZATION OF STANDARDS COMMITTEE PLENARY GROUP MEETING**

- **S1 ACOUSTICS**—U.S. Technical Advisory Group (TAG) for IEC/TC 29 Electroacoustics and ISO/TC 43 Acoustics
- **S2 MECHANICAL VIBRATION AND SHOCK**—U.S. Technical Advisory Group (TAG) for ISO/TC108 Mechanical Vibration and Shock
- **S3 BIOACOUSTICS**—U.S. Technical Advisory Group (TAG) for ISO/TC 43 Acoustics, IEC/TC 29 Electroacoustics, and ISO/TC 108/SC4 Human Exposure to Mechanical Vibration and Shock
- **S12 NOISE**—U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC1 Noise

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S3, and S12, to take place in the following sequence on the same day:

<b>S12</b>	9:45 a.m. to 11:00 a.m.	Union B Room
<b>S1</b>	2:00 p.m. to 3:30 p.m.	Union B Room
<b>S3</b>	3:45 p.m. to 4:45 p.m.	Union B Room

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees, plus a review of the international standardization (U.S. TAG) activities including reports on recent meetings and planning for forthcoming meetings.

Members of S2 on Mechanical Vibration and Shock (and U.S. TAG for ISO/TC 108 and five of its Subcommittees, SC1, SC2, SC3, SC5 and SC6) are also encouraged to attend the Standards Committee Plenary Group meeting, even though the S2 meeting will take place one day earlier, on Wednesday, 3 November 1999 at 9:00 a.m.

The U. S. Technical Advisory Group (TAG) Chairs for the various international Technical Committees and Subcommittees under ISO and IEC, which are parallel to S1, S2, S3 and S12 are as follows:

<u>U.S. TAG Chair/Vice Chair</u>	<u>TC or SC</u>	<u>U.S. TAG</u>
<b>ISO</b>		
P. D. Shomer, Chair H. E. von Gierke, Vice Chair	<b>ISO/TC 43</b> Acoustics	S1 and S3
P. D. Shomer, Chair H. E. von Gierke, Vice Chair	<b>ISO/TC 43/SC1</b> Noise	S12
D. Reynolds, Chair	<b>ISO/TC 108/SC4</b> Human Exposure to Mechanical Vibration and Shock	S3
D. J. Evans, Chair	<b>ISO/TC 108</b> Mechanical Vibration and Shock	S2
R. H. Mehta, Chair K. Won, Vice Chair	<b>ISO/TC 108/SC1</b> Balancing, including Balancing Machines	S2
A. F. Kilcullen, Chair	<b>ISO/TC 108/SC2</b> Measurement and Evaluation of Mechanical Vibration and Shock as Applied to Machines, Vehicles and Structures	S2
D. J. Evans, Chair	<b>ISO/TC 108/SC3</b> Use and Calibration of Vibration and Shock Measuring Instruments	S2
D. J. Vendittis, Chair R. F. Taddeo, Vice Chair	<b>ISO/TC 108/SC5</b> Condition Monitoring and Diagnostics of Machines	S2
G. Booth, Chair	<b>ISO/TC 108/SC6</b> Vibration and Shock Generating Systems	S2
<b>IEC</b>		
V. Nedzelnitsky, U.S. TA	<b>IEC/TC 29</b> Electroacoustics	S1 and S3

4a THU. AM

**Meeting of Accredited Standards Committee (ASC) S12 on Noise**

P. D. Schomer, Chair S12, and Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC1, Noise  
*2117 Robert Drive, Champaign, Illinois 61821*

R. W. Hellweg, Vice Chair, S12  
*Compaq Computer Corporation, MS PK02-1/J60, Maynard, Massachusetts 01754*

H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC1, Noise  
*1325 Meadow Lane, Yellow Springs, Ohio 45387*

**Accredited Standards Committee S12 on Noise.** Working group chairs will report on their progress for the production of noise standards.

**Scope of S12:** Standards, specifications and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control; including biological safety, tolerance and comfort and physical acoustics as related to environmental and occupational noise.

**Session 4pAA****Architectural Acoustics and Musical Acoustics: Archeological Acoustics II**

David Lubman, Chair  
*David Lubman & Associates, 14301 Middletown Lane, Westminster, California 92683*

***Invited Papers*****1:00**

**4pAA1. Acoustics of ancient Chinese bells.** Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Many ancient Chinese bells, some more than 3000 years old, remain from the time of the Shang and Zhou dynasties. Most of these bells, being oval or almond shaped, sound two distinctly different musical tones, depending upon where they are struck. Studies of these bells have provided us with much knowledge about the musical culture of Bronze age China. Results are reported from our investigations of original bells in the Shanghai Museum and the Sackler Gallery at the Smithsonian Institution, as well as replica bells in our own laboratory. Modes of vibration in original bells, obtained by experimental modal testing in the museum, compare very well with holographic interferograms of replica bells. Our results are compared with our catalogue of data on ancient Chinese bells gathered by other investigators. Although acoustical studies have revealed a great deal about ancient Chinese music and casting practices, several questions still remain unanswered.

**1:30**

**4pAA2. Acoustics of Karen bronze drums.** Laura M. Nickerson and Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Bronze drums are important to the culture of the Karen people, who live mainly in Burma and the mountainous region between Burma and Thailand. A Karen bronze drum, which is cast in one piece except for the animal adornments, consists of an overlapping tympanum that may range from 9–30 in. in diameter and a cylinder that is slightly longer than the diameter. Bronze drums have a wide variety of ritual use, both musical and nonmusical. The sound spectrum of a drum shows a dense collection of partials out to about 3 kHz. The corresponding modes of vibration, recorded using electronic TV holography, sometimes show the strongest vibration in the tympanum and sometimes in the cylindrical shell. The sound of the drum and the manner of playing will be illustrated by videotape.

2:00

**4pAA3. The carnyx, the lur, and other ancient European horns.** D. Murray Campbell (Dept. of Phys. and Astron., Univ. of Edinburgh, Edinburgh EH9 3JZ, UK, D.M.Campbell@ed.ac.uk)

Modern European brass instruments are the result of a long period of evolution. Relatively little is known about their ancestors, the lip-excited horns which were widely used in Europe several thousand years ago. Various attempts have been made to reconstruct playable versions of such instruments as the Scandinavian lur and the Celtic carnyx; although there is a considerable speculative element in all of these reconstructions, they offer the intriguing possibility of carrying out both musical and acoustical tests on these long-vanished instruments. In the last few years a project funded by the National Museums of Scotland has resulted in the reconstruction of two specimens of a carnyx, based on fragments found in Deskford, in Scotland, in the nineteenth century. Only the bell section, in the shape of a boar's head, had survived, but from other pictorial evidence it is known that the instrument had a tube about 2 m long, and was held vertically. This paper reports on a series of acoustical tests which have been carried out on the two reconstructions, and compares the acoustical and musical properties of the carnyx with those of reconstructed lurs and other ancient horns. [Work supported by EPSRC(UK).]

THURSDAY AFTERNOON, 4 NOVEMBER 1999

KNOX ROOM, 2:00 TO 3:30 P.M.

### Session 4pBB

## Biomedical Ultrasound/Bioresponse to Vibration: Nonlinear Acoustics in Medicine

Robin O. Cleveland, Chair

*Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215*

### Invited Papers

2:00

**4pBB1. Full wave simulations of pulsed focused fields in a nonlinear tissue mimic.** Thomas L. Szabo, Frances Clougherty (Hewlett Packard, MS-095, 3000 Minuteman Rd., Andover, MA 01810), Gregory L. Wojcik, John C. Mould, and Laura M. Carcione (Weidlinger Assoc., Los Altos, CA 94022)

In order to simulate focused ultrasound fields in nonlinear tissue, a full wave pseudospectral finite difference model has been developed. Unlike models based on the KZK equation, this model can predict wide-angle propagation effects as well as tissue backscatter, essential for pulse-echo medical imaging. This simulation program is capable of including broadband pulse propagation, causal attenuation, nonlinearity (B/A), and inhomogeneities. Harmonics beyond the solver's Nyquist limit are removed numerically and quantified. To facilitate comparison to data and to expedite calculation, the model has been adapted for circularly symmetric transducers. As input to the program, waveforms are taken from hydrophone scans close to the faces of 2.25- and 5-MHz focusing transducers. Simulations are compared to data taken in water and in front of and behind a tissue mimicking cylinders of tofu, which are sliced repeatedly to give transmitted and reflected pressure data versus thickness. Tissue inhomogeneities are simulated by voids in the tofu. Acoustic properties of the mimic are measured separately. B/A is extracted numerically from observed harmonic generation versus input pressure and range.

### Contributed Papers

2:30

**4pBB2. The acoustic field of a clinical pulsed Doppler ultrasound system near lung.** Christy K. Holland and Jeffrey T. Flament (Dept. of Radiol., M.L. 0742, Univ. of Cincinnati, 234 Goodman St., Cincinnati, OH 45219-2316, Christy.Holland@uc.edu)

Several animal models have exhibited thresholds for petechial hemorrhage in lung exposed to clinical levels of diagnostic ultrasound. Extravasation of red blood cells into the alveolar spaces has occurred consistently near the visceral pleura (at the outer surface of the lung). In order to quantify the ultrasound exposure conditions under which damage occurs, the pressure field of an ATL HDI 3000 in Doppler mode was measured near *ex vivo*, aerated rat lung in a water bath. The pressure output of the ATL L10-5 linear array in pulsed Doppler mode (6.0 MHz center frequency) was interrogated near the lung with a Sonic Technologies 0.4-mm bilaminar membrane-type PVDF hydrophone. A Velmex 3-axis translation stage was utilized for automated positioning of the hydrophone. Virtual instruments were developed using LABVIEW v.5.0.1 on a Macintosh 8100 to sample the pressure waveforms from a LeCroy 9350CL digital oscilloscope via an IEEE 488 interface and National Instruments data acquisition boards. Images of the pressure field in the axial, transverse, and eleva-

tional planes near the surface of the lung will be compared to the same measurements in the free field (without the lung present).

2:45

**4pBB3. Path-integrated Goldberg number as a predictor of enhanced heating from finite-amplitude therapeutic ultrasound sources.** Ibrahim M. Hallaj, Kullervo Hynynen (Brigham and Women's Hospital, Harvard Med. School, 221 Longwood Ave., Boston, MA 02115), and Robin O. Cleveland (Boston Univ., Boston, MA 02215)

Several theoretical and experimental studies have demonstrated that the higher-frequency harmonics generated by finite-amplitude propagation from therapeutic ultrasound sources result in enhanced heating of tissue. However, published results vary in their assessments of the significance of the nonlinearity on tissue heating. The variations in the results are due to the use of different sources and propagation paths. It would be useful to have an easily computed estimator of finite-amplitude effects for therapeutic ultrasound sources. The present study describes a dimensionless quantity, which takes the propagation path into account, to estimate the excess temperature rise due to nonlinearity. The dimensionless parameter is based on the Goldberg number which is a measure of the importance of nonlin-

ear effects to absorption effects. The relationship between the enhanced heating of tissue and the path-integrated Goldberg number for water-tissue paths is presented for typical therapeutic devices. [Work sponsored by National Cancer Institute.]

3:00

**4pBB4. Holographic interferometric visualization of weak shock waves in castor oil packed with equal-diameter gelatine spheres.** S. Hamid R. Hosseini (Shock Wave Res. Ctr., Inst. of Fluid Sci., Tohoku Univ., 2-1-1 Katahira, Aoba, Sendai, 980-8577 Japan, hosseini@ceres.ifs.tohoku.ac.jp), S. Moosavi Nejad (Tohoku Univ., Sendai, Japan), Tasneem B. F. R. Nahaboo, and Kazuyoshi Takayama (Tohoku Univ., Sendai, Japan)

The attenuation of shock waves in castor oil in which equal-diameter gelatine spheres were packed was measured with hydrophones at various stand-off distances. The motion of shock waves was quantitatively visualized by using double exposure holographic interferometry. Diameters of gelatine spheres under study were  $3.0 \text{ mm} \pm 1\%$ ,  $4.5 \text{ mm} \pm 1\%$ , and  $6.0 \text{ mm} \pm 1\%$ . Shock waves were produced at the center of a chamber in which gelatine spheres of total 4 L in volume were densely immersed in castor oil. In order to generate shock waves, irradiation of a pulsed YAG laser beam on silver azide pellets were used. The weight of silver azide pellets ranged from  $3 \mu\text{g}$  to 10 mg, with their corresponding energy of 4.5 mJ to 15 J. The effect of the nonuniformity on the shock-wave attenuation

created due to the multiple interaction of shock waves with the gelatine spheres was experimentally clarified. The result will be useful to establish a model of human tissue in numerically simulating the complex interaction of underwater shock waves with human tissue which takes place in shock-wave therapies.

3:15

**4pBB5. Ultrasound field correction in trans-skull ultrasound surgery using a focused array.** Gregory Clement, Jie Sun, Tonia Giesecke, and Kullervo Hynynen (Brigham & Women's Hospital and Harvard Med. School)

Correction of the distortion due to diffraction, reflection, and attenuation in trans-skull ultrasound propagation is investigated. Human skull fragments are positioned between an underwater focused transducer array and a PVDF needle hydrophone. A positioning system is used to scan the hydrophone along a plane normal to the transducer face to provide amplitude and phase delay information as a function of position. The hydrophone measurements are then used to manually correct the driving signal phase of each element in the array. The field is remeasured and the correction evaluated in terms of the obtained focus and the occurrence of grating lobes. The measurements are repeated at different orientations about each skull sample. Finally, skull information is obtained using MRI and CT scans. Variation of the measured ultrasound field as a function of skull thickness, density, and transducer location is examined.

THURSDAY AFTERNOON, 4 NOVEMBER 1999

HARRISON ROOM, 1:30 TO 5:00 P.M.

### Session 4pEA

## Engineering Acoustics: Transducer Design and Calibration

R. Daniel Costley, Jr., Chair

National Center for Physical Acoustics, MILTEC, P.O. Box 878, University, Mississippi 38677

### Contributed Papers

1:30

**4pEA1. Application of differential-capacitance accelerometers to underwater acoustic velocity sensing.** Thomas B. Gabrielson (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, tbg3@psu.edu)

The differential-capacitance accelerometer is relatively common in medium- to high-grade commercial accelerometers but has been largely ignored for application to underwater acoustic velocity or acceleration sensors. The differential-capacitance configuration is capable of good long-term stability and exceptional low-frequency noise performance. The effective sensor impedance is determined not by the signal frequency but by the ac bias frequency. The low sensor impedance associated with high-frequency ac bias relaxes the requirements on signal cables. In addition, the critical preamplifier characteristics are those at the ac bias frequency so low-frequency ( $1/f$ ) noise can be avoided. The differential-capacitance accelerometer does not have an inherent roll-off in response at low frequency so these devices can be used in self-orienting three-axis sensors for deployable systems. In machinery monitoring applications, synchronous detection of the ac-modulated signal yields a high degree of immunity to external interference and is amenable to straightforward multiplexing and optical or rf transmission. This paper summarizes the design principles associated with the differential-capacitance accelerometer through two illustrations: modification of a low-grade commercial product for improved signal-to-noise performance and multisensor multiplexing in the context of acoustic velocity sensing. [Work supported by the Office of Naval Research.]

1:45

**4pEA2. Generalized state-variable model for the design of a closed-loop differential capacitive accelerometer.** Barry J. Doust, Thomas B. Gabrielson (Grad. Prog. in Acoust. and the Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804), and Jeffrey L. Schiano (The Penn State Univ., University Park, PA 16802)

In an effort to develop a high-performance, miniature, directional acoustic sensor, the advantages of closed-loop operation are investigated. A nonlinear state-variable model of a generalized sensor system is presented. The sensor model is differential capacitive with electrostatic actuation for force rebalance of the proof mass. The physical model parameters are chosen to replicate scales and tolerances of typical MEMS devices with particular attention to nonlinearities associated with the electrostatic force and large deflections. Modeling and simulation will be presented for various control schemes and results will be compared in terms of important operational criteria such as dynamic range, bandwidth, linearity, and noise. [Work supported by the ONR AASERT program.]

2:00

**4pEA3. Fiber-optic flexural-disk accelerometer.** Andrew J. Doller, Karl M. Reichard, and Steven L. Garrett (Grad. Prog. in Acoust. and Appl. Res., The Penn State Univ., P.O. Box 30, University Park, PA 16804, adoller@psu.edu)

A fiber-optic, interferometric accelerometer incorporates a centrally supported flexural disk [U.S. Patent No. 5,368,485 (1994)] to convert vibration into differential strains on two spiral coils of optical fiber bonded to the disk's top and bottom surfaces. The sensor coils form two legs of a



Mach-Zender interferometer, terminated in a  $3 \times 3$  coupler, whose outputs were used in a passive homodyne symmetric signal demodulator [U.S. Patent No. 5,313,266 (1994)]. A new digital implementation of this demodulation algorithm used fast Fourier transforms to perform the necessary algebraic operations on time-windowed data segments, eliminating drifts. The sensor/demodulator system bandwidth was measured from 42 Hz to 5 kHz. A minimum detectable signal of 7 mrad was obtained with a 10-dB signal-to-noise ratio. At the lowest frequency, this corresponded to a dynamic range of 105 dB, which is 9 dB greater than the theoretical maximum for the 16-bit digitizer, due to interferometric conversion of amplitude modulation into frequency modulation. Thick-plate elastic models agreed with measured disk resonance frequencies. The measured sensitivity of 17 rad/g below resonance was less than the theoretical value of 23 rad/g, probably due to imperfect strain transfer through the epoxy adhesive used to bond the fibers to the disk's surface. [Work supported by a grant from the Office of Naval Research.]

2:15

**4pEA4. Minimum point search method for improvements in absolute calibration of accelerometers at high frequencies using the fringe-disappearance method.** Lixue Wu (Acoust. Standards, Inst. for Natl. Measurement Standards, Natl. Res. Council Canada, Ottawa, ON K1A 0R6, Canada)

The measurement uncertainty in the absolute calibration of accelerometers at high frequencies using the fringe-disappearance method was reduced by a minimum point search method. Outputs of a bandpass filtered photodetector and an accelerometer were simultaneously recorded by two true rms digital voltmeters. Displacement amplitudes were finely controlled by a precision attenuator. Minimum points of the output of the photodetector and the corresponding sensitivities of the accelerometer were found from the recorded data. During data processing, a simple model for the minimum points was proposed and a least mean squares algorithm was used to estimate parameters of the model. Results of calibrations were compared with those obtained using fringe-counting method at lower frequencies. It was found that the minimum point search method could compare with the fringe-counting method in measurement uncertainties. Good measurement reproducibility was also observed that resulted in the reduction of total measurement uncertainty.

2:30

**4pEA5. Alternating-current calibration of solid-state gyroscopes by reciprocity.** Thomas B. Gabrielson (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, tb3@psu.edu)

One of the hot topics in MEMS research is the development of miniature solid-state angular-rate sensors (gyroscopes). These devices have a number of potential applications including navigation systems for divers and unmanned vehicles, interpolators for periods of GPS loss of signal, and automobile skid sensors. For many of these applications, constant-rate (dc) calibrations are not representative of the expected excitations. Fortunately, high-quality ac-rate calibration can be done with an inexpensive apparatus by means of reciprocity. A suspended platform is driven in torsional oscillation by a pair of geophones acting as a force couple. Another pair of geophones is used as an angular velocity sensor. Since both of these transducer pairs are reciprocal, they can be used in conjunction with the device under test to perform a complete reciprocity calibration. The platform produces very smooth ac-rate motion from 1 to 20 Hz at rates of below 0.1 deg (rms)/h to 200 deg/h. The transfer impedance used in the reciprocity calibration is determined by the polar moment of inertia of the platform. This can be determined accurately by adjusting the separation of a symmetrically placed pair of masses and measuring the change in natural frequency. [Work supported by the Office of Naval Research.]

**4pEA6. Thin, low-frequency sound projectors for use in shallow water.** James F. Tressler and Thomas R. Howarth (Naval Res. Lab., 4555 Overlook Ave., SW, Washington, DC 20375-5350, tressler@acoustics.nrl.navy.mil)

A thin, low-frequency acoustic projector is being developed at the Naval Research Laboratory for shallow water applications. The active panel consists of cymbal-type flexensional driver elements arranged individually behind an acoustic radiating plate. Variants of two panel designs have been calibrated underwater. The acoustic projector sandwich is nominally 6.35 mm thick and is intended to operate between 1 and 10 kHz. The calibration results that will be presented (TVR, SPL, directivity response) will focus primarily on a panel consisting of 12.7-mm-diam cymbal elements arranged in an  $8 \times 8$  square configuration. [This research is being supported by Code 321SS of the Office of Naval Research.]

3:00-3:15 Break

3:15

**4pEA7. Reduction of measurement variation: Small acoustic chamber screening of hard disk drives.** Tami Ogle (Western Digital Corp., 822 Creek Dr., San Jose, CA 95125) and Lee Hornberger (Santa Clara Univ.)

Several small acoustic chambers on the manufacturing floor are used at Western Digital to screen out unacceptably loud disk drives. The screening test initially had large variation which made its effectiveness questionable. It was unclear whether this variation was coming from the drives or the chambers. In this study statistical methods were used to determine the source of the variation, and a design of experiments approach was used to determine the factors which significantly contributed to the variation. It was determined that the measurement chambers, not the drives, were the source of the large variation in testing. Additionally, the significant factors affecting the screening test were found to be the microphone calibration and microphone height inside the chamber.

3:30

**4pEA8. A low-cost sound level meter based on personal computer.** Guillermo de Arcas, Juan M. Lopez (Dept. of S.E.C., E.U.I.T. Telecomunicacion, Universidad Politecnica de Madrid, Crta. Valencia km 7, Madrid 28031, Spain, garcas@sec.upm.es), Manuel Recuero, and Alberto Martin (I.N.S.I.A., Madrid 28031, Spain)

In this abstract a low-cost sound level meter is developed with the help of digital signal processing techniques and an integrated programming environment. The instrument is based on the use of a general purpose data acquisition PC card to acquire the signal and a powerful digital signal processing algorithm based on digital filters to compute a one-third octave band analysis conforming to ANSI S1.11-1986 in real time. The objective is made possible thanks to the use of optimized signal processing routines from Intel Signal Processing Library. The use of this library and an instrumentation specific integrated programming environment such as LabWindows/CVI (National Instruments) makes it possible to design a low-cost instrument in a very short time. The algorithm is a multirate filter bank implementation of the standard. The discussion is completed with several simulations and benchmarks are presented for different design situations. Execution time is also compared to that obtained with a popular digital signal processor (TMS820C30 from Texas Instruments). [For Engineering Acoustics Best Student Paper Award.]

**4pEA9. Transducer electronic data sheet implementation within traditional precision microphones and array microphone systems.** Michael J. Lally, Mike J. Dillon (The Modal Shop, Cincinnati, OH 45212), Gunnar Rasmussen, and Peter Wulf Andersen (The Modal Shop, Cincinnati, OH 45212 and GRAS Sound and Vib., Denmark)

Consistent with the IEEE1451 standard development of smart sensors, microphones have been developed that can digitally report critical transducer characteristics. This new, standardized communication ability automates communication of calibration scaling and eliminates potential book-keeping errors in large channel testing. Discussion includes the framework of the evolving standard, implementation in preamps/array microphone systems and practical examples of error reduction and time savings.

**4pEA10. Problems associated with driving an efficient Tonpilz transducer with broadband signals.** Alan H. Lubell (Lubell Labs., Inc., 21 N. Stanwood Rd., Columbus, OH 43209)

A symmetrical piezoceramic-stack-driven Tonpilz makes a very efficient underwater loudspeaker for use in any body of water where human activity takes place. The large variation of impedance with frequency makes driving this type of transducer a unique problem. Transformer leakage inductance together with stack capacitance causes excessive current to be drawn at the series resonant frequency, which must be controlled with an appropriately sized series resistor. This paper presents measured and computed frequency responses and impedance magnitude versus frequency graphs for several of Lubell Labs.' underwater acoustic transducers. Included are recent calibrations made at the USN Dodge Pond facility. These are used to determine suitable values of leakage inductance and series resistance. A reasonable size series resistance is one that will not cause excessive loss at the fundamental transducer resonance. Excessive leakage inductance must be avoided because of its effect on high frequency response. An expression for the current drawn when the transducer is driven with bandlimited pink noise has been derived by the author and will be given herein.

**4pEA11. Binaural equalization for loudspeaker systems.** Peng Wang and Wee Ser (Ctr. for Signal Processing, S2-B4-08, School of EEE, Nanyang Technolog. Univ., Singapore 639798, p145255056@ntu.edu.sg)

On-axis equalization methods for loudspeaker systems can be found everywhere. In these schemes, the axial impulse responses are taken as the objects to be equalized. However, the generated equalizers will display unsatisfactory off-axis distortions and different equalization effects at left and right ears. In this paper, a new criterion, which combines the considerations of the distortions at left and right ears and the binaural difference, is presented. A binaural equalization method applying this criterion is introduced and the comparison with normal on-axis equalization methods is also given.

**4pEA12. A novel dual-band equalizer design for loudspeaker systems.** Peng Wang, Wee Ser, and Ming Zhang (Ctr. for Signal Processing, S2-B4-08, School of EEE, Nanyang Technolog. Univ., Singapore 639798, p145255056@ntu.edu.sg)

Many equalization methods based on DSP technology have been proposed during the past decade. However, nearly all the methods demand low-frequency components to be processed separately because of deficient resolution. A dual-band equalization method for loudspeaker systems is presented in this paper. In this method, the conventional equalization based on deconvolution of the system response and a recently introduced method based on warped filter design are combined, to achieve a refined performance in the whole audio frequency region with lower order filter. Computer simulations are given to demonstrate the effectiveness.

**4pEA13. Calibration or conformance check. What is the difference.** Ernst Schonthal (The Modal Shop, 1775 Mentor Ave., Cincinnati, OH 45212)

Both measurements are performed using certified equipment. A certificate is issued after both measurements. This paper will address the difference between the two certificates and why calibration is important. Also this paper will explain some instances where one has to do both.

**Session 4pID****Interdisciplinary: Distinguished Lecture on Acoustics at the End of the 20th Century—An Overview of the State of the Art**

William J. Cavanaugh, Chair

*Cavanaugh Tocci Associates, Inc., 327F Boston Post Road, Sudbury, Massachusetts 01776***Chair's Introduction—3:25*****Invited Papers*****3:30****4pID1. Acoustics at the end of the 20th Century—An overview of the state of the art.** Malcolm J. Crocker<sup>a)</sup> (Mech. Eng. Dept., Auburn Univ., Auburn, AL 36849)

Recent years have seen rapid advances in digital computers, miniaturization of electronics, and the development of new materials. These advances have led to improved knowledge in many scientific fields including a large number of different areas of acoustics. In many cases the developments have been synergistic; new experimental knowledge has led to improved theoretical models and approaches and vice versa. This paper begins with a brief introduction reviewing some of the technological developments that are helping to accelerate the increase in knowledge in different areas of acoustics. It then continues with a general review of the state of the art in a number of specific fields such as: use of finite-element and boundary-element methods in acoustics, computational aeroacoustics, sonochemistry, thermoacoustic engines, active noise and vibration control, sound-intensity measurements, techniques of speech coding and speech recognition, ultrasonics in medical acoustics, and cochlear mechanics. Finally, some projections about future scientific and technological developments in different areas of acoustics are made. <sup>a)</sup>The author is the editor-in-chief of the recently published *Encyclopedia of Acoustics*.

**Session 4pMUa****Musical Acoustics: Music, Rhythm and Development II**

Caroline Palmer, Cochair

*Department of Psychology, The Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210*

Mari Riess Jones, Cochair

*Department of Psychology, The Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210****Contributed Papers*****1:30****4pMUa1. Perception of rhythm by children in multitimbral, multimetric contexts.** Punita G. Singh (Sound Sense, 20-A Aurangzeb Rd., New Delhi 110011, India)

Perception of meter by children aged 6 to 12 was studied using sequences of 12 tones as stimuli. Either no accents, or physical accents at positions implying triple, quadruple, or both meters were provided by changing the spectral locus of four harmonics  $n$ ,  $n+1$ ,  $n+2$ ,  $n=3$ , of tones  $L_n$  (where  $n=2$  or 6) or the rise/decay times of their temporal envelopes (95+5 ms, vs 5+95 ms). Listeners reported if they perceived a triple, quadruple, or ambiguous meter, or no accents at all. Children were

easily able to perceive the metrical structure of sequences with accents on triple or quadruple meter positions alone. Mixed meters were hard to parse in a single-timbre context. In mixed sequences with accents provided by different timbre features at quadruple and triple meter positions, listeners tended to follow the meter implied by tones L2 rather than L6. Temporal envelope variables were not effective in facilitating parsing of mixed meters. Results indicate that young listeners are similar to adults in using timbral information to parse complex meters. Lower harmonics seemed to be more effective carriers of accents than higher harmonics. Spectral resolving power may thus be an important aspect of not just pitch, but rhythm perception as well.

1:45

**4pMUa2. Rhythm, expectancy, and time judgments: Some “surprising” results.** J. Devin McAuley (Dept. of Psych., Ohio State Univ., Columbus, OH 43210, mcauley.11@osu.edu)

A series of time-judgment experiments were performed that manipulated the beginning and ending of a to-be-remembered time interval (the standard) relative to a preceding rhythmic context. Beginning and ending times of the tones marking out the standard were examined for conditions that were on-time, early, and late, relative to the implied beat of the context tones. Consistent with an entrainment model, time judgments about the standard time interval showed a “shorter” perceptual bias in early conditions and a “longer” perceptual bias in late conditions, and an overall expectancy profile, with better performance observed when the standard was on-time than when it was early or late. The most interesting feature of the data was an unexpected interaction between the beginnings and ending manipulations. Interpreted with respect to the entrainment model, the interaction suggests that a tracking oscillator’s phase is more affected by unexpected beginnings, whereas its period is more affected by unexpected endings.

2:00

**4pMUa3. Simulating meter perception in acoustic signals.** Edward W. Large (Ctr. for Complex Systems, Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431)

A model of meter perception is presented in which an acoustic signal provides parametric input to a pattern-forming dynamical system. Under rhythmic stimulation, the system undergoes bifurcations that give rise to patterns of self-sustained oscillations. The temporal structure of these patterns reflects the perceived temporal organization of acoustic signals that has been described by both linguists and music theorists [M. Liberman and A. M. Prince, *Ling. Inq.* **8**, 249–336 (1977); F. Lerdahl and R. Jackendoff, *Generative Theory of Tonal Music* (MIT, Cambridge, 1983)]. These patterns are stable, yet flexible: They can persist in the absence of input and in the face of conflict, yet they can reorganize given a strong indication of a new temporal structure. Both continuous and discrete time models will be discussed and their application to acoustic signals will be demonstrated.

2:15

**4pMUa4. Motor learning, meter, and rhythm in music performance.** Rosalee K. Meyer (Psych. Dept., Ohio State Univ., Columbus, OH 43210)

What do musicians learn when they perform music: motor-independent knowledge specifying abstract concepts or motor-specific knowledge designating motor movements? Whereas some theories of se-

quence production suggest that timing is motor independent, others have suggested a closer relationship between timing and motor movements. Previous research suggests that skilled pianists’ knowledge of how to perform simple melodies is motor independent [C. Palmer and R. K. Meyer, *Psychol. Sci.* (in press)]. However, the relationship of motor movements to temporal structures such as meter (periodic accent structure) and rhythm (duration pattern) were not addressed in this study. Two experiments reported here investigate whether skilled pianists’ knowledge of meter and rhythm is motor independent or motor specific. In a transfer of learning paradigm, pianists performed a short musical piece under speeded conditions and then transferred to another musical piece with identical pitches but the same or different motor movements. In experiment 1, the transfer melody contained the same or different meter, and in experiment 2, the transfer melody contained the same or different rhythm. Changes in total sequence duration indicate the importance of motor-specific knowledge in the context of music that differs in temporal dimensions such as meter and rhythm. [Work supported by NIMH.]

2:30

**4pMUa5. Timing relationships in rhythmic tapping and music performance with delayed auditory feedback.** Steven A. Finney (Dept. of Psych., Ohio State Univ., 142 Townsend Hall, 1885 Neil Ave., Columbus, OH 43210, finney.17@osu.edu)

Delayed auditory feedback (DAF) disrupts performance in speech, music, and rhythmic tapping, with a characteristic symptom in all domains being insertions or repetitions of material (resembling stuttering). One recent attempt to explain DAF impairment, the Node Structure Theory [D. G. MacKay, *The Organization of Perception and Action* (Springer, New York, 1987)] has proposed that such repetitions are due to the auditory feedback causing reactivation of the node which produced that feedback, i.e., the sound is a stimulus which precedes (and causes) the repetition. The current research investigated the effects of various auditory delays on tapping simple rhythms, and analyzed the precise temporal relationships between sounds and taps. For the delays causing the most impairment, sound onset was found to follow rather than precede the onset of inserted taps, counter to the requirements of the Node Structure Theory. DAF impairment in a previous experiment in musical performance [S. A. Finney, *Music Perception* **15**, 153–174 (1977)] was reanalyzed from this perspective, and alternative explanations of the repetition behavior are discussed.

2:45–3:00

### Panel Discussion



**Session 4pMub****Musical Acoustics: The Ohio State University Marching Band Performance**

James M. Pyne, Chair

*The Ohio State University, 316 Weigel, 1866 College Road, Columbus, Ohio 43201***Chair's Introduction—4:25**

THURSDAY AFTERNOON, 4 NOVEMBER 1999

UNION E ROOM, 1:30 TO 2:50 P.M.

**Session 4pNSa****Noise: Status of Noise Regulations**

Bennett Brooks, Chair

*Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, Connecticut 06066***Invited Paper****1:30****4pNSa1. Local ordinance targeted to low-frequency noise.** Bennett Brooks (Brooks Acoustics Corp., 27 Hartford Turnpike, Vernon, CT 06066)

The local authorities of a suburban town wanted to control the low-frequency noise emissions due to musical entertainment at night clubs and taverns. Complaints from residents about booming bass guitar and drum sounds had to be balanced with the need for tavern owners to reasonably conduct business. It was recognized that the existing town code, which placed limits on A-weighted noise levels at property boundaries, was not an effective means to control low-frequency noise emissions. A revised ordinance was developed to address this issue, which now places additional limits on the allowable noise levels in specific octave bands. In particular, this includes the low-frequency (bass) octave bands. The octave band limits were selected such that the summation of band levels equals the A-weighted overall noise limit.

**Contributed Paper****1:50**

**4pNSa2. Prediction and attenuation of noise resulting from construction activities in major cities: The case for modernizing construction noise codes.** Daniel R. Raichel (Dept. of Mech. and Aero Eng. and the School of Architecture and Environ. Studies, City College of CUNY, New York, NY 10031) and Michael Dallal (Cooper Union, New York, NY 10003)

Noise arising from operation of construction machinery in major cities is generally most annoying at street and below-ground levels, where the loudest equipment is used. Both residential and business neighborhoods are adversely affected by the presence of excessive noise, particularly if construction activity occurs during early hours. In some cases, vehicular traffic patterns become altered with a resultant change in the traffic noise levels. A survey was conducted as to which type of machinery is most commonly used in New York City, the amount of noise pollution generated by specific equipment, and commercially available means of attenuating the noise, preferably through retrofitting of existing equipment. On-site sound-level measurements were taken and the results compared with published data. Methods of predicting noise levels through the use of computer models, based on the effect of surrounding topology, were also evaluated. Noise codes in several major cities in the United States and abroad were compared with respect to their compliance with U.S. federal regulations. This was done for the purpose of developing a proposal for a

new set of construction noise regulations for the City of New York. [Work supported by the New York City Department of Design and Construction.]

**2:05**

**4pNSa3. CNEL at two sites in Del Mar, due to a hundred helicopter flybys.** Robert W. Young (1696 Los Altos Rd., San Diego, CA 92109)

In the Settlement Agreement in early 1999 of a lawsuit about helicopter noise, in which the City of Del Mar was one of the plaintiffs and the Marine Corps was one of the Federal defendants, there was a requirement that the community noise equivalent level (CNEL) be included as a standard for measuring noise significance. This is a preliminary report about CNEL measured continually at two sites in south Del Mar, California. In both cases the residents had filed official complaints about the noise of helicopters. These residences are distance from heavily traveled streets and highways. The automatic noise monitor consists of a Computer Engineering, Ltd., 493-2 integrating sound-level meter feeding a CEL-238 secondary processor, all in a weather-proof case. The outfit prints, and puts in memory, the one-hour average sound level during a preceding hour and at the end of a day the 24-hour average sound level (24 HL), the day-night average sound level (DNL), and the CNEL. At 22839 Via Grimaldi, Friday to Wednesday 5–10 May 99, the CNELs were 55.8, 55.5, 54.6, 55.9, 58.7, and 58.7 dB. At 421 Ocean Drive, Tuesday to Saturday 11–15 May 99, the CNELs were 58.4, 56.5, 59.6, 58.6, and 60.5 dB.

**4pNSa4. A radiosity-based model for simulating sound propagation in urban streets.** Jian Kang (The Martin Ctr., Cambridge Univ., 6 Chaucer Rd., Cambridge CB2 2EB, UK)

To study the fundamental characteristics of the sound field in urban streets resulting from diffusely reflecting boundaries, a theoretical/computer model has been developed using the radiosity technique. The model divides building facades and ground into a number of patches and then simulates the sound propagation in a street by energy exchange between the patches. Computations in a typical street with a single noise source show that with diffusely reflecting boundaries the sound attenuation along the length is significant, and the sound distribution in a cross section is generally even unless the cross section is very close to the source. The effectiveness of some architectural changes and urban design options on further increasing sound attenuation along the length, such as by strategically arranging buildings in a street or adding absorption patches on facades or ground has been analyzed. It has also been demonstrated that by replacing geometrically reflecting boundaries with diffusely reflecting boundaries, the sound attenuation along the length becomes greater, and the reverberation time becomes shorter. This suggests that from the viewpoint of urban noise reduction, it is better to design the street boundaries as diffusely reflective rather than acoustically smooth. [Work supported by the Lloyd Foundation.]

**4pNSa5. Propagation of noise from petrochemical activities in the Taranaki region of New Zealand.** Daryl Prasad (Acoust. Res. Ctr., Architecture Bldg., The Univ. of Auckland, Private Bag 92019, Auckland, NZ, d.prasad@auckland.ac.nz)

Noise is a significant environmental concern facing the petrochemical industry in New Zealand. Typical inland oil and gas exploration involves operations that can run continuously for 70 days. Exposure to the resulting steady and impulsive noise produces undue stress and loss of amenity to neighboring residents. This paper reports on an investigation into the propagation of noise in the Taranaki region of New Zealand. Of the many factors that influence the propagation from drilling, atmospheric conditions (especially wind, temperature, and humidity) create the most variability. This study aims to provide a better understanding into the way noise is propagated in this region. Results of wind and temperature profile measurements will be presented as well as the related frequency and seasonality of events such as temperature inversions. Actual noise propagation data combined with the meteorological data will be used to test a model developed for the prediction of noise levels based on the CONCAWE and ISO 9613-2 methods of calculation.

THURSDAY AFTERNOON, 4 NOVEMBER 1999

UNION E ROOM, 3:05 TO 4:20 P.M.

### Session 4pNSb

#### Noise: Hearing Protection

Elliott H. Berger, Chair

*E-A-R/Aearo Company, 7911 Zionsville Road, Indianapolis, Indiana 46268-1657*

#### Contributed Papers

3:05

**4pNSb1. Occupational and leisure noise exposures as risk factors for adult hearing impairment: The Nord Trondelag (NT) Hearing Study.** Howard J. Hoffman, Robert H. MacTurk, George W. Reed (Epidemiology, Statistics & Data Systems Branch, NIDCD, NIH, Bethesda, MD 20892), Kristian Tambs, Jostein Holmen (Natl. Inst. of Public Health (Folkehelsa), Oslo, Norway), and Hans M. Borchgrevink (The Natl. Hospital, Oslo, Norway)

In the NT Hearing Study (1995–1997), air-conduction thresholds were obtained using sound attenuation booths. This report is based on 27 997 subjects selected in a nested, case-control study to receive follow-up questionnaires on work and leisure noise exposures. Cases with low-frequency hearing loss (HL)—speech frequencies 500, 1000, and 2000 Hz—were defined by a pure-tone average (PTA) exceeding 30 dB in the better ear. A second case series with only high-frequency HL was defined by PTA exceeding 30 dB at 3000, 4000, and 6000 Hz. Controls were defined by PTAs less than 30 dB in both ears. Multivariate logistic regression analyses were stratified by sex and age. Family history of HL and recurrent ear infections in childhood were significant risk factors across all sex and age strata. Among men under 50, an odds ratio (OR) of 1.87 [95% confidence interval (CI): 1.53–2.28] for high-frequency HL was found for loud noise at work, an OR=1.35 (CI: 1.08–1.69) for use of noisy equipment, and OR=1.09 (CI: 1.02–1.17) for hunting/sport shooting. Low-frequency HL was related to occupational noise exposures in older (50+ years), but not younger men. Tinnitus was associated with occupational noise exposure at all ages.

3:20

**4pNSb2. Noise dosimetry for telephone users.** Philippe Moquin (Mitel Corp., 350 Legget Dr., Kanata, ON K2K 1X3, Canada, philippe\_moquin@mitel.com)

Due to recent F.C.C. regulations, all telephones in North America will be required to have at least 12 dB of amplification. While this is a definite asset to the hard of hearing community, there are fears that normal hearing workers who use the telephone most of the day could suffer noise-induced hearing loss. Theoretical calculations validate this hypothesis. Measurements of actual levels in a call center were taken. Finally, a method of providing noise dosimetry within a digital telephone set is presented. Simultaneous measurements show good correlation.

3:35

**4pNSb3. Progress on a rating system for hearing protector attenuation.** William J. Murphy and John R. Franks (Bioacoustics and Occupational Vib. Section, NIOSH MS C-27, 4676 Columbia Pkwy., Cincinnati, OH 45226-1998)

The cumulative distribution of A-weighted noise reduction (CDNRA) has been proposed as an alternative to the noise reduction rating for hearing protectors [W. J. Murphy and J. R. Franks, *J. Acoust. Soc. Am.* **105**, 1131 (1999)]. Individuals' REAT data were A-weighted and summed across frequencies to determine the statistical distribution. The cumulative distribution of REAT data was fit to a combination of two logistic curves to capture the features of well-fit and poorly-fit hearing protectors. Previously, the CDNRA model was shown to yield Pearson's correlation coef-

ficients of 0.99 or better. In this paper, the CDNRA theory will be reviewed and the method will be applied to the subject-fit data from six HPDs. The Bilsom UF-1 earmuff was rated at 23-dB attenuation. The V-51R was rated at 6-dB attenuation. The EP100 was rated at 7-dB attenuation. And the EAR Classic was rated at 23-dB attenuation. Subject-fit data from the EAR Express and Howard Leight Max earplugs will be presented as well.

3:50

**4pNSb4. Evaluation of a FitCheck hearing protector test system.** William J. Murphy, John R. Franks, and Dave A. Harris (Bioacoust. and Occupational Vib., NIOSH, 4676 Columbia Pkwy., Cincinnati, OH 45226-1998)

The measurement of attenuation provided by a hearing protector is an important element in the development of an effective hearing-loss prevention program. Three systems were evaluated for accuracy compared with a sound-field test of real-ear attenuation at threshold (REAT) [J. R. Franks, J. Acoust. Soc. Am. **105**, 1129 (1999)]. The FitCheck method used with headphones was the most feasible system to use for field-test situations. However, the FitCheck method produced REAT values in the low frequency that were less than those produced by the sound-field method when used with the EAR Express earplug. A subsequent study has been conducted which evaluated the EAR Classic and the Howard Leight MAX earplugs with the FitCheck and sound-field methods. Twenty-four subjects were instructed to fit the devices without experimenter intervention (*re*: ANSI S12.6 Method B subject-fit protocol). Subjects were not allowed to adjust the earplugs between test methods. Mean REAT data from both

methods will be compared. Preliminary data suggest that the FitCheck method yields lower REAT estimates for frequencies below 1 kHz, but the differences were not statistically significant.

4:05

**4pNSb5. Changes in attenuation in personal hearing protector over time during use.** Michael P. Valoski and George Durkt, Jr. (MSHA, Tech. Support, P.O. Box 18233, Pittsburgh, PA 15236)

Ideally, implementing engineering controls reduces workers noise exposures below the permissible exposure level (PEL). However, some occupations continue to have exposures above the PEL. To protect the hearing sensitivity of these workers, personal hearing protectors (PHP) must be worn. Manufacturers test PHPs under ideal conditions in a laboratory. For these tests, highly motivated and trained subjects remain stationary. Furthermore, a researcher helps the subject fit the PHP. Shortly after determining that the PHP was properly fitted, the tests are conducted. Under these conditions the optimum attenuation afforded by the PHP is achieved. Mining conditions differ substantially from those under which the laboratory attenuations are determined. Often, workers wear their PHPs for extended periods without removal. Moreover, a worker's constant movement may disrupt the seal of a PHP. Once the seal is broken, the PHP provides degraded performance. This project examines any change in earmuff attenuation over a 3-h wearing time in a laboratory. After donning an earmuff unaided, a subject wore the earmuff for 3 h while walking and stopping to lift containers from the floor to a table in a large reverberation chamber. It was found that the PHP's attenuation did not change substantially during the test.

THURSDAY AFTERNOON, 4 NOVEMBER 1999

MADISON ROOM, 1:30 TO 5:45 P.M.

## Session 4pPA

### Physical Acoustics: Thermoacoustics

Robert A. Hiller, Cochair

*Department of Physics and Astronomy, University of Mississippi, University, Mississippi 38677*

Philip S. Spoor, Cochair

*Los Alamos National Laboratory, MST-10, MS K764, Los Alamos, New Mexico 87545*

### Contributed Papers

1:30

**4pPA1. Nonresonant referenced laser-induced thermal acoustics (LITA) for measurement of the speed of sound in gases.** Allan J. Zuckerwar, Roger C. Hart, Jeffrey Balla, and Gregory C. Herring (NASA Langley Res. Ctr., M.S. 236, Hampton, VA 23681)

Nonresonant laser-induced thermal acoustics (LITA) is an effective tool for measuring the speed of sound in gases and is suitable for operation at elevated temperatures. Counterpropagating sound waves of known, fixed wavelength are generated electrostrictively by crossing two split laser beams which are generated from a short-pulse pump laser. The sound waves form a Bragg grating, which is illuminated by a second, long-pulse probe laser. A small fraction of the probe beam is diffracted to a detector, which permits accurate measurement of the sound frequency. Since the method comprises a free-field measurement, the corrections required of a conventional acoustic interferometer or resonator are not necessary here. The presentation will include the principle of operation of LITA, comparison with other methods of measuring the speed of sound in gases, and sample measurements at temperatures up to 650 K.

1:45

**4pPA2. Attenuation in tertiary gas mixtures with two relaxing components.** Yefim Dain and Richard M. Lueptow (Dept. of Mech. Eng., Northwestern Univ., Evanston, IL 60208)

Vibrational relaxation accounts for absorption and dispersion of acoustic waves in gases that can be significantly greater than the classical absorption mechanisms related to shear viscosity and heat conduction. This vibrational relaxation results from retarded energy exchange between translation and intramolecular (vibration) degrees of freedom. Theoretical calculation of the vibrational relaxation time of gases based on the theory of Landau and Teller [Phys. Z. Sowjetunion **10**, 34 (1936)] and Schwartz *et al.* [J. Chem. Phys. **20**, 1591 (1952)] has been applied at room temperature to the polyatomic mixtures of nitrogen with water vapor and methane as relaxing components. The theory accounts for vibrational-vibrational coupling of all constituents together with vibrational-translational energy transfer. For negligible concentrations of methane and small vibrational-vibrational coupling the analytical expression matches previous theoretical and experimental results for the binary nitrogen-water mixture. For tertiary nitrogen-water-methane mixtures, the humidity and methane dependence of the nitrogen relaxation frequency differs from that of binary

mixtures of either component with nitrogen due to coupling between the components. The significance of vibrational–vibrational coupling of binary pairs in the tertiary mixture has been analyzed.

2:00

**4pPA3. Time-of-flight measurements on high-frequency surface acoustic phonons in silicon.** D. M. Photiadis (Naval Res. Lab., Washington, DC) and J. Ding (ThermaWave, Inc., San Francisco, CA)

Results from time-of-flight measurements of high-frequency surface acoustic phonons (50–500 GHz) in (001) silicon wafers at low temperatures are reported. Pulsed laser excitation is used for the generation of the high-frequency surface acoustic phonons and aluminum edge bolometers are used as detectors. The times defined by the fast initial response of the edge bolometer correspond to surface acoustic delays in silicon. Late time responses consistent with echoes from the back face of the wafer were also observed. These results and the accompanying analysis will be discussed. [This research was supported by the Office of Naval Research.]

2:15

**4pPA4. Acoustics on the planet Mars: A preview.** Victor W. Sparrow (Grad. Prog. Acoust., Penn State Univ., 157 Hammond Bldg., University Park, PA 16802, sparrow@helmholtz.acs.psu.edu)

Assuming no operational difficulties, NASA's Mars Polar Lander will settle on the Martian surface in December 1999 and will transmit back to Earth sound samples from an audio frequency microphone designed and built by the University of California Space Sciences Laboratory and funded by The Planetary Society. Information on this MARS MICROPHONE is available on the World Wide Web at [plasma2.ssl.berkeley.edu/marsmic/](http://plasma2.ssl.berkeley.edu/marsmic/) and [planetary.org](http://planetary.org). It is hoped that the microphone will detect natural environmental sounds on Mars as well as monitor operational noises made by the lander itself. This talk presents what little is known about the acoustic environment of Mars. Using available knowledge of the Martian surface atmosphere, the acoustics between Earth and Mars can be easily compared. Although the sound speeds on Earth (343 m/s) and Mars (228 m/s) are comparable, Mars' ambient density is only 1.4% of that on Earth. The resulting difference in characteristic impedances implies that a machine vibrating on Mars will radiate sound at a level approximately 20 dB lower than on Earth. Additional comparisons will be presented. [The helpfulness of Dr. Janet Luhmann and Dr. Greg Delory of the Berkeley SSL is appreciated.]

2:30

**4pPA5. Design and construction of a small thermoacoustic refrigerator.** Tamra S. Underwood, Dana M. Smith, and Ralph T. Muehleisen (Dept. of Civil, Environ., and Architectural Eng., Univ. of Colorado, Boulder, CO 80309)

While there has been much recent work on the development of medium to large scale (hundreds to thousands of watts) thermoacoustic devices, there has been relatively little work done on small thermoacoustic devices. Larger devices are often desired in order to get large cooling powers, but such devices can be expensive and difficult to construct. Small devices on the other hand, can be both inexpensive and easy to construct. To more carefully investigate small thermoacoustic devices, a refrigerator approximately 1 in. in diameter and 5 in. long has been constructed with a target of 10 W of cooling power and a temperature span of 25 °C. To achieve higher cooling powers, several of the devices can be operated in parallel. The performance of the actual device and the comparison to the predicted values will be presented. [Work supported by the Office of Naval Research.]

3:00

**4pPA6. Thermal performance of heat exchangers for thermoacoustic refrigerators.** Yuwen Chen and Cila Herman (Dept. of Mech. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218, herman@titan.me.jhu.edu)

Heat exchangers in thermoacoustic refrigerators extract heat from the refrigerated volume and reject it to the surroundings. Eight essential heat-transfer processes coupled in the energy migration from the cold-side heat exchanger through the stack to the hot-side heat exchanger were identified, and a simplified computational model describing them was developed. Geometrical and operational parameters as well as thermophysical properties of the heat exchangers, the stack plate, and the working medium were organized into dimensionless groups. Heat transfer in the transverse direction, resulting from temperature differences between the inlet and outlet temperatures of the transport fluid in the heat exchangers, was accounted for in the simulations. Two types of boundary conditions between the thermoacoustic working fluid and the heat exchangers were considered: (1) constant temperature of the thermoacoustic working fluid over the heat exchangers with constant temperature difference along and across the stack and (2) constant heat flux from the thermoacoustic working fluid to the heat exchanger. Nonlinear temperature distributions and heat fluxes near the edge of the stack plate were observed. Effects of different parameters on the thermal performance of the heat exchangers were investigated. [Work supported by the Office of Naval Research.]

3:15

**4pPA7. Use of electrodynamic drivers in thermoacoustic refrigerators.** Ray Scott Wakeland (Grad. Prog. in Acoust., Penn State Univ., ARL, P.O. Box 30, State College, PA 16804, wakeland@psu.edu)

Some issues involved in matching electrodynamic drivers to thermoacoustic refrigerators are examined using an equivalent circuit model. Conclusions are that the driver should be chosen to have a large product  $(Bl)^2/(R_e R_m)$ . The suspension stiffness should be chosen to make the combined impedance of the mechanical and acoustical parts of the system entirely real at the operating frequency. The piston area should be selected to maximize electroacoustic efficiency or other desired parameters by matching the acoustical load to the optimum mechanical load for the particular driver. Alternately, if the piston area is fixed, the operating frequency can be adjusted to make this same match. Measurements made at this laboratory [R. W. M. Smith *et al.*, *J. Acoust. Soc. Am.* **105**, 1072(A) (1999)] provide experimental support for the conclusions of this paper concerning the attainable maximum electroacoustic efficiency. [Work supported by NSF and ONR.]

3:30

**4pPA8. Thermoacoustic properties of nonuniform stacks.** Gabriela Petculescu and Larry A. Wilen (Dept. of Phys. and Astron., Ohio Univ., Athens, OH 45701, wilen@helios.phy.ohiou.edu)

Thermoacoustic stacks of uniform cross section can be characterized completely in terms of the single thermoviscous function  $F(\lambda)$  [W. P. Arnott, H. E. Bass, and R. Raspet, *J. Acoust. Soc. Am.* **90**, 3228–3237 (1991)]. For nonuniform stacks, one must treat thermal diffusion and viscosity separately, and it may be possible to “tune” the performance of the stack by taking advantage of the extra geometric freedom. Measurements have been performed to determine the thermoviscous characteristics of pin-like stacks for which the elements are oriented either along, or perpendicular to, the temperature gradient. The measurements characterize terms in the thermoacoustic equations responsible for thermal dissipation and thermoacoustic gain. The results for pins of varying size, shape, and spacing will be discussed. Time permitting, applications of novel designs to compact lumped element oscillators will be demonstrated. [Work supported by the Office of Naval Research.]



**4pPA9. An acoustic impedance measurement of a parallel-pore stack used to determine the thermoviscous functions of the stack.** Timothy G. Simmons, Richard Raspet (Dept. of Phys. and Astron., and Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, tgsimmon@olemiss.edu), and Jeremy Brewer (Rhodes College, Memphis, TN 38112)

Recently, thermoviscous functions have been measured for single pores of various cross-sectional geometry [L. A. Wilen, *J. Acoust. Soc. Am.* **103**, 1406–1412 (1998)]. This investigation consists of measuring the acoustic impedance through a stack with parallel pores, inverting to yield the thermoviscous functions  $F(\lambda)$  and  $F(\lambda_T)$ , and determining the accuracy with which the inversion can be accomplished. The measurement shall be performed using a standard impedance tube with the typical measurements taken, i.e., pressure maximum, pressure minimum, and the positions of two pressure minima. In order to be able to solve for the thermoviscous functions, a second measurement must be taken with the stack positioned away from the rigid termination, leaving a region of air. The impedance translation theorem (modified to account for the stack porosity) will have a functional form containing both  $F(\lambda)$  and  $F(\lambda_T)$  and shall be employed for both stack placements to provide two equations which can be solved for the two unknowns,  $F(\lambda)$  and  $F(\lambda_T)$ , for a given frequency. If the method proves successful, modifications shall be made to the experimental setup so that it can accommodate a temperature gradient imposed across the stack. [Work supported by ONR.]

4:00

**4pPA10. Thermoacoustics of moist air with wet stacks.** William V. Slaton and Richard Raspet (Natl. Ctr. for Physical Acoust., University, MS 38677, wmslaton@meta3.net)

Recent work in porous media has led to an analytical treatment of sound in wet-walled tubes [Raspet *et al.*, *J. Acoust. Soc. Am.* **105**, 65–73 (1999)]. This work has been extended to examine thermoacoustic refrigerators which utilize moist air at atmospheric pressure as the working fluid as described by Hiller [Thermoacoustic Review Meeting 1999, National Center for Physical Acoustics]. In the described system the moisture in the air stream is cooled below its dew point and appears as either liquid on the stack surface or as a fog. This physical system can be adequately described by ignoring thermal diffusion of the water vapor (Soret effect) and neglecting temperature variations in the tube wall. The effects on sound propagation and efficiency due to the presence of liquid water in the system described above will be discussed. [Work supported by ONR.]

4:15–4:30 Break

4:30

**4pPA11. Experimental investigation of the acoustical properties of narrow tubes filled with a mixture of gas and condensable water vapor.** D. Felipe Gaitan, Richard Raspet, Craig J. Hickey, and William Slaton (Univ. of Mississippi, NCPA Coliseum Dr., University, MS 38677)

It is generally believed that using the latent heat of condensable vapors will increase the efficiency and energy density of thermoacoustic refrigerators. This phenomenon has been recently considered in the literature [Raspet *et al.*, *J. Acoust. Soc. Am.* **105**, 65 (1999)] in which the equations describing the acoustic properties of wet pores are described. In order to investigate some of the assumptions made in these equations (e.g., the boundary conditions at the walls of the wet pore), we have measured the acoustic damping in a small (1.3-mm-radius) tube filled with a mixture of gas and condensable water vapor. The explored parameter space included ambient pressure (0.1–2 bars), temperature (20–100 °C), and acoustic frequency (2–10 kHz). Results of these measurements will be reported. [Work supported by the Office of Naval Research.]

**4pPA12. Heat transport by acoustic streaming in a resonance tube.** George Mozurkewich (Ford Res. Lab., Maildrop 3028/SRL, P.O. Box 2053, Dearborn, MI 48121-2053)

The steady fluid flow known as acoustic streaming is capable of transporting heat advectively. This heat-transport mechanism has occasionally been applied for beneficially cooling a warm object, but in other contexts (such as thermoacoustic cooling) its effects can be detrimental. Gopinath and Mills have computed heat transport by streaming between two plane sections of a cylindrical resonator with insulating walls. In contrast, the experiments described here treat the interaction between streaming gas and thermally conductive cylinder walls. The walls of a horizontally oriented resonance tube were instrumented with a heater and more than 20 thermocouples to permit quantification of axial conduction through the solid walls and radially outward natural convection. By applying conservation of energy to the tube walls, the local, radially inward heat current,  $q(x)$ , was deduced. The quantity  $q(x)$  is related to streaming transport within the tube. The measurements delineate regions where heat is added to and removed from the streaming gas. Although the magnitude of  $q$  initially increases with increasing acoustic amplitude, it quickly saturates. The fluid-dynamical implications of amplitude-independent  $q(x)$  will be discussed.

5:00

**4pPA13. Streaming in thermoacoustic engines.** Gordon Smith, Richard Raspet, and Henry Bass (Natl. Ctr. for Physical Acoust., Coliseum Dr., University, MS 38655)

Thermoacoustics utilizes an acoustic process to transport heat. Interestingly enough, sound sources that move essentially in a sinusoidal manner often do not yield flow fields of a sinusoidal nature. In addition to the oscillation of each fluid element contributing to the acoustic system, there can exist a pattern of steady vortices or other time-independent circulation regions known as acoustic streaming. Many different models for acoustic streaming have been developed, each stemming from a unique specific premise to the problem. Within a thermoacoustic device, acoustic streaming provides a means for a thermal short, in which a direct fluid flow would be established between two heat exchangers, permitting a much easier path for thermal transport than that established by a thermoacoustic process. However, little empirical data exist to provide a means of estimating the amount of thermal loss from such a process. Theoretical work to date will be presented, as well as experimental data from a single stage prime-mover thermoacoustic device. [Work supported by ONR.]

5:15

**4pPA14. Characterization of the nuisance heat load in a thermoacoustic refrigeration demonstration device.** Ray Scott Wakeland (Grad. Prog. in Acoust., Penn State Univ., ARL, P.O. Box 30, State College, PA 16804, wakeland@psu.edu)

The heat load in a thermoacoustic refrigeration demonstration device is studied. In the demo, which acts only to create a low temperature at the position of a thermocouple, the only heat load is nuisance heat, which is not well known. The method described here for estimating experimentally the size of this heat load assumes that heat transport associated with acoustics depends on acoustic amplitude  $p$  and on the magnitude of the deviation from ambient temperature  $|\Delta T|$ , but is independent of the sign of  $\Delta T$ . That is, the heat load due to acoustics-induced transport is taken to have the form  $\dot{Q}_{\text{transport}} = -F(p)\Delta T$ , where  $F(p)$  is some function of  $p$  only, and  $\Delta T$  can be either positive (refrigerator) or negative (heater). Direct thermoviscous heating at the site of the thermocouple  $\dot{Q}_{\text{TV}}(p)$  is also a function of  $p$  only. At steady state,  $\dot{Q}_{\text{heater}} = F(p)\Delta T - \dot{Q}_{\text{TV}}(p)$ . In the experiment, the refrigeration stack is replaced by an electric heater, the power of which is easily measured. The steady-state temperature rise above ambient  $\Delta T$  is measured as a function of heater power and acoustic amplitude. The experiment generates a family of straight lines, the slopes of which determine the function  $F(p)$ , and the intercepts give  $\dot{Q}_{\text{TV}}$ . [Work supported by ONR and NSF.]

5:30

**4pPA15. Separation of gas mixtures in the acoustic boundary layer.** Philip S. Spoor and Gregory W. Swift (Condensed Matter and Thermal Phys. Group MST-10, Los Alamos Natl. Lab., Los Alamos, NM 87545)

Spatial separation of binary gas mixtures in the presence of a sound wave has been reported by various experimenters in the last half-century,

and is usually attributed to static pressure or temperature gradients. The authors suggest an additional mechanism, a "concentration pumping" in the acoustic boundary layer analogous to thermoacoustic heat pumping. Data from a simple acoustic mixture-separation experiment will be compared with predictions of this new theory. Observation of mixture separation between two acoustically coupled thermoacoustic engines will also be discussed.

THURSDAY AFTERNOON, 4 NOVEMBER 1999

MARION ROOM, 2:00 TO 4:30 P.M.

### Session 4pPP

## Psychological and Physiological Acoustics: Pitch and Complex Stimuli

Gary R. Kidd, Chair

*Department of Speech and Hearing Science, Indiana University, Bloomington, Indiana 47405*

### Contributed Papers

2:00

**4pPP1. Pitch of envelope-modulated rippled noise.** Stanley Sheft and William A. Yost (Parmlly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, ssheft@luc.edu)

Both amplitude modulation (AM) and the delay-and-add process of rippled-noise (RN) generation can impart a pitch to wideband Gaussian noise. For AM, and possibly also RN, the basis of the pitch percept derives from temporal processing of the stimulus. In the present study, the pitch of envelope-modulated rippled noise was matched to filtered pulse trains. RN was generated with a delay of either 5 or 10 ms with either one or eight iterations and a gain of either plus or minus one. AM rates ranged from 50 to 200 Hz. Without rippling the carrier, accuracy in pitch matching to AM rate was fairly consistent over the range of rates studied with most deviations occurring as "octave errors." With envelope modulation of RN carriers, pitch matches were distributed between pitches associated with the AM rate and fine-structure delay, in some cases with a noticeable increase in "octave errors." Even with eight iterations in RN generation, an influence of AM rate on pitch matches was obtained. Pitch-match distributions can be roughly accounted for by consideration of the largest positive peak of both the fine-structure and envelope autocorrelation functions. [Work supported by NIH and AFOSR.]

2:15

**4pPP2. Pitch strength of regular interval stimuli as a function of duration and filtering.** William A. Yost (Parmlly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

The ability of listeners to discriminate between a random interval noise and a regular interval stimulus, iterated rippled noise (IRN), was measured as a function of duration and bandpass filtering. IRN is generated by delaying (by  $d$  ms) a noise and adding it back to the undelayed noise after attenuation ( $g$ ) in an iterative process (where  $n$  is the number of iterations). A two-alternative, forced-choice adaptive procedure was used in which stimulus duration was varied adaptively. Different values of  $g$ ,  $d$ ,  $n$ , and filter condition were used. In general, the duration required to discriminate IRN from regular interval noise increased as the filter condition provided less low-frequency energy. Duration thresholds were higher for high values of attenuation. The smallest duration thresholds were obtained for  $d$  in the range of 2–5 ms. The results will be discussed in terms of temporal models of the pitch of complex stimuli. [Work supported by NIDCD and AFOSR.]

2:30

**4pPP3. Finding the missing fundamental: A connectionist model of harmonic detection.** Clifford F. Lewis, Stephan B. Fountain (Dept. of Psych., Kent State Univ., P.O. Box 5190, Kent, OH 44242-0001, clewis@kent.edu), and Michael K. McBeath (Arizona State Univ., Tempe, AZ 85287-1104)

A major shortcoming of the place-coding principle of pitch perception has been that it does not account for the phenomenon of virtual pitch, or the missing fundamental, because there is no physical energy on the basilar membrane at the location corresponding to the fundamental frequency. A connectionist model provides an example of a simple mechanism for virtual pitch that is consistent with place-coded input. An autoassociator network was trained using place-coded amplitudes of harmonics that corresponded to tones produced by a variety of musical instruments. After training, the network exhibited the missing fundamental illusion. Specifically, activation was present in the output unit corresponding to the fundamental frequency, even when the input energy at the fundamental was removed. This activation was approximately an order of magnitude larger than the activation of other units receiving no input energy, and it was the same order of magnitude as the activation when the fundamental frequency was present. The model demonstrates how virtual pitch can result from learned associations between harmonics that typically occur simultaneously in nature. When one of the harmonics is missing, the network simply fills in the missing part of the pattern.

2:45

**4pPP4. On loudness and pitch of "complex noise" in relation to the factors extracted from the autocorrelation function.** Shin-ichi Sato, Hiroyuki Sakai, and Yoichi Ando (Grad. School of Sci. and Technol., Kobe Univ., Rokkodai, Nada, Kobe, 657-8501 Japan)

In a previous study, it was shown that the loudness of bandpass noise within the critical band changes increasing the effective duration of the envelope of the normalized autocorrelation function (ACF)  $\tau_e$  changing the filter slope of bandpass noise and the subsequent reverberation time  $T_{\text{sub}}$  of sound field [Y. Ando, *Architectural Acoustics—Blending Sound Sources, Sound Fields, and Listeners* (AIP Press/Springer-Verlag, New York, 1998), Chap. 6]. In this study, the loudness of the complex signals including bandpass noises with different frequency (complex noise) is examined. The complex noises consist of the bandpass noises whose center frequencies are the harmonics of the fundamental frequency. The bandwidth of each component is changed within the critical band by use of a sharp filter of more than 1000 dB/octave. Subjects judge the loudness of complex noise by comparing the tone of its perceived pitch frequency, the "missing fundamental." The results are discussed with the four factors

extracted from the autocorrelation function, which are  $\tau_c$ , the energy represented at the origin of the delay  $\Phi(0)$ , and the amplitude and the delay time of the first peak of ACF  $\phi_1$  and  $\tau_1$ . [Work supported by ACT-JST and Grant-in-aid for Scientific Research, JSPS, Japan.]

3:00

**4pPP5. Absolute pitch is demonstrated in speakers of tone languages.** Diana Deutsch (Dept. of Psych., Univ. of California, San Diego, La Jolla, CA 92093), Trevor Henthorn (Univ. of California, San Diego, La Jolla, CA 92093), and Mark Dolson (E-mu Creative Technol. Ctr., Scotts Valley, CA 95067)

In two experiments, Vietnamese and Mandarin speakers manifested a precise form of absolute pitch in reading lists of words. In one experiment, seven Vietnamese speakers read out a list of ten words twice, on two separate days. The average pitch of each word was determined by computer analysis, and for each speaker comparisons were made between the average pitches produced by the same word on different days. Remarkable correspondences were obtained: For most speakers, the signed pitch difference, averaged across words, was well within a semitone. In another experiment, 15 Mandarin speakers read out a list of 12 words twice within a session, with the 2 readings separated by roughly 20 s; and in 2 sessions which were held on different days. Again, comparing across sessions, for most speakers the signed pitch difference, averaged across words, was well within a semitone. Furthermore, the differences were no greater when the pitches were compared across sessions than within sessions. These findings demonstrate that tone language speakers are able to refer to a remarkably precise and stable absolute pitch template in producing words. Relationships to other work on absolute pitch are discussed, and the findings are illustrated by sound recordings.

3:15

**4pPP6. Infants' discrimination of spectral slope.** Christine D. Tsang and Laurel J. Trainor (Dept. of Psych., McMaster Univ., Hamilton, ON L8S 4K1, Canada)

Spectral slope is a global property of the spectral envelope representing the change in energy across spectral frequency. Spectral slope is a dimension for discriminating speech, voices and musical instrument timbres. The average spectral slope of speech and music is about  $-6$  dB/octave. If spectral slope is important in timbre perception, enhanced sensitivity for spectral slopes around  $-6$  dB/octave would be expected, and poor discrimination would be expected for spectral slopes very different from this value. Eight-month-old infants were tested across a range of spectral slopes ( $-16$  to  $+16$  dB/octave) to determine whether infants' slope discrimination is best near  $-6$  dB/octave. The stimuli were generated according to the following equation:  $dB_i = b \log_2 i$ , where  $i$  is the harmonic number,  $b$  is the spectral slope value, and  $dB_i$  is the relative amplitude of harmonic  $i$ . The results show that infants were only sensitive to differences in spectral slope in a limited range between  $-10$  and  $-4$  dB/octave, despite the fact that local intensity cues were smallest in this range. This suggests that infants are using spectral slope as a dimension for discriminating timbre and that the auditory system is tuned to sounds with real-world spectral slopes. [Research supported by NSERC.]

3:30

**4pPP7. Sound quality judgments of everyday sounds.** Gary R. Kidd and Charles S. Watson (Communication Disorders Technology, Inc., 501 N. Morton St., Ste. 215, Bloomington, IN 47404)

An experiment was conducted to determine the psychological dimensions that underlie listeners' judgments of the qualities of everyday sounds and to identify the stimulus properties on which such judgments are based, using Osgoods' semantic differential technique. A collection of 145 common sounds was presented to 32 listeners who rated the sounds on 20 seven-point rating scales. The sounds were of a variety of common objects and events (e.g., dishwasher, vacuum cleaner, door closing) as well as ambient sounds that sometimes involved several objects and events (e.g., traffic, cafeteria noise, rain). Rating scales included subjective adjective

pairs (e.g., happy-sad, relaxed-tense) and objective pairs (e.g., hard-soft, large-small). A principal components analysis of the rating data indicated that average judgments of the listeners were associated with four dimensions, roughly corresponding to large-small, compact-scattered, harsh-mellow, and overall quality (high-low). However, the level of agreement across subjects varied considerably for the different sounds and rating scales. The relations between several static (averaged over the entire waveform) and dynamic (time-varying) stimulus properties and listeners' judgments will be discussed. [Work supported by NIH/NIMH.]

3:45

**4pPP8. Observer weighting of monaural level information in a pair of tone pulses.** Mark A. Stellmack and Neal F. Viemeister (Dept. of Psych., Univ. of Minnesota, 75 E. River Rd., Minneapolis, MN 55455)

In a cued single-interval task, listeners were presented a pair of standard pulses (each 10 ms, 1 kHz) of fixed level followed by a pair of similar tone pulses, the comparison, separated by a fixed interpulse interval (IPI) of 2–256 ms. Listeners were instructed to attend to either the first or second pulse of the comparison and to indicate whether the target pulse was higher or lower in level than the corresponding pulse of the standard. The levels of the pulses in the comparison were selected randomly and independently in each trial ( $\mu = 75$  dB SPL,  $\sigma = 8$  dB). The relative weight given to each pulse was obtained by computing the correlation between the level of the first or second pulse in dB and the listener's responses. Listeners adjusted their weighting strategy most appropriately for the given task at the largest IPI. As IPI decreased, listeners gave increasingly similar weight to the two pulses for both target positions. This suggests that level information was integrated across the pulses. Thus, it appears that in certain situations integration of level information occurs over much larger time intervals than previously supposed. The data are not consistent, however, with simple temporal integration of the pulses. [Work supported by NIH DC 00683.]

4:00

**4pPP9. Subjective evaluation of audio compression algorithms: Do conservatory students have a "golden ear?"** Sam Carrier (Dept. of Psych., Oberlin College, Oberlin, OH 44074) and Schuyler Quackenbush (AT&T Labs-Res., 180 Park Ave., Bldg. 103, Florham Park, NJ 07932)

Digital audio compression algorithms are typically based upon psychoacoustic models of hearing; thus they are known as perceptual coders. For commercial applications such as providing music over the Internet, where bit-rate reduction is essential, these lossy codecs have customarily been evaluated through subjective testing using professional audio engineers as listeners. This study replicated the MPEG-2 AAC Stereo Verification Tests (ISO/IEC JTC1/SC29/WG11 N2006) with the 20 listeners drawn from undergraduates enrolled in the Oberlin Conservatory of Music. These tests used the triple-stimulus hidden-reference double-blind methodology with eight codecs and 10 test stimuli. In this procedure, listeners compare two test signals (one coded, the other not) to an uncoded reference, then rate the "audio quality" of both test stimuli on a 1–5 scale in 0.1-step increments. The test stimulus selected as the hidden reference is graded 5.0, the coded signal graded on the degree of impairment. Difference scores thus reflect the discriminability of the stimuli. Taking the magnitude of Studentized  $t$ -statistic values as a measure of acuity, MPEG listeners averaged 4.45, while Oberlin listeners averaged 9.86. While there were some methodological differences between the two experiments, subjects differed several ways Oberlin listeners were younger and were engaged in formal musical training.

4:15

**4pPP10. The relations between auditory problems and cognitive, linguistic, and intellectual measures in first grade children.** Gary R. Kidd, David A. Eddins, and Charles S. Watson (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

The relations between hearing ability and several sensory, perceptual, and cognitive abilities were determined for all children entering first grade in the Benton County, Indiana school system for three consecutive years. This research was part of a multi-disciplinary, epidemiological, longitudi-

nal study of the factors that influence the development of reading and language skills in the early grades. As a result of an auditory evaluation that included otoscopy, tympanometry, pure-tone screening, and speech audiometry, two groups of children with auditory problems were identified who also had lower scores on nonauditory measures. In the total sample of 459 children, 24 had hearing thresholds excess of 20 dB HL and 51 had either excessive ear wax or evidence of ear infection based on otoscopy.

Scores for both groups on measures of reading, cognitive/intellectual abilities, and on the SCAN central auditory test battery were significantly lower than for children with normal auditory function. At entry into second grade, however, reading achievement for both groups was within normal limits. The association between these measures of auditory function and socio-economic status will be discussed. [Work supported by the Benton County School System and by a grant from Indiana University.]

THURSDAY AFTERNOON, 4 NOVEMBER 1999

MORROW ROOM, 1:30 TO 4:30 P.M.

### Session 4pSA

## Structural Acoustics and Vibration: Vibration and Radiation—Simple Structures

Richard P. Szwerc, Cochair

*Naval Undersea Warfare Center, Code 7250, 9500 MacArthur Boulevard, West Bethesda, Maryland 20817*

Hari S. Paul, Cochair

*Acoustical Foundation, 47/20 First Main Road, Gandhi Nagar Adyar, Chennai 600 020 Tamil Nadu, India*

### Contributed Papers

1:30

**4pSA1. Wave propagation experiments using point load excitation in a multi-layered structure.** Chunnan Zhou, Jan D. Achenbach (Ctr. for Quality Eng. and Failure Prevention, Northwestern Univ., Evanston, IL 60208), and John S. Popovics (Drexel Univ., Philadelphia, PA 19104)

Transient wave propagation measurements on a single plate, a two-layer plate and a two-layer-on-a-half-space structure are reported. The experiments are performed using point load excitation and point detection at the surface of the structure. The experimental results are used to verify the accuracy of a new model for wave propagation in layered structures. The basis of the model is first introduced. Experimental measurements using different wave sources, plate materials and wave sensors are compared with each other and with theoretical results. The actual source functions for pencil-lead-break wave sources are obtained. Next, experiments performed on an aluminum plate bonded to a stainless steel plate are reported. Both the cases of stainless steel atop aluminum and aluminum atop stainless steel are considered. The experimental results show good agreement with the theoretical predictions. The effects of the second layer on the transient displacement of the top surface are discussed. Finally, a two-layer-on-a-half-space structure consisting of a stainless steel plate as the top layer, an aluminum plate as the second layer and a thick acrylic resin block as the half-space is studied. The presented experimental transient measurements agree with the predicted results, thus verifying the accuracy of the model.

1:45

**4pSA2. A review and critique of theories for piezoelectric laminates.** V. G. Senthil, Vasundara V. Varadan, and Vijay K. Varadan (Ctr. for the Eng. of Electron. and Acoust. Mater., Penn State Univ., University Park, PA 16802, vvvesm@engr.psu.edu)

This paper presents a three-dimensional (3-D) complete field solution for active laminates based on a modal, Fourier series solution approach that is used to compute all the through thickness electromechanical fields near the dominant resonance frequency of a beam plate with two piezoelectric (sensor and actuator) and one structural layers. Then a detailed review of the extant laminate models used for piezoelectric laminates emphasizing the underlying assumptions in each case is presented. The non-zero, through thickness field components are computed under these assumptions. The results of the 3-D model and FSDT model are compared

for two aspect ratios (AR thickness to width of layers). An AR of 20 is at the limit of FSDT and an AR of 50 well within the assumptions of FSDT. It is concluded that for moderate aspect ratios, several of the approximations of FSDT are questionable at resonance frequencies, resulting in large errors in the estimation of the elastic and electric field components.

2:00

**4pSA3. A novel technique for the prediction of the displacement of a line-driven plate with discontinuities.** Daniel DiPerna and David Feit (Carderock Div., Naval Surface Warfare Ctr., CDNSWC, 9500 MacArthur Blvd., West Bethesda, MD 20817)

A novel technique is presented for obtaining approximate analytic expressions for an inhomogeneous line-driven plate. The equation of motion for the inhomogeneous plate is transformed, and the transform of the total displacement is written as a sum of the solution for a homogeneous line-driven plate plus a term due to the inhomogeneity. This expression may in general be solved numerically. However, by introducing a small parameter into the problem, it may be solved approximately using perturbation techniques. This series may not be convergent, but its convergence may be improved using the Padé approximation. Results are presented for the case of an inhomogeneity concentrated at a single point, and a distribution of points along the line-driven plate. [Work sponsored by the CDSNWC ILIR program and ONR, Code 334.]

2:15

**4pSA4. Structural intensity in a point-excited infinite elastic plate in the high-frequency range.** Sabih I. Hayek and Jungyun Won (Active Vib. Control Lab., Dept. of Eng. Sci. and Mech., 227 Hammond Bldg., Penn State Univ., University Park, PA 16802)

An infinite elastic plate is excited by a mechanical point force, which generates a vector active structural intensity field in the plate. In an earlier paper [Hayek *et al.*, *J. Acoust. Soc. Am.* **105**, 1299 (1999)], the structural intensity was evaluated analytically for a plate using the Bernoulli–Euler theory, which is valid in the low-frequency range. The paper then employed the same theory for the active control of total structural intensity by use of co-located point forces and moments. In this paper, the evaluation of the vector structural intensity field due to point forces and moments is modeled by use of the more exact Mindlin plate theory, which is valid in a much higher frequency range. The main objective is to achieve active



control of the total structural intensity in infinite elastic plates using the Mindlin plate theory through a co-located point force and a point moment actuator at an arbitrary location on the plate.

2:30

**4pSA5. The measurement of structural intensity on a plate with nonuniform thickness.** Richard Szwerz, Henry Chang (Naval Surface Warfare Ctr., Code 725, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700, szwercrp@nswccd.navy.mil), and Courtney Burroughs (Penn State Univ., State College, PA 16804)

A method to measure the structural intensity on a plate with non-uniform thickness is presented and demonstrated with experimental results. The method is an extension of the wave decomposition method, previously demonstrated on uniform thickness beams. It treats a plate as a superposition of beams, and relies only upon local rather than global vibration measurements. The average thickness in the local region of the measurement points is the plate thickness used in the calculations of the power levels.

2:45–3:00 Break

3:00

**4pSA6. Estimating radiation modes using wave-number filtering.** Scott D. Sommerfeldt and Dong Lin (Dept. of Phys., Brigham Young Univ., Provo, UT 84602, s\_sommerfeldt@byu.edu)

Acoustic radiation modes have been investigated in recent years as a means of estimating the far-field radiated acoustic power from a structure. The authors have been investigating two possible methods of estimating the amplitudes of the radiation modes associated with radiation from a baffled beam. The first method, based on using distributed sensors shaped according to the radiation modes, has been reported previously [S. D. Sommerfeldt and D. Lin, *J. Acoust. Soc. Am.* **105**, 1088 (1999)]. This paper will report a second method, based on using distributed sensors designed to act as low-pass wave-number filters. The spatial filtering provided by the sensors allows the array to estimate the radiation mode amplitudes with only a small number of sensors, while still avoiding spatial aliasing problems. Numerical analysis indicates that the method is capable of providing good estimates of the radiated acoustic power for a normalized frequency range extending up to  $kl=10$ , using an array of only six sensors. Although the radiation mode shapes are frequency dependent, the method has demonstrated the ability to provide good modal estimates throughout this entire frequency range. Broadband estimates of the acoustic radiation are typically within 1 dB of the actual value.

3:15

**4pSA7. A computational acoustic field reconstruction process based on an indirect variational boundary element formulation.** Zhidong Zhang, Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109-2145), S. T. Raveendra (Automated Analysis Corp., 2805 S. Industrial, Ann Arbor, MI 48104), T. Allen, and K. Y. Zhang (Ford Motor Co., 20901 Oakwood Blvd., MD299, Dearborn, MI 48121)

The objective of this work is to develop a computational capability based on the indirect variational boundary element method (IVBEM) to evaluate appropriate velocity boundary conditions on an assembly of piston-type sources such that they can recreate a prescribed acoustic field. Information for the acoustic pressure of the original acoustic field at certain field points constitutes the input to the developed process. Several new developments associated with the IVBEM are necessary for completing the field reconstruction process. The developed computational capability targets exterior radiation applications. Thus a new formulation for treating irregular frequencies in the IVBEM is developed and implemented in the field reconstruction process. The computation of the appropriate velocities for the piston-type sources is based on deriving transfer

functions between the field points, where the acoustic pressure of the original field is prescribed, and the velocity assigned on each element of the generic source. A singular value decomposition solver is integrated with the IVBEM computations in order to evaluate the velocity boundary conditions from the transfer functions. Finally, an algorithm that identifies the optimum field points from a set of candidate points for prescribing the acoustic pressure of the original field is developed. [Research supported by a University Research Program award from Ford Motor Co.]

3:30

**4pSA8. Wave-number solution for a string with distributed mass properties.** Mauro Pierucci (Dept. of Aerosp. Eng., College of Eng., San Diego State Univ., San Diego, CA 92182)

A structure with spatially variable properties when excited at a single frequency will excite many wave numbers. In this paper the vibration of a string with a mass distribution  $m(x)$  given by a Fourier series expansion is analyzed. The equation governing the wave-number solution  $w(k)$  for the string displacement is obtained via Fourier transform. The final result is a finite-difference equation. The finite-difference equation relates all of the displacement wave numbers to each other. The solution to this equation is obtained by rewriting it as a system of simultaneous equations that are then solved by standard linear algebra techniques. Wave-number spectral solutions are obtained for different mass distribution. In one case the mass distribution is assumed to vary in a discontinuous manner between adjacent bays, in another case the mass is assumed to vary in a "triangular" manner. The wave-number solutions are presented in 3-D graphs so that the magnitude of the displacement can be seen as a function of the wave number and either the magnitude of the additional mass, the area over which the mass is distributed, or the length of the bay. In the limiting case as the mass distribution becomes very concentrated, the classical pass/stop band, familiar with periodic structures, is recovered.

3:45

**4pSA9. Love wave surface acoustic wave sensor for ice detection on aircraft.** Vasundara V. Varadan, Sunil Gangadharan, and Vijay K. Varadan (Ctr. for the Eng. of Electron. and Acoust. Mater., Penn State Univ., University Park, PA 16802, vvvesm@engr.psu.edu)

This paper presents the design, fabrication, experimental results, and theoretical validation of a Love wave surface acoustic wave sensor for detecting the phase change from liquid water to solid ice. The sensing of this phase transition is due to the shear horizontal nature of Love waves which couple to a solid (ice) but not to a liquid (water). An SiO<sub>2</sub> film of thickness 3.2  $\mu\text{m}$  deposited on an ST cut quartz wafer via plasma-enhanced chemical-vapor deposition acts as the guiding layer for Love waves. Testing is carried out with the water or ice placed directly in the propagation path of Love waves. An oscillation frequency shift of 2 MHz is observed when water on the sensor is frozen and melted cyclically. The contribution to the frequency shift is explained in terms of the acousto-electric effect (high permittivity and conductivity of water relative to ice), mass loading, and elastic film formation (solid ice). An arrangement for wireless interrogation of the sensor is proposed which is particularly attractive for aircraft and rotorcraft applications obviating the need for complex wiring and local power sources.

t

4:00

**4pSA10. Bifurcation and chaos in flexural vibration of a baffled plate in mean flow.** Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202) and Jason Zhu (General Motors Truck Group, Milford, MI 48380)

This paper presents the results of an investigation of flexural vibration of a finite plate clamped to an infinite baffle in mean flow. The plate flexural displacements are solved by the Galerkin method. The critical mean flow speed is determined by a general stability theory and the Routh

algorithm. Results show that the instabilities of a baffled plate are mainly caused by an added stiffness due to acoustic radiation in mean flow, but controlled by structural nonlinearities. The added stiffness is shown to be negative and increase quadratically with the mean flow speed. Without the inclusion of structural nonlinearities, the plate has only one equilibrium, namely, its undeformed flat position. Under this condition, the amplitude of plate flexural vibration grows exponentially in time everywhere, known as absolute instability. With structural nonlinearities, the plate may possess multiple equilibria. When the mean flow speed exceeds the critical value, the plate becomes unstable around all equilibria, jumping from one equilibrium position to another. Since this jumping is random, the plate flexural vibration may seem chaotic. This chaotic behavior disappears when viscous damping is introduced. Accordingly, the plate settles down to a “buckled-down” position owing to the hydraulic fluid-loading effect. [Work supported by ONR.]

**4pSA11. Transverse wave motion of piezoelectric (622) layer due to step potential.** C. Anandam, R. Natrajan (Dept. of Mathematics, A. C. College of Technol., Anna Univ., Chennai-600 025, India), and H. S. Paul (Acoustical Foundation, Gandhi Nagar, Adyar, Chennai-600 020, India)

The propagation of transverse waves in a piezoelectric layer of class (622) is investigated, when excited by either (a) step electric potential, keeping zero stress or (b) step shear stress, keeping zero potential. The resulting electric potential or shear stress is presented graphically. For case (a), electric potential linearly increases as the thickness of the plate increases, whereas stress is maximum near nondimensional thickness  $r = 0.1$  and decreases gradually and becomes zero at the top surface. For case (b), electric potential becomes negative and linearly decreases up to nondimensional thickness,  $r = 0.87$  and then increases and becomes zero at the top surface with slight oscillatory nature as the thickness of the plate increases. It is also observed that mere application of time dependent electric potential in piezoelectric (622) class generates transverse waves, which is not possible in any other piezoelectric material.

THURSDAY AFTERNOON, 4 NOVEMBER 1999

DELAWARE A & B ROOM, 1:30 TO 4:30 P.M.

### Session 4pSC

#### Speech Communication: Speech Perception: Audio and/or Visual Cues (Poster Session)

John W. Hawks, Chair

*School of Speech Pathology and Audiology, Kent State University, Kent, Ohio 44242*

#### Contributed Papers

To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m. To allow for extended viewing time, posters will be on display from 9:00 a.m. to 10:00 p.m.

**4pSC1. Perceptual effects of place of articulation on voicing for audiovisually discrepant stimuli.** Lawrence Brancazio, Joanne L. Miller, and Matthew A. Paré (Dept. of Psych., Northeastern Univ., 125 NI, 360 Huntington Ave., Boston, MA 02115, brancazio@neu.edu)

This research investigated effects of the visually specified place of articulation on perceived voicing. It is known that the /b-/p/ boundary along a voice-onset-time (VOT) continuum falls at a shorter VOT than the /d-/t/ boundary. Green and Kuhl [Percept. Psychophys. **45**, 34–42 (1989)] demonstrated that tokens from an auditory /ibi-/ipi/ continuum dubbed with an /igi/ video and perceived as /idi/ and /iti/ due to the “McGurk effect” had a voicing boundary at a longer VOT than when presented only auditorily and perceived as /ibi/ and /ipi/. We extended this finding in two directions. First, using an auditory /bi-/pi/ series with a video /ti/ for which the McGurk effect did not always occur, we compared visually influenced (/d,t/) and visually uninfluenced (/b,p/) responses for these audiovisually discrepant stimuli. The /d-/t/ boundary was at a longer VOT than the /b-/p/ boundary, affirming the boundary shift’s perceptual, not stimulus-based, origin. Second, we tested the generalizability of Green and Kuhl’s findings using auditory /di-/ti/ tokens with /ti/ and /pi/ videos. Preliminary results suggest that boundaries for /b-/p/ and /bd-/pt/ percepts (with video /pi/) may occur at shorter VOTs than for /d-/t/ percepts (with video /ti/). Thus, the visually induced voicing boundary shift apparently replicates across different stimulus configurations. [Work supported by NIH/NIDCD.]

**4pSC2. Talker variability effects in auditory-visual speech perception.** Arlene E. Carney, Bart R. Clement (Dept. of Commun. Disord., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, carne005@umn.edu), and Kathleen M. Cienkowski (Univ. of Connecticut, Storrs, CT 06268)

Talker variability effects have been demonstrated for many aspects of syllable, word, and sentence recognition for stimuli presented in the auditory-only modality. In the current investigation, talker variability effects were examined in multiple modalities—visual-only, auditory-only, and auditory-visual. Auditory-visual presentations were of two types: consonant [auditory and visual were of the same token—/visual bi and auditory bi/] or disparate [/visual gi and auditory bi or visual bi and auditory gi/]. The disparate condition has elicited “the McGurk effect,” in which listeners may report a fused response that is neither the auditory or the visual component of the stimulus. Eleven talkers (five male and six female) of varying cultural backgrounds were videotaped. Their productions of CV syllables were presented to adult listeners who reported their percepts. Individual talkers elicited different degrees of syllable fusion among listeners, despite their equivalent auditory intelligibility. In addition, individual listeners varied considerably in their overall perceptions, from complete fusion for all talkers to almost no fusion for any talker. Results suggest that listeners may employ different types of processing strategies for bimodal stimuli. [Research supported by NIDCD.]

**4pSC3. Audibility and visual biasing in speech perception.** Bart R. Clement and Arlene E. Carney (Dept. of Commun. Disord., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, clem0025@umn.edu)

Evidence of an integrative interaction between vision and audition comes from studies of audiovisual disparity, in which conflicting auditory and visual speech tokens may elicit a fused perceptual response that is different from that represented in either modality alone (i.e., the “McGurk Effect”). Visual information in speech can bias and in some cases even dominate the auditory percept. Hearing-impaired listeners may be more susceptible to visual biasing in speech perception than normal-hearing listeners because of an increased reliance on visual information. It is argued here that the degree of audibility of the auditory component of the stimulus must be equated across the two groups before true differences in visual biasing can be established. In the current investigation, the effects of visual biasing on speech perception are examined as audibility is systematically degraded for 12 normal-hearing listeners and two male talkers. Results demonstrate that the degree of visual biasing is in fact related to degree of audibility. Further, individual normal-hearing listeners may differ distinctly with regard to how they integrate the two modes of information. Both of these factors must be considered when this phenomenon is studied in hearing-impaired listeners. [Research supported by NIDCD.]

**4pSC4. Testing the relative salience of audio and visual cues for stop place of articulation.** Stephen J. Winters (Dept. of Linguist., Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, swinters@ling.ohio-state.edu)

Recent OT analyses have appealed to putatively universal rankings of perceptual entities in order to motivate analyses of phonological phenomena. J. Jun [WCCFL 14, 221–237 (1995)] bases a universal constraint ranking for susceptibility of stops to place assimilation on the inherent “salience” of place cues. E. Hume [WCCFL 17 (1998)] appeals to the “perceptual vulnerability” of labial stops in order to account for their cross-linguistic tendency to undergo metathesis. Both phonologists motivate their analyses with speculative and unquantified assumptions about place cue salience. This study used an objective definition of salience to empirically establish the relative strengths of audio and visual cues for stop place of articulation. Comparing results between audio-visual and audio-only groups showed that visual cues are strongest for labials but are also significantly higher for coronals than dorsals. Adding acoustic information between speech reception threshold and comfortable listening level conditions had weaker effects than adding visual information, but contributed most strongly to the intelligibility of dorsals. Overall results suggested that labial stops have the strongest cues for place regardless of modality, contrary to what Hume, Jun and other researchers would predict. [This material is based upon work supported by a National Science Foundation Graduate Fellowship.]

**4pSC5. Auditory-visual integration in younger and older adults.** Mitchell S. Sommers (Dept. of Psych., Washington Univ., Campus Box 1125, St. Louis, MO 63130) and Nancy Tye-Murray (Central Inst. for the Deaf, St. Louis, MO 63130)

This study examined age-related changes in the ability to integrate auditory and visual speech information. Visual enhancement, the advantage afforded by seeing as well as hearing a talker, compared with listening alone, was measured in younger (18–25) and older (over age 65) adults. In addition, dichotic identification of filtered speech (auditory–auditory speech integration) and binaural gap detection (auditory–auditory nonspeech integration) were tested to evaluate whether auditory-visual integration was related to a more global ability to integrate sensory signals, be they speech or nonspeech. Visual enhancement was significantly impaired in older, compared with younger adults. Importantly, visual-only (lipreading) performance of younger and older adults was nearly identical,

suggesting that age-related declines in visual enhancement were not a result of reduced reception of visual information. Older adults also exhibited deficits in the ability to integrate two channels of auditory information and these deficits were similar for both speech (sentences) and nonspeech (gap detection) stimuli. Regression analyses indicated significant correlations between the three measures of sensory integration, visual enhancement, dichotic sentence identification, and binaural gap detection. Taken together, the findings suggest that age-related declines in visual enhancement are a result of a generalized decline in sensory integration abilities.

**4pSC6. Geometrical displays of speech spectra as aids to lipreading.** Steven L. Tait, Jr. and William J. Strong (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, strongw@acoust.byu.edu)

Many syllables are ambiguous to lip readers because of their similar appearance on the lips. An aid is necessary to make distinction of these syllables possible. In a previous study [E. J. Hunter and W. J. Strong, *J. Acoust. Soc. Am.* 102, 3166–3167 (1997)], speech spectra were processed at 5-ms intervals and displayed as sequences of irregular decagons. Subjects were asked to discriminate between pairs of decagons representing ambiguous syllables. Subjects successfully discriminated most pairs, but were inconsistent in discrimination of voiced-/unvoiced-stop pairs. This study is concerned with display formats aimed at improving discrimination of ambiguous pairs to develop a more effective aid for lip readers. In particular, “pie slice” and ray formats are evaluated against polygons. Filled shapes are evaluated against open shapes. Five-band speech spectra are evaluated against ten-band spectra. Display color is evaluated against line thickness for representing overall sound level.

**4pSC7. A real-time audio–video front-end for multimedia applications.** Dmitry Zotkin, Ramani Duraiswami, Ismail Hariatoglu, Larry Davis (Inst. for Adv. Computer Studies, Univ. of Maryland, College Park, MD 20742, ramani@umiacs.umd.edu), and Takahiro Otsuka (ATR Media Integration & Commun. Res. Labs., Kyoto, Japan)

A real-time system combining auditory input received by a microphone array with video input is designed and implemented as a prototype smart video-conferencing system. The audio part of the system uses accurate algorithms for localization of the source in typical noisy office environments, and performs dereverberation and beamforming to enhance signal quality. The coordinates of the sound sources obtained are used to control the pan, tilt and zoom of a video camera. The video-processing part segments people in the image using an obtained background model, and can label and track them and get a close-up view of the head of the talker. Heuristic algorithms for intelligent zoom and tracking of people in video-conferencing scenarios are incorporated into the system. The current implementation is performed on a commercial off-the-shelf personal computer equipped with a data acquisition board and does not require expensive hardware. The speed, precision and robustness of the system are superior to existing ones. Demonstrations of the system in different scenarios will be presented, along with quantitative measures of its performance. The system is designed as a general purpose front-end to systems for speech and person recognition, and for newer human-computer interfaces. [Work supported in part by DARPA.]

**4pSC8. Examining processor and stimulus list differences in word recognition by children using a cochlear implant.** Stefan Frisch (Prog. in Linguist., Univ. of Michigan, Ann Arbor, MI 48109-1285), Ted A. Meyer, and David B. Pisoni (Indiana Univ. Med. Ctr.)

Boothroyd and Nittrouer [*J. Acoust. Soc. Am.* 84, 101–114 (1988), *J. Acoust. Soc. Am.* 87, 2705–2715 (1990)] and Rabinowitz *et al.* [*J. Acoust. Soc. Am.* 92, 1869–1881 (1992)] have demonstrated a power-law



relationship between the accuracy in phoneme recognition and the accuracy in word recognition by normal hearing adults, older adults, children, and adults using cochlear implants. The power-law relationship provides a simple mathematical model of context effects on speech perception. An analysis of spoken word recognition performance by children using cochlear implants reveals an analogous power-law relationship between phoneme and word recognition. Further, comparison between groups using two different speech processors (SPEAK and MPEAK) shows that, while overall performance is affected by the processor, the power-law relationship is unaffected. By contrast, comparison between different stimulus lists finds the power-law relationship is affected by the phonotactic structure of the 20 lexical items used and their familiarity to the listener. [Work supported by NIH, AAO-HNS.]

#### **4pSC9. Vocal tract size and the intelligibility of competing voices.**

Peter F. Assmann (School of Human Development, Univ. of Texas at Dallas, Richardson, TX 75083)

When two people speak at the same time, it is easier to understand what either is saying if the voices differ in fundamental frequency ( $F_0$ ). Talkers differ in vocal tract size as well as  $F_0$ , and their formant frequencies occupy different ranges, providing a further basis for voice segregation. To investigate the effects of vocal tract size, a set of declarative English sentences was produced by an adult male and processed using a speech vocoder. Sentences were presented in pairs, with the spectrum envelope shifted up or down by a fixed percentage in one of the sentences. A +20% shift (sufficient to shift the formants of an adult male into the female range) led to a 6% increase in word recognition accuracy. A -20% shift led to a 9% drop in accuracy, consistent with listeners' informal reports that the "larger" voice sounded muffled. Benefits of upwards shifts and adverse effects of downward shifts were restricted to the "shifted" member of the pair. Upward spectral shifts were accompanied by increased spectral tilt, while downward spectral shifts led to reduced spectral tilt. Hence the observed effects may be due to spectral masking, rather than sensitivity to vocal tract size per se.

#### **4pSC10. The role of modulation spectrum amplitude and phase in consonant intelligibility.**

Steven J. Aiken, Donald G. Jamieson, Vijay Parsa, and Prudence Allen (Hearing Health Care Res. Unit, UWO, School of Commun. Sci. and Disord., Elborn College, London, ON N6G 1H1, Canada)

This paper considers an acoustic basis for speech intelligibility and evaluates various acoustically based speech intelligibility prediction algorithms. Earlier research indicates that speech intelligibility does not require preservation of spectral and temporal fine-structure, but is highly dependent on the preservation of the amplitude component of the modulation spectrum [R. Drullman, *J. Acoust. Soc. Am.* **97**, 585-592 (1995)]. This study assessed the importance of the phase component of the modulation spectrum using a 21-alternative forced-choice consonant perception test. Temporal and spectral fine-structure were removed by modulating a white noise carrier with 50 Hz low-pass filtered speech amplitude envelopes in 4, 8, or 24 discrete bands. Modulation spectrum phase was distorted by imposing a random delay in each discrete band. Behavioral results are discussed in light of intelligibility predictions generated by the articulation index [N. R. French and J. C. Steinberg, *J. Acoust. Soc. Am.* **19**, 90-119 (1947)] and the speech transmission index [H. J. M. Steeneken and T. Houtgast, *J. Acoust. Soc. Am.* **67**, 318-326 (1980)].

#### **4pSC11. Assessment of intelligibility of digitized speech by young adults under different listening conditions.**

Ramesh N. Bettagere (Dept. of Special Education and Commun. Sci. and Disord., Campbell Hall, No. 217, SLU879, Southeastern Louisiana Univ., Hammond, LA 70402)

The present study evaluated the intelligibility of tape-recorded speech sentences and digitized speech sentences digitized at low (4 kHz), moderate (8 kHz), and high (16 kHz) sampling rates by young adults under

different listening conditions (quiet versus noise). Twenty-four young adults participated as subjects. The tape-recorded speech sentences and digitized speech sentences were presented to each subject in quiet and in the presence of background speech babble at a signal-to-noise ratio of 0 dB. The subjects were instructed to immediately write down each sentence soon after they heard it. A  $4 \times 2$  analysis of variance with repeated measures was performed to assess the effects of mode of speech and listening conditions on total number of words correctly transcribed. Results showed that the main effects for mode of speech and listening condition were statistically significant. The interaction effects of mode of speech by listening condition were also statistically significant. Pairwise comparisons showed that significantly more words were transcribed in the quiet condition than in the corresponding noise condition for all the modes of speech except for the digitized speech at high sampling rate (16 kHz). The implications of the results are discussed.

#### **4pSC12. Vowel intelligibility in clear and conversational speech.**

Sarah Hargus Ferguson and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

Several studies have demonstrated that instructions to speak clearly yield significant improvements in speech intelligibility along with a wide variety of acoustic changes relative to conversational speech. The current study explored the relationship between acoustic properties of vowels and their identification in clear and conversational speech, as well as the effects of hearing loss on this relationship. The goals were, for both normal and impaired listeners: (1) to determine what acoustic factors underlie the intelligibility differences observed between speaking styles and (2) to explore the relative contribution of various acoustic cues to vowel identification, using stimuli that vary naturally in intelligibility. Monosyllabic words were excised from sentences spoken either clearly or conversationally and presented in a background of 12-talker babble for vowel identification by young normal-hearing and elderly hearing-impaired listeners. Vowel identification performance was correlated with the results of acoustic analyses to assess the relative importance of spectral, durational, and dynamic cues to vowel perception. Preliminary analyses suggest that spectral target and dynamic formant information are primary cues to vowel identity, while the importance of duration is significant but small. [Work supported by NIHDCD-02229 and the ASHA Foundation.]

#### **4pSC13. Effects of talker differences in vowel production on vowel identifiability.**

Amy Neel (Dept. of Audiol. and Speech Sci., Purdue Univ., West Lafayette, IN 47907-1353, atneel@purdue.edu)

Normal talkers differ in speech intelligibility, and vowel production characteristics contribute to intelligibility. This study investigated the relation between talker's acoustic vowel characteristics and the identification of their vowels by normal-hearing listeners. Static and dynamic characteristics were measured for ten American English vowels produced in /dVd/ context by two female and two male talkers. Listeners identified four types of resynthesized vowels for each talker, ranging from nearly natural tokens with dynamic formants and appropriate duration values to relatively impoverished tokens containing only static formants and no duration cues. Significant differences in identification scores among talkers were found. The talker with the best-identified vowels, a male, had the largest and most dispersed vowel space with little overlap of formant trajectories in the dynamic  $F1 \times F2$  space. The worst talker, also male, had a smaller vowel space with greater overlap of formant trajectories. For all four talkers, dynamic-formant vowels were better identified than static-formant vowels, and there were greater differences among talkers for dynamic-formant



vowels than for static-formant vowels. The results confirm that formant movement contributes to vowel identification and that greater separation in the dynamic  $F1 \times F2$  space contributes to better vowel identifiability.

**4pSC14. Dialectal differences in diphthong perception.** Matthew J. Makashay (Dept. of Linguist., Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, makashay@ling.ohio-state.edu)

This study attempts to determine if speech perception varies across dialects as production does. Synthetic stimuli based on one talker from Binghamton, NY (northern US) and one from Birmingham, AL (southern US) were presented to subjects from both regions. The stimuli were 18 vowel continua in CVC context. These 10-point continua had initial tokens containing nonhigh vowels, such as *hot* /hat/, with formant values and durations manipulated to result in final tokens containing diphthongs, such as *hot* /hit/. Other continua contained endpoints such as *sad* /sæd/ and *side* /said/, or *bought* /bt/ and *bout* /baut/. The production of diphthongs differs between these dialects, varying from the Canadian raising of the North to the monophthongization of the South. The question to be resolved in this study is whether perception of diphthongs differs between the dialects. It was found in a direct boundary estimation task that southern subjects perceived southern tokens as diphthongs earlier in the continua than northern subjects did, while there was no significant difference between these groups for the northern diphthongs. [This material is based upon work supported under a National Science Foundation Graduate Fellowship.]

**4pSC15. The perception of noise vocoded prevocalic stop bursts and formant transitions.** Michael Kieffe (Univ. of Alberta, Edmonton, AB, Canada)

Theories of prevocalic stop consonant perception differ with respect to the amount of spectral resolution required by the model, e.g., gross spectral shape theories predict that human speech perception is more robust against decreased frequency selectivity where detailed cues such as vocalic formant transitions break down. Previous results in the perception of noise vocoded speech [M. Kieffe, Int. Conf. Phon. Sci. (1999)] showed that, although listeners' performance worsened significantly at bandwidths as low as 500 Hz, a spectral shape model using cepstral coefficients was better able to predict subjects' responses. However, a simulation which showed that release bursts require much less spectral resolution than vocalic formant transitions complicates the interpretation. Therefore, in this experiment, subjects were asked to identify the place of articulation for gated release bursts independently of vocalic portions for both unprocessed and noise channel vocoded speech. Preliminary results show that correct identification of stop place from only the release burst does not change significantly up to a channel bandwidth of 2000 Hz. Results suggest that responses from the previous experiment can be explained by the robustness of the burst in those stimuli. Gross versus detailed cue models are also evaluated with respect to responses to formant only stimuli. [Work supported by SSHRC.]

**4pSC16. Is the perceptual magnet effect a matter of auditory sensitivity?** Sreedivya Radhakrishnan, John W. Hawks, and Robert J. Otto (School of Speech Pathol. and Audiol., Kent State Univ., Kent, OH 44242)

Efforts to validate and extend Kuhl's [P. K. Kuhl, Percept. Psychophys. 50, 93–107 (1991)] perceptual magnet effect for vowel prototypes in adult listeners has been somewhat elusive and particularly difficult to replicate in other than the /i/ vowel space. However, there is converging evidence that suggests the magnet effect, attributed to assimilation by a prototype, may more likely be due to differences in auditory sensitivity across the vowel space. Hawks' [J. W. Hawks, J. Acoust. Soc. Am. 95,

1074–1084 (1994)] data on discrimination of formant changes for 17 synthetic vowels supported this notion, indicating that difference limens (DLs) based from an exemplar token for /i/ was among the largest (poorest) found, while tokens neighboring /i/ yielded the smallest DLs. If the perceptual magnet effect is a demonstration of differences in auditory sensitivity and Hawks' data adequately reflects auditory sensitivity in the vowel space, then a likely location to find a magnet effect is within the space for /u/, which yielded the poorest DLs in Hawks' study. The results of an experiment investigating this possibility through replication of Kuhl's original protocol in the vowel space for /u/ will be presented.

**4pSC17. Cues to vocal-fold and vocal-tract lengths interact in talker-sex perception from vowel segments.** Michael J. Owren (Dept. of Psych., 224 Uris Hall, Cornell Univ., Ithaca, NY 14853) and Jo-Anne Bachorowski (Vanderbilt Univ., Nashville, TN 37240)

Even brief sounds can be richly informative about talker characteristics. Earlier acoustic analysis confirmed predictions that the best statistical classification of vowels by talker sex occurred using acoustic correlates of dimorphism in vocal-fold and vocal-tract lengths [J.-A. Bachorowski and M. J. Owren, J. Acoust. Soc. Am. (in press)]. In follow-up perceptual testing, 24 participants in experiment 1 classified talker sex from 180 short segments that balanced fundamental frequency ( $F_0$ ) and estimated vocal-tract length (VTL) cues. Responses were 98% correct, with a mean latency of 478 ms. However, latencies were longer for male talkers that combined high  $F_0$  and short VTL values, as well as for female talkers with low  $F_0$  and long VTL. In experiment 2, 24 participants heard stimuli composed of small numbers of waveform cycles. Classification "threshold" was 1.8 cycles, with male talkers being more accurately classified than females. These results show that indexical cueing is an inherent component of even the shortest possible vowel segments, and that the acoustic features most closely related to sexual dimorphism interact in predictable fashion in influencing perceptual processing.  $F_0$  and VTL, the particular acoustic cues in question, are fundamentally related to features also known to influence speech processing.

**4pSC18. The perceptual consequences of overlap in /s/ and /ʃ/ productions within a talker.** Sheryl A. Clouse, Jessica L. Burnham, and Rochelle S. Newman (Dept. of Psych., Univ. of Iowa, E11 Seashore Hall, Iowa City, IA 52242)

A primary issue in speech perception is the apparent lack of invariance between the acoustic information in a signal and the listeners perception. Different intended phonemes may be produced with identical acoustic values. Previously, we examined fricative centroids and frication peaks for over 100 utterances beginning with /s/ and /ʃ/ from each of 20 different speakers, and found substantial overlap across talkers. In the present study, we examine the effect of this overlap on perception. Listeners in a phoneme identification task heard natural productions of /s/ and /ʃ/ syllables from speakers with either little or great overlap between categories. Although labeling accuracy was near ceiling, effects were found in listeners' reaction times. Listeners needed more time to interpret the speech of talkers who had substantial overlap in their fricative centroids, even when these talkers showed no overlap in their frication peaks. Listeners also had slower reaction times for talkers with overlap in their frication peaks, but not their centroids. This suggests that category overlap has measurable consequences to perception, and that peaks and centroids are sensitive to different aspects of the speech signal, both of which are perceptually important to listeners. [Work supported by NIDCD Grant R01-DC00219 to SUNY at Buffalo.]

**4pSC19. An fMRI investigation of feature distinctions in consonant identification.** Tobey L. Doeleman (Cornell Phonet. Lab., Cornell Univ., Ithaca, NY) and Joy Hirsch (MSKCC/Cornell Univ., College of Medicine, New York, NY)

Differences in neural activation associated with the voicing and place distinctions of phonemes were investigated using functional magnetic resonance imaging (fMRI). Two types of synthetic stimuli were presented auditorily in a forced-choice identification task: clear CV syllables and CV syllables whose initial consonants were ambiguous with regard to either place or voicing. The stimuli consisted of the consonants [p, b, t, d, k, g] followed by the vowels [a, i, u]. Subjects were asked to match each auditory stimulus with one of two visually presented choices representing either voicing-feature or place-feature alternatives. Results consistently showed more widespread activation for all subjects in Wernicke's area (Brodmann's area 22) associated with determining the voicing feature versus the place feature of a phoneme. This finding is consistent with previous behavioral results [T. L. Doeleman, *J. Acoust. Soc. Am.* **101**, 3111(A) (1997)] showing more voicing than place errors in consonant identification and suggests a difference in the relative difficulty of these featural decisions. The fMRI design also allowed for a comparison of activation associated with identifying clear versus ambiguous phonemes. Results show a trend of greater activation, especially in the right hemisphere, associated with the ambiguous stimuli.

**4pSC20. Contrast effects with sinewave analog anchors.** Robert A. Fox and Sarah England (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210-1002, fox.2@osu.edu)

An area of interest for several decades has been the effect of the immediate phonetic context on vowel perception—especially in terms of the phonetic contrast effect. One of the questions regarding this easily replicated effect is whether it is a phonetic-level effect or an acoustic-level effect. The current experiment was designed to determine whether the same contrast effect on the identification of a /i/-/I/ synthetic vowel continuum would be obtained with synthetic vowel (SV) anchors (representing one of the two endpoints) and with sinewave analog (SA) vowels (substituting tone glides for vowel formants). The stimuli included a nine-point /i/-/I/ SV continuum (produced with the Klatt program) and a correlated SA continuum. There were three basic anchoring conditions: anchoring (1) with the SV endpoints, (2) with the SA endpoints (subjects identified these tokens as tones), and (3) with the SA endpoints after subjects had been trained to identify them as vowels. Results showed a significant anchor effect with SV anchors (as expected) and no anchor effects with the SA anchor prior to training. Subjects did show a significant anchoring effect with the SA anchors after training, but the effect was significantly smaller than with SV anchors.

**4pSC21. The effects of high-intensity speech on consonant feature transmission in normal-hearing subjects.** Benjamin W. Y. Hornsby, D. Wesley Grantham, Ralph N. Ohde, Daniel H. Ashmead (Dept. of Hearing and Speech Sci., Vanderbilt Univ., 1114 19th Ave. South, Nashville, TN 37212, ben.hornsby@vanderbilt.edu), and Timothy D. Trine (Starkey Labs., Inc., Eden Prairie, MN 55300)

The effect of high-speed presentation levels on consonant recognition and feature transmission was assessed in normal-hearing subjects. Consonant recognition of C/i/ nonsense syllables was measured at five overall speech levels ranging from 65 to 100 dB SPL. Audibility remained constant by mixing speech stimuli with a speech-shaped noise at a constant 0-dB signal-to-noise ratio. Consistent with the work of others, overall percent correct performance decreased as the presentation level of speech

increased [Studebaker *et al.*, *J. Acoust. Soc. Am.* **105**, 2431–2444 (1999)]. Confusion matrices were analyzed in terms of relative percent information transmitted at each speech presentation level, as a function of feature. Six feature sets (voicing, place, nasality, duration, friction and sonorance) were analyzed. Results showed the feature duration (long consonant duration fricatives) to be most affected by increases in level while the voicing feature was relatively unaffected by increases in level. In addition, alveolar consonants were substantially affected by level while palatal consonants were not.

**4pSC22. Detection of consonant voicing: A module for a hierarchical speech recognition system.** Jeung-Yoon Choi (Res. Lab. of Electron., MIT, Cambridge, MA 02139)

This research describes a module for detecting consonant voicing in a hierarchical speech recognition system. In this system, acoustic cues are used to infer values of features that describe phonetic segments. A first step in the process is examining consonant production and conditions for phonation, to find acoustic properties that may be used to infer consonant voicing. These are examined in different environments to determine a set of reliable acoustic cues. These acoustic cues include fundamental frequency, difference in amplitudes of the first two harmonics, cutoff first formant frequency, and residual amplitude of the first harmonic, around consonant landmarks. Classification experiments are conducted on hand and automatic measurements of these acoustic cues for isolated and continuous speech utterances. Voicing decisions are obtained for each consonant landmark, and are compared with lexical and perceived voicing for the consonant. Performance is found to improve when measurements at the closure and release are combined. Training on isolated utterances gives classification results for continuous speech that is comparable to training on continuous speech. The results in this study suggest that acoustic cues selected by considering the representation and production of speech may provide reliable criteria for determining consonant voicing.

**4pSC23. Learning to discriminate reduced-channel /ba/-/wa/ syllables in normal-hearing adults.** Cammy L. Bahner, Thomas D. Carrell, and T. Newell Decker (Commun. Disord., Univ. of Nebraska, Lincoln, NE 68583-0738)

Acoustic information delivered to individuals with cochlear implants differs in many ways from the same information delivered via a normally functioning cochlea. One difference is that the number of frequencies that can be distinguished is sharply reduced in implanted listeners. This results in poor frequency selectivity where only sounds that fall into different frequency bands are distinguishable (all other things being equal). Shannon [*Science* **270**, 303–304 (1995)] used an electrical filtering and rectification system to simulate this aspect of cochlear implant function in normal-hearing listeners. In the present study software filtering was combined with envelope-shaped-noise (ESN) processing to accomplish the same end. The immediate goal was to examine the efficacy of a simple technique in training listeners to distinguish between reduced-channel versions of syllables. The stimuli tested were synthetically produced /ba/ and /wa/ syllables, identical in every aspect except for their onset duration. These syllables were filtered into eight frequency bands and ESN processed. Following this they were once-again filtered at the same frequencies to create reduced-channel versions of the original syllables. The resulting stimuli were presented to 16 normal-hearing adult subjects in behavioral and electrophysiological tasks. Results demonstrated that the training rapidly improved listeners' abilities to identify reduced-channel /ba/ and /wa/.

**4pSC24. Effects of first formant onset properties on voicing judgments of prevocalic stops without F1 cutback.** José R. Benki (Prog. in Linguist., Univ. of Michigan, 1076 Frieze Bldg., Ann Arbor, MI 48109-1285, benki@umich.edu)

This study examines the effects of first formant (*F1*) onset properties on voicing judgments of prevocalic stop consonants in contexts that do not exhibit the *F1* cutback covariation present in pretonic and utterance-initial voicing contrasts. Previous research [K. R. Kluender and A. J. Lotto, *J. Acoust. Soc. Am.* **95**, 1044–1052 (1994)] strongly suggests that *F1* onset frequency effects on voicing judgments are due to auditory factors, and not to listener experience with natural covariance with *F1* cutback and *F1* onset frequency. The present study extends the previous findings to voicing contrasts in contexts that are not signaled by *F1* cutback, such as before unstressed vowels in English. Categorization data were collected for a continuum between *rabid* (/b/) and *rapid* (/p/) in which both closure duration and *F1* onset properties were manipulated using LPC resynthesis of natural speech. Preliminary results indicate that lower *F1* onset frequencies of the unstressed vowel condition more voiced percepts of the preceding stop consonant, consistent with findings for utterance-initial stop consonants.

**4pSC25. Effects of stimulus uncertainty, consonantal context and training on formant frequency discrimination.** Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405, kewley@indiana.edu)

Words in natural speech vary in their predictability, e.g., stimulus uncertainty ranges from low to very high across sentences. The ability to discriminate formants is strongly affected by longer phonetic context [Kewley-Port and Zheng (in press)]. This study investigates the effects of stimulus uncertainty from minimal to high uncertainty and the phonetic contexts /V/ or /bVd/. The primary formants were from four female vowels /i, ε, æ, ə/. Using adaptive tracking, DLs for discriminating a small change in a formant was calculated in Delta Barks. In experiment 1, performance for five listeners, optimized by extensive training, began with minimal uncertainty, subsequently increasing uncertainty from 8- to 16- to 22 formants per block. Effects of higher uncertainty were less than expected, only decreasing performance by about 33%, although DLs for CVCs were 25% poorer than for isolated vowels. In experiment 2, performance in the 22-formant condition was tracked over 1 h for 37 listeners without formal laboratory training. DLs for untrained listeners were about 300% worse than for trained listeners, and comparable to untrained listeners in Kewley-Port and Zheng (in press). Results indicate that the effects of longer phonetic context degrade formant frequency discrimination more than higher stimulus uncertainty. [Work supported by NIHDCD-02229.]

**4pSC26. Perception of stress and speaking style for selected elements of the SUSAS database.** Robert S. Bolia and Raymond E. Slyph (Air Force Res. Lab., 2255 H St., Wright-Patterson AFB, OH 45433)

The SUSAS database [J. H. L. Hansen and S. E. Bou-Ghazale, *EUROSPEECH 97*, 1743–1746 (1997)] is a collection of utterances recorded under conditions of simulated or actual stress, the purpose of which is to allow researchers to study the effects of stress and speaking style on the speech waveform. The aim of the present investigation was to assess the perceptual validity of the simulated portion of the database by determining the extent to which listeners classify its utterances according to their assigned labels. Seven listeners performed an eight-alternative, forced-choice, judging whether monosyllabic or disyllabic words spoken by talkers from three different accent classes (Boston, Generic Midwest, New York) were best classified as “angry,” “clear,” “fast,” “loud,” “neutral,” “question,” “slow,” or “soft.” Mean percentages of “correct” judgments were analyzed using a 3 (accent class) × 2 (number of syllables) × 8 (speaking style) repeated measures analysis of variance. Results indicated that, overall, listeners correctly classify the utterances only 58% of the time, and that percentage of correct classifications varies as a function of all three independent variables.

**4pSC27. Prosodic disambiguation of syntactic ambiguity in discourse context.** Shari R. Speer, Shari B. Sokol (Univ. of Kansas, 3031 Dole, Sp.-Lang.-Hear., Lawrence, KS 60045), Amy J. Schafer (UCLA, Los Angeles, CA), and Paul Warren (Wellington Univ. of New Zealand)

A cooperative boardgame task was used to examine how native speakers use prosodic structure to resolve syntactic ambiguity in discourse context. The game task required two speakers to use utterances from a predetermined set to negotiate the movement of gamepieces to goal locations. In one condition, the discourse contained two situations that had to be described using the same syntactically ambiguous word sequence. In the other condition, an identical syntactically ambiguous structure was used to describe only one situation. Sentences that could describe two situations had an ambiguous prepositional phrase attachment as in “I want to move the square with the triangle,” in which the move involved either a combined square-and-triangle piece or a triangle pushing a square to another position. Sentences describing only one situation involved only a cylinder pushing the square as in “I want to move the square with the cylinder.” Phonological analyses and phonetic analyses of duration and fundamental frequency were compared. Results indicate that speakers used prosodic phrasing to reflect situational syntactic ambiguity. Examples of relatively high and low variability in production will be discussed. [Work supported by NIH grant MH-51768, NZ/USA Cooperative Science Programme grant CSP95/01, Marsden Fund grant VUW604, and NIH research DC-00029.]

**4pSC28. What is a neighbor in a neighborhood effect? Items that mismatch on the first phoneme still produce neighborhood effects.** Rochelle S. Newman (Dept. of Psych., Univ. of Iowa, E11 Seashore Hall, Iowa City, IA 52242, rochelle-newman@uiowa.edu), James R. Sawusch, Paul A. Luce (State Univ. of New York at Buffalo, Buffalo, NY 14260), and Alicia Healy (Univ. of Iowa, Iowa City, IA 52242)

Previously, we presented results suggesting that neighborhood density could influence perception in phoneme identification tasks. In a typical task demonstrating this effect, subjects might hear two series, one ranging from *beysh* to *peysh*, while the other varied from *beyth* to *peyth*, where *beyth* is similar to more real words than *peyth*, and *peysh* is similar to more real words than *beysh*. Subjects were more likely to classify the ambiguous stimuli from each series as members of the category which makes it more wordlike. In the present series, similar series were created where the only neighbors for the endpoints were ones that differed in their initial consonant. We created two pairs of series: in the first, *foif*–*toif* and *fof*–*tof*, the neighbors which drove the bias were all ones which matched on the initial phoneme (e.g. *choke*). In the second series, *zUf*–*zUf* and *zE*–*zE*, the only neighbors for the items were ones that mismatched on the initial phoneme (e.g., *push*, *mesh*, *fetch*, etc.). Similar neighborhood effects were found for both series, suggesting that words in memory which do not match the beginning of a perceived item are still activated and can still influence perception.

**4pSC29. Inhibition in phonological priming: Lexical or strategic effects?** Lisa C. Shoaf and Mark A. Pitt (Dept. of Psych., The Ohio State Univ., 1885 Neil Ave., Columbus, OH 43210, contos.1@osu.edu)

Facilitatory (speeded) and inhibitory (slowed) response times are found in phonological priming experiments, in which the amount of word-initial phoneme overlap between a prime and target is varied (e.g., *mark*–*must*). While facilitation appears to be strategic, there is debate as to the nature of the inhibitory priming found when high-overlap prime-target pairs are used (e.g., *musk*–*must*). Some researchers propose that this inhibitory priming is due to lexical competition between simultaneously activated candidates; others suggest it is due to the use of a response strategy. Experiments in our laboratory attempted to resolve this debate. We tested for the presence of strategic effects by manipulating variables thought to influence strategy acquisition (i.e., *isi*, proportion of overlap trials), and then by examining participants’ RTs to trials of varying over-



lap (e.g., prone–must, mark–must, muff–must, musk–must) collected over the course of the experiment. Results suggest that inhibition in the phonological priming paradigm is determined in large part by the development of a response strategy.

**4pSC30. Phonemic effects in spoken-word recognition in Japanese.** Anne Cutler (MPI for Psycholinguist., P.O. Box 310, 6500 AH Nijmegen, The Netherlands, anne.cutler@mpi.nl) and Takashi Otake (Dokkyo Univ., Soka, Saitama, 340 Japan)

Previous studies have shown that Japanese listeners are sensitive to the moraic structure of speech, and find it easier to manipulate or respond to morae than phonemes. We further examined moraic processing via two word reconstruction experiments, in which Japanese listeners heard three- or four-mora nonwords which could be changed into real words by substitution of a single mora. In experiment 1, listeners had to change the first mora of the nonword, in experiment 2 the final mora. We compared three types of substitution for CV morae: substitution preserving the C (e.g., *kimera* or *kamere* for, respectively, the first and last mora of the word *kamera*, which has three morae: ka-me-ra), substitution preserving V (*namera*, *kamena*), or substitution preserving neither (*nimera*, *kamene*). When C or V was preserved, responses were significantly faster and more accurate than when neither was preserved. In initial position, there was no difference between C- and V-preserving substitutions, but in final position, preservation of the C led to faster and more accurate responses than preservation of the V. These results confirm that spoken word recognition in Japanese is sensitive to vocabulary structure and similarity (inter alia at a submoracic level) between words.

**4pSC31. Spoken word recognition and the influence of lexical competitors at the beginning of a word.** Nadia Duenas and Michael Vitevitch (Speech Res. Lab., Dept. of Psych., Indiana Univ., Bloomington, IN 47405, mvitevitch@indiana.edu)

Marslen-Wilson and Welsh (1978) suggest that the initial portion of a word activates multiple lexical candidates in memory, forming a “cohort” of competitors. Marslen-Wilson (1987) further states that the number of competitors in the cohort does not affect the speed and accuracy of spoken word recognition. Monosyllabic CVC words with the same number of lexical competitors (using a substitution-based metric to estimate competitor set size) but varying in the percentage of competitors sharing the same initial phoneme as the target word were presented in an auditory naming task and an auditory lexical decision task. Words with a smaller proportion of competitors sharing the same initial phoneme as the target word were responded to more quickly than words with a larger proportion of competitors sharing the same initial phoneme as the target word. These results suggest that the number of candidates activated in memory does affect spoken word recognition, and that the initial portion of a word is important in processing. The implications of these results for models of spoken word recognition are discussed. [Work supported by NIH-NIDCD Training Grant DC-00012.]

**4pSC32. The basic units of rate normalization.** Jessica L. Burnham and Rochelle S. Newman (Dept. of Psych., Univ. of Iowa, E11 SSH, Iowa City, IA 52242, rochelle-newman@uiowa.edu)

We investigated the units of rate normalization, and found that they are not based on phonemes, but on segments with obvious acoustic boundaries, emphasizing the role of basic auditory processing in speech recognition. Individuals vary their speaking rate, and listeners use the duration of adjacent segments to adjust for these changes [J. Miller and A. Liber-

man, *Percept. Psychophys.* **25**, 457–465 (1979)]. However, it has not been clear what these segments actually are. We examined whether two-phoneme sequences would produce two separate rate normalization effects (one for each phoneme), or only a single effect. We first used the series *shkas–chkas*, and examined effects of /k/ and /a/ duration on perception of the initial contrast. Altering the /k/ duration resulted in a rate normalization effect separate from that caused by varying the duration of the sequence as a whole. This suggests /k/ and /a/ are treated as separate units of speech. A /w/+vowel sequence showed a different result. For sequences such as *ka*, with obvious acoustic boundaries, each phoneme has a separate rate normalization effect. However, for sequences such as *w*, without such obvious cues, the sequence is treated as a single (larger) segment by the rate normalization process.

**4pSC33. The effect of amplitude modulation on periodic- and aperiodic-formant-based sentences.** Thomas D. Carrell, Camille C. Dunn, and Jennifer Gutzwiller (Commun. Disord., Univ. of Nebraska, Lincoln, NE 68583)

Amplitude modulation has been shown to increase the intelligibility of tone-analog sentences by 30%–60% [T. D. Carrell and J. Opie, *Percept. Psychophys.* **52**, 437–445 (1992)]. Tone-analog sentences are constructed from three or four time-varying sinusoidal waveforms that trace the formant frequencies of the first several formants of a natural utterance [Remez *et al.*, *Science* **212**, 947–950 (1981)]. Because these sentences have no fundamental frequency an important auditory grouping cue is missing. It has been argued that the observed intelligibility increments were based on the ability of amplitude modulation to supply auditory grouping cues similar to those missing from tone-analog sentences. This explanation was tested with whispered sentences and with reduced-channel envelope-shaped noise (ESN) sentences [C. L. Bahner and T. D. Carrell, 138th meeting Acoust. Soc. Am. (1999)]. In an experiment comparing the effect of amplitude modulation on tone-analog sentences, whispered sentences, and reduced-channel ESN sentences, it was found that the intelligibility of tone-analog sentences was improved whereas the intelligibility of the other two types of sentences was reduced. This pattern of results calls for a distinction between the effects of amplitude modulation on sentences constructed with periodic-based formants versus those constructed with aperiodic-based formants.

**4pSC34. Interpreting Garner interference.** Shawn Weil and Mark Pitt (Dept. of Psych., Ohio State Univ., 1827 Neil Ave., Columbus, OH 43210, weil.17@osu.edu)

The Garner (1974) speeded classification procedure has long been used in psychoacoustic, speech perception, and music perception research to assess the relative integrality or separability of stimulus dimensions. Interference between dimensions is indicated by a drop in performance between an orthogonal condition (two dimensions varied independently) and a control condition (one dimension held constant, the other varied). In a meta-analysis, the ubiquity of Garner interference was examined to understand better the meaning of interference and the theoretical claims that can be drawn from it. Within this context, a set of experiments revisited the source of interference found between talker and phonetic dimensions (Mullennix and Pisoni, 1990). Individual speaker variability (simulated by multiple single-speaker tokens) and word-initial consonant (/b/ vs /p/) were manipulated.



**4pSC35. Laughter: Perceptual evaluations of laugh sounds with known acoustic variability.** Jo-Anne Bachorowski (Dept. of Psych., Vanderbilt Univ., Nashville, TN 37240) and Michael Owren (Cornell Univ., Ithaca, NY 14853)

Acoustic analysis of laughter recorded under controlled social contexts revealed significant variability on a number of source- and filter-related acoustic measures. Here, the results of three experiments used to test listeners' ( $n=84$ ) evaluations of this acoustic variability are described. The 80 laugh stimuli used in each experiment were selected on the basis of

laughter sex, percentage of voicing, F0 contour, and overall laugh duration. Listeners evaluated each stimulus twice using one of three 4-point rating schemes (goodness for inclusion in a laugh track, sincerity, and the extent of positive emotional responding). The results indicate that listeners prefer laughs with voiced, song-like qualities, and suggest that particular patterns of acoustic features in this nonlinguistic human vocal signal are more likely to engender positive emotional responses in listeners. Further analyses will examine the links between listener evaluations, variation in laugh vowel sounds, and acoustic correlates of laughter body size. [Work supported by NSF.]

THURSDAY AFTERNOON, 4 NOVEMBER 1999

GARFIELD ROOM, 1:00 TO 2:45 P.M.

## Session 4pSPa

### Signal Processing in Acoustics: Signal Processing I: Multi-Channel Techniques

John Impagliazzo, Chair

Naval Undersea Warfare Center, Code 8212, Building 679, Newport, Rhode Island 02841

#### Contributed Papers

1:00

**4pSPa1. Exact solutions for the problem of source location from measured time differences of arrival.** Ramani Duraiswami, Dmitry Zotkin, and Larry Davis (Inst. for Adv. Computer Studies, Univ. of Maryland, College Park, MD 20742, ramani@umiacs.umd.edu)

Source location using the difference in time of arrival of a signal (TDOA) at an array of receivers is used in many fields including speaker location. Using three TDOA values from four noncollinear receivers one can, in principle, solve for the unknown source coordinates in terms of the receiver locations. However, the equations are nonlinear, and in practice signals are contaminated by noise. Practical systems often use multiple receivers for accuracy and robustness, and improved S/N, and solutions must be obtained via minimization. A new family of exact solutions, for the case of four receivers located in a plane, is presented. These solutions can be evaluated using a small number of arithmetic operations. The performance of these solutions for several practical situations is examined via simulation, and possible geometries of receiver locations suggested. For the case where multiple receivers ( $>4$ ) are used, a new formulation is presented, that incorporates the present solutions, imposes additional constraints on source location, and enables use of a constrained  $L_1$  optimization procedure to achieve a robust estimate of the source location. The present estimator is compared with ones from the literature, and found to be robust and accurate, and more efficient. [Work partially supported by DARPA.]

1:15

**4pSPa2. Blind estimation of the relative travel times and amplitudes of multipath using auto- and cross-correlation functions.** John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 3440 Market St., Ste. 400, Philadelphia, PA 19104-3325, johnsr@sas.upenn.edu)

A location problem is considered where sound propagates along multipath which are impractical to model because the environment is poorly known. The acoustic bandwidth is assumed to be large enough so that the cross-correlation functions between pairs of receivers contain multiple peaks from multipath. The highest peak may not correspond to the difference in path lengths between the source and the receivers. Using similarities in the patterns of peaks in auto- and cross-correlation functions, an algorithm is developed to identify which cross-correlation peak corresponds to the difference in first arrivals, which can be used for locating the source if these arrivals are straight. The similarities are expressed with new "correlation equations." The number of lag-type correlation equa-

tions is  $O(\mathcal{R}^2 N^2)$ , where  $N$  is the typical number of multipath at each of  $\mathcal{R}$  receivers. The correlation equations may be impractical to solve exactly. Accurate solutions are found in simulations for the numbers, relative travel times, and amplitudes of all the multipath with the aid of a new fourth-moment function which is a cross correlation of non-negative lags of an auto-correlation function with lags from a cross-correlation function. The technique relies on time series which are filtered to yield one dominant source.

1:30

**4pSPa3. Race car trajectory determination by multi-source acoustic emission analysis.** Matthew Barga, Yann G. Guezennec, and Giorgio Rizzoni (Dept. of Mech. Eng. and Ctr. for Automotive Res., The Ohio State Univ., 206 W. 18th Ave., Columbus, OH 43210-1107, guezennec.1@osu.edu)

This talk presents a signal processing methodology to extract racecar trajectory and engine speed information from a multiple microphone array placed at track side. Phase differences, as well as Doppler shifts of the engine acoustic emissions from an array of microphones, provide multiple sources from which a robust determination of the vehicle trajectory can be extracted. Furthermore, time-frequency analysis of the Doppler-corrected signals provide a rich source of information about the instantaneous engine speed. The combination of both methods provides valuable information about the vehicle and engine characteristics. The analysis method used in this work is applied to acoustic emission data recorded by a track-side array of four microphones. The data analyzed include synthetic validation and calibration data, as well as actual race data. Special attention was devoted to optimizing the microphone array configuration for maximum resolution given an array size. By carefully exploiting the inherent redundancies in the data, a robust and computationally efficient analysis method was successfully developed. The results demonstrate that valuable vehicle and engine parameters can be extracted from such an approach.

1:45-2:00 Break

2:00

**4pSPa4. Normalizing inter-band scaling in minimum variance (MV) beamformer estimates of variable-bandwidth filtered wideband signals.** Laura A. Drake, Aggelos Katsaggelos (Elec. & Computer Eng. Dept., Northwestern Univ., Evanston, IL 60208-3118), and Jun Zhang (Univ. of Wisconsin-Milwaukee, Milwaukee, WI 53201)

Narrow-band minimum variance (MV) beamforming can be used to estimate a wideband signal when the signal is first filtered into several narrow-band signals. This paper identifies and solves a new problem for narrow-band MV beamforming of wideband signals. "Inter-band scaling"

can occur when narrow-band MV beamforming is applied to a wideband signal filtered by a variable-bandwidth filterbank. This situation can occur when the wideband signal to be estimated is speech. Then, it is often desirable to use a filterbank with higher-frequency resolution in the low, than in the high-frequency bands since speech has more information in low frequency than high. In this case, the power out of the MV beamformer in each band can depend not only on the actual signal power in that band, but also on the bandwidth. The inter-band scaling normalization method presented here is tested with both white Gaussian noise, and a segment of a speech signal. The tests show that the method is effective and does not distort the speech signal. Finally, this method should be extendable to other problems (such as other adaptive array processing methods) that require estimates of statistical measures of variable-bandwidth filtered signals.

2:15

**4pSPa5. A spread spectrum technique for studying sound propagation in forested areas.** Michelle E. Swearingen and David C. Swanson (Grad. Prog. in Acoust., P.O. Box 30, State College, PA 16804, swear@sabine.acs.psu.edu)

Spread spectrum methods are relatively new to acoustics, but have been shown to be quite useful, especially in determining the path taken by a particular sound pulse. Because of this property, spread spectrum signals should give a rough picture of the scattering off tree boles that occurs in a forest. The arrival times will determine the path taken, and the level changes will determine how much attenuation that path contributes. The experimental setup and preliminary data will be presented.

THURSDAY AFTERNOON, 4 NOVEMBER 1999

2:30

**4pSPa6. Constrained maximum likelihood estimation of the principal curvatures on the object surface in an ultrasonic tactile sensing.**

Kenbu Teramoto (Dept. of Mech. Eng., SAGA Univ., SAGA, 8408502 Japan) and Noriko Mori (SAGA Univ., SAGA, 8408502 Japan)

The maximum likelihood estimation (MLE) provides a robust solution in matched-field imaging (MFI), which has evolved from various underwater acoustic applications successfully. In the identification of the principal curvatures during the medical or robotic applications of acoustic tactile sensing, matched field processing offers a reasonable detection of the reflected wavefront. Two major difficulties, however, are there in the principal curvature identification. The first reason is that the wavefront reflected by the elliptic paraboloidal or the hyperbolic paraboloidal surface cannot be described in the combination of the plane waves nor the spherical waves strictly. The second is that the principal curvature identification process becomes ill posed due to the nonlinear relationship between the principal curvatures and propagation time of flight. The MFI scenario, therefore, can solve the nonlinear optimization problem in order to identify the curvature. In this paper, the proposed identification algorithm seeks the unique KKT (Karush–Kuhn–Tucker) point in the augmented Lagrange function of the constrained likelihood function, which is defined over the observed signal field. Furthermore, several acoustical experiments show that the proposed tactile sensor can identify the principal curvatures of the following surface cases: (1) plane, (2) paraboloid, (3) elliptic paraboloid, and (4) hyperbolic paraboloid.

GARFIELD ROOM, 3:00 TO 4:15 P.M.

**Session 4pSPb**

**Signal Processing in Acoustics: Signal Processing II: Applications in Acoustics**

Charles F. Gaumont, Chair

*Naval Research Laboratory, Code 7142, Acoustics Division, Washington, DC 20375-5320*

**Contributed Papers**

3:00

**4pSPb1. A minimum-variance frequency-domain algorithm for binaural hearing aid processing.** Michael E. Lockwood, Douglas L. Jones, Mark E. Elledge, Robert C. Bilger, Marc Goueygou, Charissa R. Lansing, Chen Liu, William D. O'Brien, Jr., and Bruce C. Wheeler (Beckman Inst., Univ. of Illinois at Urbana–Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

A new algorithm has been developed that allows optimal filtering methods to be applied individually to different narrow frequency bands. Using this technique it is possible to process binaural signals in a manner which dramatically reduces the amplitude of signals originating away from a desired receive direction. The algorithm was tested on artificially combined anechoic and reverberant signals with varying numbers of interfering sound sources. In addition to being computationally efficient, the algorithm was able to produce output which had a consistently positive intelligibility weighted SNR gain [Link and Buckley, *J. Acoust. Soc. Am.* **91**, 1662 (1992)], and in which the desired talker was noticeably easier to understand. Results of a paired-comparison listening test confirmed these results, and allowed an estimate to be made of the algorithm parameters which provided the best speech intelligibility for the output. These parameters included the window length, length of a filtered block, and amount of data which must be stored in memory. Collectively, the results show that

this algorithm may be the first building block for a binaural hearing-aid processing algorithm which suppresses off-axis interference. [Research supported by the Beckman Institute.]

3:15

**4pSPb2. The effect of reverberation on a new binaural noise cancellation algorithm for hearing aids.** Marc Goueygou, Michael E. Lockwood, Mark E. Elledge, Robert C. Bilger, Douglas L. Jones, Charissa R. Lansing, Chen Liu, William D. O'Brien, Jr., and Bruce C. Wheeler (Beckman Inst., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

A new algorithm was developed to emulate the “cocktail party effect” in hearing aids [Liu *et al.*, *ASA Symposium* (1997)]. Speech signals are received binaurally; the interferers and the desired source are localized; and the signal from the strongest interferer is cancelled by frequency and time selective beamforming. Here we evaluate the algorithm’s performance in real conditions including reverberation. Binaural recordings of speech uttered simultaneously by three different speakers around the receivers were made in three different rooms (an anechoic room, a mildly and a highly reverberant room). Noise cancellation was evaluated by the intelligibility weighted SNR (IWSNR) gain between processed and unprocessed signals [Link and Buckley, *J. Acoust. Soc. Am.* **91**]. Algorithm performance is severely degraded as the reverberation time increases: the IWSNR gain falls from 8–9 dB in the anechoic room to 1–2 dB in the highly reverberant room. Also, knowing of the exact location of the

sources does not add significant improvement. The results suggest that the room impulse response be deconvolved from the received signals to suppress the multiple reflections of sound. We suggest a new strategy combining noise cancellation with blind identification of the room impulse response. [Research supported by the Beckman Institute.]

3:30

**4pSPb3. A real-time dual-microphone signal-processing system for hearing aids.** Mark E. Elledge, Michael E. Lockwood, Robert C. Bilger, Marc Goueygou, Douglas L. Jones, Charissa R. Lansing, William D. O'Brien, Jr., and Bruce C. Wheeler (Beckman Inst., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

A real-time dual-microphone-based signal-processing system has been developed that suppresses off-axis interference from multiple sound sources. The system employs optimal filtering methods to attenuate the amplitude of all signals other than that from a desired receive direction. The algorithm was implemented on a Texas Instruments TMS320C62 fixed-point digital signal-processor evaluation board installed in a PC. Input/output was accomplished via a 16-bit codec controlled by optimized assembly language routines. To simplify algorithm coding, the main algorithm was written in C and executes using approximately 55 processor time, allowing for future additions to enhance performance. It has a delay of 15 ms. The system can significantly improve the intelligibility of speech in noisy environments, as evidenced in off-line tests by a marked intelligibility weighted SNR gain [Link and Buckley, *J. Acoust. Soc. Am.* **91**, 1662 (1992)]. Initial real-time testing with actual binaural signals supports the conclusion that this performance gain will be maintained. The results show that this system may be used as a main building block for hearing aids to be used in noisy environments. [Research supported by the Beckman Institute.]

3:45

**4pSPb4. Multimedia model for speech transmission through walls.** Jiann-Ming Su and J. H. Ginsberg (The George W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The paper will describe how an algorithm previously developed to study the distortion of a planar stress wave as it propagates at normal incidence through a multi-layered wall was adapted to study transmission of speech signals. The algorithm is a time domain numerical implementa-

tion of the method of characteristics [Ginsberg and Kim, *J. Appl. Mech.* **11**(2), S145-S151 (1992)]. The algorithm is used to produce a digitized waveform of the exiting signal corresponding to an incident signal consisting of a speech fragment. To assess the effects of distortion resulting from internal reflections, a speech pattern is digitized with the aid of standard multimedia tools and used as the input to the transmission algorithm. The transmitted signal is computed and converted to a convenient audio format for playback. To ensure that distortion in the output signal is truly the effect of the medium, the signal is processed again using a noise reduction routine developed by Boll [*IEEE Trans. ASSP ASSP-27*(2), 113-120 (1979)]. In a blind test, listeners are requested to identify, from both the raw and processed output, the original speech fragment. A demonstration will conclude the presentation.

4:00

**4pSPb5. Further development of time delay spectrometry.** Bradley Finch, Gareth Cook, and Anthony Zaknich (Ctr. for Intelligent Information Processing (CIIPS), Univ. of Western Australia, Nedlands 6009, Australia, finch@ee.uwa.edu.au)

The accuracy of conventional time delay spectrometry (TDS) measurements are limited by constraints on signal parameters. These constraints ultimately limit the resolution with which the frequency response can be measured. Poletti, Cook, and others have shown that more exact measurements, without these constraints, are possible. These theories require that the system response to the full complex chirp be known. This is usually done by exciting the system with two separate orthogonal sweep signals. In this paper, this theory is developed in a manner similar to that done by Vanderkooy for the conventional TDS measurement, to show that the more exact system response can be deduced from a single linear sweep. The new development is supported by experimental results comparing the measured system response of a series resonant circuit against the conventional TDS results originally reported by Vanderkooy. This method provides practitioners more convenience in making exact TDS measurements, and allows for further exploration of the application of this technique in time varying environments.

THURSDAY AFTERNOON, 4 NOVEMBER 1999

FAIRFIELD ROOM, 1:30 TO 5:15 P.M.

### Session 4pUW

## Underwater Acoustics, Acoustical Oceanography and Animal Bioacoustics: The Effect of Man-Made Sound on Marine Mammals II

Peter L. Tyack, Chair

*Department of Biology, Woods Hole Oceanographic Institution, 45 Water Street, Woods Hole, Massachusetts 02543-1049*

Chair's Introduction—1:30

### Invited Papers

1:35

**4pUW1. Acoustic responses of Baleen whales to low-frequency, man-made sounds.** Christopher W. Clark (Bioacoust. Res. Prog., Cornell Univ., Sapsucker Woods Rd., Ithaca, NY 14850), Peter L. Tyack (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and William T. Ellison (Marine Acoustics, Inc., P.O. Box 340, Litchfield, CT 06759)

In the last 5 years, two projects were undertaken to evaluate impacts of man-made sounds on whales. Baleen whales were identified as "at risk" because of their use of low-frequency sound for communication, their endangered status, and studies showing responses to continuous noises at exposure levels  $>120$  dB,  $re: 1 \mu\text{Pa}$ . To evaluate the potential impact of operational ATOC (195-dB source intensity), humpback whales off Kauai were studied during the breeding season. To evaluate the potential impact of

U.S. Navy SURTASS LFA sonar, playback experiments were conducted on four species at exposures of 120–155 dB. LFA research was designed to obtain responses during feeding (blue and fin whales, southern California, September–October), migration (gray whales, central California, January) and breeding (humpbacks, Hawaii, March). For the ATOC source, humpbacks showed statistically significant but subtle responses over small time and spatial scales. There were no changes in singing, or larger-scale changes in distribution or relative numbers. For LFA experiments off southern California, whales did not change vocal rates or leave the testing area, and there were no immediately observable responses, even at exposure levels up to 150 dB. Tyack (this session) will discuss results from the gray and humpback whale LFA experiments.

2:05

**4pUW2. Responses of Baleen whales to controlled exposures of low-frequency sounds from a naval sonar.** Peter L. Tyack (Biol. Dept., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, [ptyack@whoi.edu](mailto:ptyack@whoi.edu))

Playback experiments were conducted from 8–27 January 1998 using a moored SURTASS-LFA sound source to study behavioral responses of gray whales migrating off Pt. Buchon, California. Shore stations operated for over 150 h during 18 days, tracking about 1200 migrating whales. When the source was moored inshore near most migrating whales, whales avoided exposure to 42-s sound stimuli in the 160–330-Hz frequency band repeated every 6 min at received levels near 120 dB (playback source levels 170–185 dB *re*: 1  $\mu$ Pa at 1 m). Responses to the source broadcasting offshore at source levels of 185–200 dB were greatly reduced. Playbacks to humpback whales singing in Hawaiian waters occurred in February–March 1998. 17 playbacks and 5 control follows of singing humpbacks were conducted from a small vessel. Singers continued to sing throughout 7/17 playbacks, and stopped during the remaining 10 playbacks. During 4/10 playbacks when singers stopped, the singer stopped when it joined with another whale. The remaining 6 cessations of song were considered possible responses to playback. Most of these whales resumed normal behavior before the hour-long playback ceased. There is no trend for the vocal response to playback to scale with exposure level; both the “vocal-response” and “no-vocal-response” categories ranged in exposure from 120–150 dB; *re*: 1  $\mu$ Pa.

2:35

**4pUW3. Marine mammal research program for the Pioneer Seamount ATOC experiment.** Daniel Costa (Inst. of Marine Sci., Univ. of California, Santa Cruz, CA 95064) and John Calambokidis (Cascadia Research Collective, Olympia, WA)

Aerial surveys and satellite and archival tags were used to test the effect of the ATOC sound source on marine mammals around the Pioneer Seamount 85 km west of San Francisco, California. Control surveys were flown at least 48 h after the end of any previous transmission cycle and experimental surveys were flown after at least 24 h of sound transmissions. Sound transmissions consisted of 20-min periods of 195-dB, that is, 1- $\mu$ P transmission repeated every 4 h. Most commonly sighted species by group were: (1) mystecetes: humpback whale (372 sightings), (2) large odontocetes: sperm whale (337 sightings), (3) small odontocetes: Pacific white-sided dolphin (306 sightings), and (4) pinnipeds: California sea lion (167 sightings). Although there were no significant differences in the number of sightings when the sound source was on or off, both humpback and sperm whales were generally seen farther from the sound source during experimental versus control surveys ( $p < 0.01$ ). The highest intensity of sound measured on tags carried by elephant seals during transmissions ranged from 118–137 dB for 60–90 Hz compared to ambient levels of 87–107 dB (60–90 Hz). On a gross level, animals did not alter return track or go to the surface, and often continued to dive closer to the sound source if on the descending segment of a dive evident. Some animals with the highest levels of exposure showing no effects while two animals showing minor changes in diving pattern had exposure levels greater than 132 dB. [Work supported by the Strategic Environmental Research and Development Program through RPA and the Office of Naval Research.]

3:05

**4pUW4. The influence of seismic survey sounds on bowhead whale calling rates.** Charles R. Greene, Jr. (Greeneridge Sci., Inc., 4512 Via Huerto, Santa Barbara, CA 93110), Naomi S. Altman (Cornell Univ., Ithaca, NY 14853), and W. John Richardson (LGL Ltd., Environ. Res. Assoc., King City, ON L7B 1A6, Canada)

During September in 1996–98, airgun arrays operated close to the coast of the Alaskan Beaufort Sea. It was of interest to know if the airgun pulse sounds affected migrating whales. Autonomous seafloor acoustic recorders were installed at various distances from shore and from the airguns. They recorded continuously for up to 22 days. Both the bowhead calls and the airgun pulses were clustered in time, requiring a greater quantity of data to isolate an influence than would be the case with nonclustered data. The consistent results from the three years were: (1) bowhead whales called frequently during their migration through the study area; (2) calling continued, although possibly at different rates, when the whales were exposed to airgun pulses; and (3) call detection rates at some locations differed significantly when airgun pulses were detectable versus not detectable. At least in 1996, patterns of call detection rates were consistent with aerial survey evidence that most bowheads were displaced from the area near the operating airguns. However, there was no significant tendency for call detection rates to change consistently at times when airgun operations started or stopped. [Work supported by BP Exploration (Alaska) and by Western Geophysical.]

3:35–3:50 Break



3:50

**4pUW5. Modeling cetacean ear filters by means of evolutionary computation.** Dorian S. Houser, David A. Helweg, Patrick W. B. Moore (SPAWARSYSCEN-San Diego, Code D351, 53560 Hull St., San Diego, CA 92152-5435), and Kumar Chellapilla (Univ. of California at San Diego, La Jolla, CA 92093-4007)

Modeling cetacean hearing is an important step in advancing models of echolocation and in predicting the potential impact of oceanic noise upon cetaceans. Modeling the response of cetacean ears to acoustic input can be a difficult task and the performance of such models is often sub-optimal. Evolutionary programming was employed as a systematic method of optimizing the response of a bank of pseudo-Gaussian filter shapes to known audiometric functions in order to model the resonant systems of the basilar membrane. The response of the filter banks was tested against normalized audiograms of the bottlenose dolphin (*Tursiops truncatus*) and the predicted frequency dependent sensitivity of the Humpback whale (*Megaptera novaeangliae*). Filter banks were created that demonstrated comparable frequency responses to tonal stimuli across the (predicted) range of hearing in both species. The potential for incorporating these ear models into models of echolocation and their use in predicting the sensitivity of baleen whales to manmade noise will be discussed.

4:05

**4pUW6. An inexpensive, portable, and rugged system for recording the low-frequency sounds of cetaceans.** Catherine L. Berchok (Grad. Prog. in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804, berchok@sabine.acs.psu.edu), Thomas B. Gabrielson, and David L. Bradley (The Penn State Univ., State College, PA 16804)

Many studies have been done on the low-frequency sounds of cetaceans. However, few of these studies have involved coincident photoidentification, biopsy, and behavioral observation. In order to make acoustical recordings in conjunction with the regular visual fieldwork done from inflatable boats on the baleen whales of the St. Lawrence, a recording system was designed to meet several requirements. It had to be very inexpensive; light enough to be carried onto the boat each day and deployed by hand; and rugged enough to survive being pounded and soaked when stored on the open boat. The resulting system consisted of two main parts: a recording system made up of DAT recorder, amplifiers, filters, batteries and voltage regulators; and a surface-motion isolation system made up of a spar buoy, omnidirectional hydrophone, cable and damping plate. A mixing circuit allowed for real-time detection of infrasonic signals through a pair of headphones. In addition to the description of the system and its calibration, samples of blue, fin, minke and humpback whale recordings will be presented, and a brief comparison of the system with and without the spar buoy will be shown. [Work supported by a graduate fellowship from NSF.]

4:50–5:15

#### Panel Discussion

4:20

**4pUW7. Displacement of migrating bowhead whales by sounds from seismic surveys in shallow waters of the Beaufort Sea.** W. John Richardson, Gary W. Miller (LGL Ltd., Environ. Res. Assoc., P.O. Box 280, King City, ON L7B 1A6, Canada), and Charles R. Greene, Jr. (Greeneridge Sci., Inc., Santa Barbara, CA 93110)

Seismic surveys for subsea oil deposits were conducted each summer, 1996–98, mainly in water <20 m deep. Airgun arrays were used, with 6–16 airguns and total volumes 560–1500 cu.in. Low-frequency sound pulses were created at intervals of 8–20 s. Effective source levels for horizontal propagation were lower than nominal source levels, but pulses were often detectable to 50+ km offshore. Westward autumn migration of bowhead whales near and offshore of the exploration area was monitored by aerial surveys flown daily, weather permitting, during the three seasons. Aerial survey data from days with and without airgun operations were compared. Most bowheads avoided the area within 20 km of the operating airguns; bowheads were common there on days without airgun operations. In 1998, numbers sighted 20–30 km away were also significantly reduced during airgun operations. Conversely, sighting rates just beyond the avoidance zone were higher on days with airgun operations. Broadband received levels of airgun pulses at 20 km were typically 120–130 dB *re*: 1  $\mu$ Pa (rms over pulse duration), lower than those previously demonstrated to cause avoidance by bowheads. Many migrating bowheads 20 to 50+ km offshore were exposed to weaker but presumably detectable pulses. [Work supported by Western Geophysical and BP Exploration (Alaska), Inc.]

4:35

**4pUW8. Predicting the effects of sound exposure on fish, marine mammals, and human beings.** Antoine David, Peter Ward, Tony Heathershaw (DERA, Southampton Oceanogr. Ctr., Southampton, UK), and Peter Varley (Element Ltd., Fareham, UK, elements@freenet.uk.com)

Fish and marine mammals are sensitive to sound and may be affected by sound energy from a wide range of sources. Current concern is focused on anthropogenic sound and in particular sound which emanates from active sonar devices. This may range from broadband short-duration events to narrow-band signals with long-time exposure. Frequencies may extend from a few hundred Hz to a few hundred kHz. In order to be able to predict the effect of sound on fish, marine mammals, and human beings, it is first of all necessary to predict the propagation of sound through the ocean environment and then to estimate the effect of a received sound pressure level on an environmental receptor. To achieve this latter objective, a generic threshold of hearing model has been developed which is suitable for use with sonars of arbitrary frequency and duty cycle. The model, which is described in this paper, relates the effects of sound in both the frequency and time domains. This enables the onset of TTS, PTS, and an equivalent damage risk criteria to be predicted as a function of the distance between an environmental receptor and a sound source.

**Meeting of Accredited Standards Committee (ASC) S1 on Acoustics**

G. S. K. Wong, Chair S1

*Institute for National Measurement Standards (INMS), National Research Council, Ottawa, Ontario K1A 0R6, Canada*P. D. Schomer, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics  
*U.S. CERL, P.O. Box 9005, Champaign, Illinois 61826-9005*

T. J. Kuemmel, Vice Chair S1

*Quest Electronics, 510 South Worthington, Oconomowoc, Wisconsin 53066*H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics  
*1325 Meadow Lane, Yellow Springs, Ohio 45387*V. Nedzelnitsky, U.S. Technical Advisor (TA) for IEC/TC 29, Electroacoustics  
*National Institute of Standards and Technology (NIST), Sound Building, Room A147, 100 Bureau Drive, Stop 8221, Gaithersburg, Maryland 20899-8221*

**Accredited Standards Committee S1 on Acoustics.** Working group chairs will report on their preparation of standards on methods of measurement and testing, and terminology, in physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound. Work in progress includes measurement of noise sources, noise dosimeters, integrating sound-level meters, and revision and extension of sound level meter specifications. Open discussion of Committee reports is encouraged.

**Scope of S1:** Standards, specifications, methods of measurement and test and terminology in the field of physical acoustics including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance and comfort.

**Meeting of Accredited Standards Committee (ASC) S3 on Bioacoustics**

R. F. Burkard, Chair S3

*Hearing Research Laboratory, State University of New York at Buffalo, 215 Parker Hall, Buffalo, New York 14214*

J. Franks, Vice Chair S3

*Robert A. Taft Laboratories, 4676 Columbia Parkway, Mail Stop C27, Cincinnati, Ohio 45226*P. D. Schomer, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics  
*U.S. CERL, P.O. Box 9005, Champaign, Illinois 61826-9005*D. D. Reynolds, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC4, Human Exposure  
to Mechanical Vibration and Shock  
*3939 Briar Crest Court, Las Vegas, Nevada 89120*H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics and ISO/TC 108/SC4,  
Human Exposure to Mechanical Vibration and Shock  
*1325 Meadow Lane, Yellow Springs, Ohio 45387*V. Nedzelnitsky, U.S. Technical Advisor (TA) for IEC/TC 29, Electroacoustics  
*National Institute of Standards and Technology (NIST), Sound Building, Room A147, 100 Bureau Drive, Stop 8221, Gaithersburg, Maryland 20899-8221*

**Accredited Standards Committee S3 on Bioacoustics.** The current status of standards under preparation will be discussed. In addition to those topics of interest, including hearing conservation, noise, dosimeters, hearing aids, etc., consideration will be given to new standards which might be needed over the next few years. Open discussion of Committee reports is encouraged.

**Scope of S3:** Standards, specifications, methods of measurement and test, and terminology in the fields of mechanical shock and physiological acoustics, including aspects of general acoustics, shock, and vibration which pertain to biological safety, tolerance, and comfort.

**Session 5aAA****Architectural Acoustics: Tribute to Buzz Towne—Classroom Acoustics**

William J. Cavanaugh, Cochair

*Cavanaugh Tocci Associates, Inc., 327F Boston Post Road, Sudbury, Massachusetts 01776*

Angelo J. Campanella, Cochair

*Campanella Associates, 3201 Ridgewood Drive, Columbus, Ohio 43026***Chair's Introduction—8:30*****Invited Papers*****8:35****5aAA1. An overview of the development for classroom acoustics guidelines.** Mike Nixon (EA. Acoust. Eng., Inc., 2810 Urbandale Ln., Plymouth, MN 55447)

This is an overview of past, current and future events that will impact the development of federal acoustical guidelines in classrooms and other facilities in the USA. In 1996 the late Buzz Towne presented clear and compelling evidence to the ASA membership of the need for a greater focus on classroom acoustics. Since that time great progress has been made to reduce acoustical barriers to more effective learning. The TCAA subcommittee on Classroom Acoustics, convened in 1996, opened a path of interdisciplinary communication and cooperation that resulted in the formation of a multidisciplinary ASA Taskforce to further the agenda. In 1997 and 1999 two highly successful interdisciplinary workshops were held in Los Angeles and New York. In 1998 ASA, in collaboration with other professional organizations, mounted a formal response to the US Access Board's Request For Information on Classroom Acoustics. In March 1999 the Access Board proposed a Course Of Action that resulted in a request to ANSI to develop acoustical guidelines for classrooms within two years. The ANSI approved S12 W/G-42 committee is already working on the development of acoustical standards. Beyond this current task, additional standards may be considered for other facilities.

**9:05****5aAA2. Acoustical consulting in the Pacific Northwest: Remembrances of Buzz Towne (1919–1998).** Herbert Chaudiere and Roy Richards (3208 15th Ave W., Seattle, WA 98119, brc@brcacoustics.com)

Robin M. (Buzz) Towne was truly the icon of acoustical consulting in the Pacific Northwest. His intense interest in acoustics led him from a mechanical engineering consultant to a career in acoustics. He formed the first full-time acoustical consulting firm in the Pacific Northwest in 1960 with just one other person, but built the firm to 15 consultants within ten years. The firm continues to prosper today under new leadership, but the legacy of Buzz remains in the minds of those that worked with him and the clients he served. Buzz was a dedicated consultant whose drive for acoustical perfection resulted in many colorful (and often times amusing) confrontations with both staff and clients. This paper recalls some of these incidents, as well as describing a few of his significant consulting projects.

**9:35****5aAA3. Good classroom acoustics are a good investment for America.** David Lubman (David Lubman & Assoc., 14301 Middletown Ln., Westminster, CA 92683-4514, dlubman@ix.netcom.com) and Louis C. Sutherland (Consultant in Acoust., Rancho Palos Verdes, CA 90275-3908)

Buzz Towne's idealism sparked ASA's current activities to improve classroom acoustics, including ANSI S12 WG43's activity to produce an American standard for classroom acoustics. But idealism alone may not suffice to realize the reforms Buzz sought. It will help if advocates can show that good classroom acoustics are a good investment for community and nation. Absent were the resources necessary for serious economic cost-benefit studies, some very informal, "back-of-the-envelope" engineering estimates were made by acousticians, audiologists, and material vendors. All assume 20 year life cycles for new and renovated classrooms. In one scenario, costs for quiet HVAC and sound absorbing ceilings are shown to be a small fraction of costs for ordinary school construction yielding substandard acoustics. In another scenario, costs for quiet classroom HVAC are shown to be small compared to annual operating costs per student. A third scenario shows that a modest but plausible assumed increase in average lifetime earnings generously will repay the initial costs for good acoustics. These scenarios do not consider the economic costs of bad acoustics, including high dropout rates, truancy, juvenile crime, and teacher burnout. The authors hope to inspire others to more fully study the economic, social, and educational benefits of good acoustics.

10:20

**5aAA4. Optimal acoustic conditions in classrooms.** Gary W. Siebein, Martin A. Gold, Michael G. Ermann (Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611-5702), Kerry N. Siebein, and Glenn W. Siebein (Siebein Assoc., Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607)

Computer models of eight classrooms were constructed with varying combinations of absorbent materials on the surfaces of the rooms using the CATT 7.0 computer program. The amount, location and cost of finish materials and the noise levels produced by the air-conditioning system were varied in the models. RASTI values were calculated for six seats in each room. By determining conditions that result in maximum RASTI values, maximum speech intelligibility conditions were obtained. Background noise levels dominated the RASTI values at levels of NC 40 and higher regardless of room finishes. At RASTI values of NC 35 and less, once an amount of absorbent material equal in area to the floor was present, only small changes in RASTI values were observed whether the material was located on the ceiling or the upper parts of the walls or a combination of both. The cost of material and locations of typical furnishings were also considered in the process.

10:50

**5aAA5. The passion behind changing the sound of education.** Karen L. Anderson (Educational Audiologist, 3252 Cranleigh Dr., Tallahassee, FL 32308)

Educational audiologists have been fighting the war of poor classroom acoustics for over 20 years. Research has amassed in the audiology literature evidencing the negative effects that inadequate acoustic treatment in classrooms have on student learning, behavior and attention. Buzz Towne dared to cross professional demarcations and take to heart the lessons of the audiology research. This presentation is made as a tribute to Buzz, by the audiologist who spent hours with him, having lively discussions as two bodies of research were compared. Buzz Towne's interest in the plight of children forced to learn in inadequate learning spaces turned into a passion. The Sound & Vibration, January 1997 article, "The Changing Sound of Education," was born of this meeting between acoustics and audiology. This passion demanded that the sound of learning was harming our children and our country's future. It forced the realization that education has changed and acoustical designs of 50 years ago are preventing today's children from achieving all that they need to learn for tomorrow. Buzz was instrumental in starting the wave of interest in classroom acoustics that has awakened acousticians across the country that will bear fruit in improved learning of millions of children.

### Contributed Paper

11:20

**5aAA6. Structural variables and their relationship to background noise levels and reverberation times in unoccupied classrooms.** Heather A. Knecht, Gail M. Whitelaw, Lawrence L. Feth (Dept. of Speech and Hearing Sci., Ohio State Univ., Columbus, OH 43210), and Peggy B. Nelson (Univ. of Maryland, MD 21201)

The classroom environment is often filled with deterrents that hamper a child's ability to listen and learn. These deterrents include background noise and reverberation leading to poor signal-to-noise ratios that can interfere with learning. By examining the acoustical conditions for speech communication in the classroom, it is evident that the acoustical environment in classrooms can affect the achievement and educational perfor-

mance of children with sensorineural hearing loss (SNHL), children with normal-hearing sensitivity who have other auditory learning difficulties, as well as elementary school children with no verbal or hearing disabilities. The purpose of this study was to measure reverberation and background noise levels in 32 different unoccupied elementary classrooms in eight different public school buildings in Central Ohio. These measurements were then compared to the external and internal criteria variables developed [Crandell *et al.*, "Sound-Field FM amplification theory and practical applications" (1995)] to determine if a relationship existed between classroom background noise levels and reverberation and the building criteria variables. Data were analyzed for overall noise level, reverberation time, and correlation between the reverberation time and internal/external criteria variables.

11:40–12:00

### Panel Discussion



## Session 5aMUa

## Musical Acoustics: Topics in Voice and Perception

Peter L. Hoekje, Chair

*Department of Physics and Astronomy, Baldwin-Wallace College, 275 Eastland Road, Berea, Ohio 44017*

## Contributed Papers

8:00

**5aMUa1. A mathematical model of singers' vibrato based on waveform analysis.** Jose A. Diaz, Howard B. Rothman, and A. Antonio Arroyo (Univ. of Carabobo, Venezuela, VLN 258, P.O. Box 025685, Miami, FL 33102-5685, jadiaz@thor.uc.edu.ve)

There is controversy regarding the best model to synthesize vibrato. Some authors include random components in their models while others suggest a deterministic model. Also, the vibrato waveform has not been studied in detail, that is, no measures have been made as to symmetry of the waveform, and parameters different from wave frequency and amplitude have not been studied. Therefore, research was undertaken to thoroughly analyze the vibrato waveform from selected singers in order to obtain a model based on the vibrato parameters. Seven premier-level singers were selected and three samples per singer were used. These samples were analyzed through the MMSV (Mathematical Model of Singers' Vibrato) software, which was specifically designed for vibrato analysis. The results obtained for all the samples were analyzed and compared, and a model fitting all of them was proposed. This model was compared against a sinusoidal model and the real wave for two different samples by the use of an error measure and visually, and it was verified that the model reproduced the variations of the real vibrato wave and significantly reduced the error measure of the sinusoidal model. The results show that a deterministic model fits all the samples under study.

8:15

**5aMUa2. The virtual environment as a musical instrument.** Jon W. Mooney (Acoustic by JW Mooney, 3912 Miami Ave., Ste. No. 2, Cincinnati, OH 45227) and Pedro Rebelo (The Univ. of Edinburgh)

Naturally occurring acoustic environments helped to formulate ancient man's musical rituals. Man-made acoustical environments shaped his musical style. Theories of architectural acoustics allowed the design of acoustical environments to suit specific musical styles. Using auralization, man's acoustic environment can now be changed at will and thus becomes a musical instrument itself. This project was carried out in two steps. In the first, the model and the listener path through the model were selected beforehand, and the music was not processed in real time. In this mode, auralization serves as a post-production mixing technique. In the second step, the musician was in control of his music, the environment and the listener attributes with the auralization feedback in real time. Here, the acoustic environment becomes a true musical instrument. Standard 3D CAD models of environments are imported into the Acoustic Rendering (TM) program which creates impulse responses of the environment as multiple functions of time, frequency, source types, locations and orientations, and listener types, locations and orientations. The musical source is convolved with the appropriate impulse responses and the response is played back through a multispeaker system based on the Dolby (R) Surround Sound Format.

8:30

**5aMUa3. Melodic expectancy in infancy.** Tonya R. Bergeson (Dept. of Psych., Univ. of Toronto at Mississauga, 3359 Mississauga Rd. North, Mississauga, ON L5L 1C6, Canada, tonya@psych.utoronto.ca)

Based on Eugene Narmour's model of melodic expectancy, it might be predicted that infants would follow the same patterns of expectancy as do adults, at least with respect to the hypothesized universal principles. In experiment 1, 6- to 7-month-old infants and adults were briefly familiarized with two melodies: one that fulfilled specific expectancies and one that violated expectancies (direction and interval size), according to Narmour's model. Infants and adults were subsequently required to detect a pitch change in the same location of each melody. Infants were able to detect the change in both melodies, and performance between melodies did not differ significantly. Adults were also able to detect the change in both melodies, but their performance was significantly better in the melody that fulfilled expectations. A confounding of expectation-fulfillment with stream segregation in the original melodies warranted further experiment. In experiment 2, as in experiment 1, 6- to 7-month-old infants detected the changes in both new melodies, and performance across conditions still did not differ. The present findings call into question Narmour's claim his bottom-up principles are innately specified.

8:45

**5aMUa4. Learning to sing a novel piece of music facilitates playing it on the violin but not the other way around: Evidence from performance segmentations.** Carolyn Drake, Stephen McAdams (Laboratoire de Psychologie Experimentale, 28 rue Serpente, 75006 Paris, France), and Alain Berthoz (College de France)

Twelve 11- to 14-year-old music students learned to perform a novel piece of music, six on the violin (initial-violinists), six singing (initial-singers), with one recording per week for five weeks. Then, without prior warning, they performed the same piece on the other instrument (sung if they had learned on the violin and visa versa). All performance measures improved over the first five performances: fewer errors, faster, and comprised of fewer, longer, more musically valid segments (as indicated by final-group lengthening). Seven children (five initial-singers, two initial-violinists) demonstrated transfer (performances did not deteriorate after changing instrument), five children (one initial-singer, four initial-violinists) did not (performance deteriorated after changing instrument). The children who transferred successfully performed slower and made more errors than those that did not transfer successfully, but their segmentations demonstrated more hierarchically structured mental representations. Thus, learning mode influenced the creation of the mental representation. Singing, an easy motor task, enables children to concentrate on musical structure and therefore to create musically appropriate mental representations specific to the music rather than the motor task, allowing transfer to other motor tasks. However, playing the violin, a complex motor task, results in task-specific representations, difficult to transfer to other motor tasks.

**5aMUa5. Formant modification of sung vowels in four languages.** Catherine I. Watson (Speech, Hearing, and Lang. Res. Ctr., Macquarie Univ., NSW 2109, Australia) and C. William Thorpe (Univ. of Sydney, Australia)

Formant frequencies of sung vowels are modified at higher pitches because of formant tuning, leading to a reduction in intelligibility as the formants converge with increasing  $F_0$ . Formant frequencies are also dependent on language-specific pronunciation requirements. The question of whether language-specific characteristics of sung vowels are maintained through pitch-effected formant tuning has not previously been examined. Sustained vowels from five similar positions on the vowel quadrilateral in four languages, English, Dutch, French, and Hungarian, were sung in a similar singing style over a range of pitches by a professional female singer who was proficient in singing each of these languages. Formant frequencies were extracted and the magnitudes of the voice partials computed from the average spectrum. The results showed an increase in  $F_1$  for the high vowels of all languages, but in the mid-low vowels there were some language differences in the direction of vowel shift as  $F_0$  increased. In addition, there were language differences in the relative magnitudes of the partials of all five vowels, particularly at higher frequencies. This suggests that some of the language-specific characteristics of sung vowels may be due to differences in style or voice characteristics.

**5aMUa6. Visual feedback in teaching singing using speech technology.** C. William Thorpe (NVC, Univ. of Sydney, NSW 2006, Australia), Jean Callaghan (Univ. of Western Sydney, Australia), and Jan van Doorn (Univ. of Sydney, Australia)

Visual feedback of voice parameters has been shown to be useful in singing training, but most current systems are designed for speech applications. Extraction of parameters such as the fundamental ( $F_0$ ) and formant frequencies can be problematic in singing because of the overlap between  $F_0$  and the formants over much of the pitch range. This has limited the application of visual feedback tools in the teaching of singing. This experiment aimed to obtain empirical data on the boundaries within which standard speech-oriented visual feedback systems can be usefully utilized for teaching singing. Four singing teachers learned to operate a simplified set of visual feedback modes on the Computer Speech Laboratory (Kay Elemetics), and then used the system in several lessons, after which they were interviewed about the utility of the system. They found that feedback of  $F_0$  and vowel formants were the most useful modes, in spite of limitations in the precision of pitch feedback and the reliability of formant extraction at high  $F_0$ . The results suggest that current speech-based visual feedback systems can be usefully integrated into singing pedagogy, but that even greater benefits would accrue if more precise  $F_0$  and reliable formant-identification algorithms were developed.

**5aMUa7. Melody-tone relation in Cantonese songs.** Patrick C. M. Wong and Randy L. Diehl (Dept. of Psych., Univ. of Texas, Austin, TX 78712)

How is fundamental frequency ( $F_0$ ) being used to signal phonemic tones when they occur in songs? In an examination of Cantonese songs, it was found that songwriters abandon the "absolute" pitch intervals of tones in speech and use "relative" intervals instead in song writing. For example, a midtone that is regularly 12% higher than a low tone in speech can be realigned as any higher  $F_0$  (but never a lower  $F_0$ ) in songs. This analysis was confirmed by a perceptual experiment. Listeners were asked to listen to six different melodies accompanied by a semantically neutral lyric: /ha6 yat1 go3 zi3 hai6 si3/ "The next word is to try," and were asked to identify the last word. (The number signals tone; "3" is a mid-tone and "6" is a low tone.) When the last pitch in the melody was higher than the note before, the last word (/si3/) was identified as either a high tone or a midtone (because they are higher tones than the low tone associating /hai6/). When it was lower in pitch, the last word was identified as a low tone. [Work supported by NIDCD.]

**5aMUa8. Chords as spectra; harmony as timbre.** Pantelis N. Vassilakis (UCLA, Music Percept. and Acoust. Lab., Dept. of Ethnomusicology, 2539 Schoenberg Hall, Los Angeles, CA 90095, pantelis@ucla.edu)

Traditionally, musical chords have been understood as combinations of discrete tones or as the layering of intervals. From the listener's point of view, however, chords often fuse into a single perceptual unit with a specific, readily recognized character (i.e., major/minor.) When chordal fusion occurs, a chord's distinct character can be understood as its timbre in terms of a single inharmonic spectral distribution, rather than a combination of discrete harmonic spectra. The present study examines the relationship between spectral distribution described in terms of (i) spectral centroid (center of amplitude-weighted frequency distribution); (ii) spectral bandwidth (spread of frequency distribution); (iii) spectral density (average number of components per critical band); and (iv) spectral inharmonicity (defined in terms of sensory consonance/dissonance curves) — and perceptual identity of complex tone combinations. It addresses the perceptual equivalence of chord inversions (root, first, second inversion of a single chord category) and the perceptual difference between different chord categories (major/minor) in terms of spectral inharmonicity rather than interval layering. Preliminary results suggest that this study can be extended to examine harmonic motion as a motion between spectral distributions with various degrees of sensory consonance/dissonance, rather than melodic motion between leading tones. [Work supported by UCLA, Graduate Division.]

**Session 5aMUb****Musical Acoustics: Free Reed Instruments I**

James P. Cottingham, Chair

*Physics Department, Coe College, Cedar Rapids, Iowa 52402***Chair's Introduction—10:15*****Invited Papers*****10:20****5aMUb1. 150 years of free reed research: The acoustics of the reed organ and harmonium from Helmholtz to the present.** James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402, jcotting@coe.edu)

Free reed instruments have been known in Asia for thousands of years, but the Western free reed instruments such as the harmonica, accordion, and reed organ have only developed within the last 2 centuries. The birth of Hermann Helmholtz in 1821 coincided rather closely with the early stages of the development of these free reed instruments in Europe and North America, so that by the time of his work in acoustics these instruments were fairly common. A summary of acoustical research on free reed instruments from the time of Helmholtz to the present will be presented, with emphasis on the reed organ and the harmonium. The properties of these and other Western free reed instruments will be contrasted with those of the Asian free reed mouth organs: the sheng, sho, and khaen. Some current and future directions for free reed research will be discussed.

**10:45****5aMUb2. Studies of accordion reed vibrations—Applications in sound synthesis.** René Caussé, Nicolas Misdariis (IRCAM-1, place Igor Stravinsky, 75004 Paris, France), and Denis Ricot (Ecole Centrale Lyon, 69131 Ecully Cedex, France)

The accordion is a free-reed instrument. The two identical reeds associated with one note are driven separately by the airflow created by the inward and outward movements of the bellows. In this present study, attention is focused on understanding the aerodynamic phenomena involved in reed vibration. The acoustical model is deduced from the variation of the flow rate through the slot in the reed plate. A variety of laboratory measurements and observations on different accordion reeds have been made with different blowing pressures. Among them: near-field sound pressure (microphones above and under the reed); displacement and velocity of the reed (capacitive sensor and laser vibrometer system); reed profile, magnified stroboscopic images, flow visualization; etc. From these experimental results, we have extracted the most important phenomena involved, which are to be implemented in a physical model accurate enough to describe the behavior of the reed and to reproduce its characteristic sound. This physical model must be as simple as possible to be computationally efficient and usable in sound synthesis. For example, at IRCAM, it is possible to program the accordion model in Modalys, a synthesizer language based on modal representation formalism. Similarities and differences between simulated and real sounds will be discussed.

**11:10****5aMUb3. Physics, phonetics, and physiology of the diatonic harmonica.** James F. Antaki (Dept. of Surgery and Bioengineering, Univ. of Pittsburgh, 300 Technology Dr., Pittsburgh, PA), Henry T. Bahnson, and Quinter C. Beery (Univ. of Pittsburgh, Pittsburgh, PA)

The harmonica is arguably the most widely played instrument in the world, yet there is a surprising scarcity of published studies of its acoustics or physical dynamics. The typical diatonic harmonica and the kinematics of its free reeds are described. The vibration of the reeds, naturally, when producing a bend, and when speaking as an overblow or overdraw is discussed and investigated by simple stopping of the reeds, by videostroboscopic analysis, and by high-fidelity displacement measurements. The reeds of the ten-hole harmonica can be made to vibrate at varying frequencies depending on the size and morphometry of the players vocal tract. Three different modes of speaking from each hole and its pair of reeds are revealed and studied: first, naturally in a closing mode, either blown or drawn; second, as a bend, either blown or drawn, with pitch in the interval between the two notes in the hole; and third, as an overblow or overdraw in an opening mode with a pitch outside the interval between the two natural notes of the hole. This dynamic interaction allows the player to speak with the instrument perhaps as with no other.

**11:35****5aMUb4. On the perception of free reed organ pipes.** Jonas Braasch (Inst. für Kommunikationsakustik, Ruhr-Univ., D-44780 Bochum, Germany, braasch@ika.ruhr-uni-bochum.de) and Christian Ahrens (Ruhr-Univ., D-44780 Bochum, Germany)

Sounds of three free reed organ stops were measured throughout the whole frequency range of the stops and compared to the measurements of a striking reed stop and a diapason stop. The steady sound spectra and the attack transients were analyzed and compared. As the results show, free reed pipes differ in a number of parameters. Especially the attack transients show characteristic differences in rise time, amplitude, and initial delays of the partial tones from striking reed pipes and diapason pipes. Interestingly, the rise time of the free reed pipes is comparable to those of the measured diapason pipes, although free reed pipes are often cited to have

a sluggish attack transient. A psychoacoustic pair comparison test was conducted to eighteen listeners, using synthesized sounds with mixed acoustical properties of the different organ stops. The results show that the presence of the chuff of the diapason pipes leads to the perception of a shorter attack duration of the diapason pipes compared to the perceptual attack duration of free reed pipes. The importance of the chuff could be stated with, psychoacoustically motivated, rate of change of mag-log amplitude plots. The chuff caused the highest rate of change of the mag-log amplitude that was measured.

### Contributed Poster Papers

Papers 5aMUB5, 5aMUB6, and 5aMUB7 will be presented in poster format. Authors will be at their posters from 12:45 p.m. to 1:45 p.m.

**5aMUB5. Experimental investigation of air-driven free reeds using a laser vibrometer system.** Michael Busha (Phys. Dept., Grinnell College, Grinnell, IA 50112, busha@ac.grin.edu) and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

Previously reported measurements of free-reed vibration using a variable impedance transducer (VIT) gave a picture of the motion of a point on the reed throughout its cycle [J. P. Cottingham, J. Acoust. Soc. Am. **105**, 940 (1999)]. A more detailed picture of the reed motion throughout its vibrational cycle has been obtained by measuring reed velocity using a laser vibrometer system. Unlike the VIT sensor, the laser vibrometer can be scanned along the reed, allowing reed profiles to be obtained. The laser system is also usable on reeds of magnetic material (e.g., steel accordion reeds) as well as reeds too small and inaccessible for the VIT. In this investigation the emphasis has been on comparison of the vibrational motion of accordion reeds with that of the reed organ reeds which had been measured in earlier studies. Some of the differences observed between the two types of reeds can be related to the structure of the reed tongues and frames.

**5aMUB6. The acoustics of the bawu.** Casey A. Fetzer, Nicole L. Brucker, Eric D. Shackelford, and James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402, jcotting@coe.edu)

The bawu employs a metal free-reed mounted in a bamboo resonating pipe with several finger holes. It is closely related in principle of operation to other Asian free-reed instruments, including the sheng of China, the sho of Japan, and the khaen of Thailand and Laos. In these instruments the reed vibration is strongly coupled to the pipe resonance, which is typically close to the natural frequency of the reed. The bawu, in contrast, employs a pipe resonator of variable effective length (using the tone holes) in which the pipe resonance and sounding frequency are both above the

natural reed frequency. The most striking result of this is a significant change in tone quality. Variations in frequency and sound spectrum with blowing pressure have been studied, and some aspects of reed design and construction have been investigated. The relationship between frequency of reed vibration and pipe length has been studied to determine the range of pipe length over which the reed can be made to sound, as well as the amount of frequency shift associated with changes in resonance frequency of the pipe.

**5aMUB7. The history of free reed organ stops.** Christian Ahrens (Musikwissenschaftliches Institut, Ruhr-Univ., D-44780 Bochum, Germany) and Jonas Braasch (Ruhr-Univ., D-44780 Bochum, Germany)

Organ stops based on free reeds have been introduced by the end of the eighteenth century. Although their origin has not been proven, it is most likely that the technique of imported East Asian free reed instruments, probably the Japanese *sho* or the Chinese *sheng*, was adapted. Organ stops with free reeds became soon popular in Europe, especially in Germany and Switzerland and often replaced stops, like the *clarinet* for example, which were traditionally established with beating reeds. With the beginning of the “*Orgelbewegung*” in the 1920s free reed stops were abandoned, because they were not in accordance to the barock organ ideal that was maintained by the protagonists of the “*Orgelbewegung*.” The development of the use of free reed pipes is compared to the development of orchestra music and instruments of the orchestra, especially to the development of the clarinet and the improvement of clarinet playing techniques. Acoustical analyses of organ stops and orchestra instruments—mainly attack transients and Fourier spectra—were done and compared with the historic investigations of Helmholtz and Töpfer/Allihn, for an attempt to explain the evolution of free reed organ stops.

FRIDAY MORNING, 5 NOVEMBER 1999

GARFIELD ROOM, 8:00 TO 10:00 A.M.

### Session 5aPAa

#### Physical Acoustics: Scattering and Elastic Wave Theory

Victor W. Sparrow, Chair

Graduate Program in Acoustics, Pennsylvania State University, 157 Hammond Building,  
University Park, Pennsylvania 16802

#### Contributed Papers

8:00

**5aPAa1. Theory of perfectly matched layer for elastic waves and their applications in cylindrical and spherical coordinates.** Q. H. Liu (Dept. of Elec. Eng., Duke Univ., Durham, NC 27708-0291)

Recently, a general theory of perfectly matched layer (PML) is developed for elastic waves in anisotropic media and is applied to waves in cylindrical and spherical coordinates through an improved scheme of complex coordinates [Liu, J. Acoust. Soc. Am. **105**, 2075–2084 (1999)]. As is

known for electromagnetic waves, Berenger’s original PML scheme does not apply to cylindrical and spherical coordinates. The straightforward extension of the complex coordinates for elastic waves to cylindrical and spherical coordinates requires extra unknowns for time-domain solutions, wasting computer memory and computation time. The main idea of the improved scheme in this work is the use of integrated complex variables. It is shown that for three-dimensional cylindrical and spherical coordinates, this improved PML scheme requires no more unknowns than in Cartesian coordinates. The number of unknowns can be further reduced



through the use of symmetry in the partial differential equations. The PML scheme allows an arbitrary inhomogeneity in the medium, and is suitable for numerical solutions of wave equations by finite-difference, finite-element, and pseudospectral methods for elastic waves in inhomogeneous media with cylindrical and spherical structures. Finite-difference time-domain (FDTD) results are shown to demonstrate the efficacy of the PML absorbing boundary condition.

8:15

**5aPAa2. Harmonic generation by Rayleigh–Lamb waves in isotropic elastic plates.** W. J. N. de Lima and M. F. Hamilton (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712-1063)

Harmonic generation by Rayleigh–Lamb waves of finite amplitude in homogeneous, isotropic, stress-free elastic plates is investigated theoretically. A bifrequency primary wave field is considered, in which the two waves propagate in single yet arbitrary Rayleigh–Lamb modes. Solutions for the second-harmonic and difference-frequency components are obtained via modal decomposition and use of reciprocity relations. Two conditions are required for generation and resonant amplification of these spectral components, power flow, and phase matching. Second-harmonic generation is considered first. Although phase matching is possible between a primary wave in the first symmetric mode and its second harmonic in the second asymmetric mode, the power flow is found to be zero. However, with one primary wave in each of the first symmetric and asymmetric modes, both phase matching and power flow may be achieved with the difference-frequency wave generated in the first asymmetric mode. In particular, resonant parametric amplification of the difference-frequency wave can be achieved under conditions where dispersion prevents efficient coupling from occurring between the primary waves and other frequency components. Explicit analytic solutions are presented and discussed. [Work supported by the Brazilian Ministry of Science and Technology, and the National Science Foundation.]

8:30

**5aPAa3. Acoustic diffraction by a half plane in a viscous medium.** Allan D. Pierce, Raymond J. Nagem, and Mario Zampolli (College of Eng., Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

A formal solution for the diffraction of sound by a half plane in a viscous medium is given and the special case of plane wave diffraction is implemented and compared to previous results. The acoustic field can be represented as a superposition of acoustic and vorticity modal fields. The two fields satisfy Helmholtz type wave equations and coupled boundary conditions on the surface of the scatterer, requiring that the fluid velocity be identically zero on the surface of the scatterer. The equations are solved by means of the Wiener–Hopf method and integral expressions for the scattered fields are given. The solutions near the edge for the particular case of an incident plane wave are expressed in terms of Fresnel integrals of complex argument. From the solutions it follows that in the vicinity of the edge there exists a boundary layer in which the fluid velocity field has no singularity. The modal components of the velocity field have square root singularities at the edge which cancel out when the two modes add to give the total fluid velocity. Outside the boundary layer the acoustic mode dominates and the solution coincides with the classical inviscid solution.

8:45

**5aPAa4. Modeling scattering from objects using a Kirchoff/diffraction method.** John A. Fawcett (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

In this paper a facet-model based upon the Kirchoff approximation is derived for high-frequency scattering from rigid objects. In the case of objects with edges, the standard approximation is modified [M. Tran Van Nhieu, *J. Acoust. Soc. Am.* **99**, 79–87 (1996)]. The resulting facet model is used to compute the backscattered time series from spheres and flat-ended cylinders, both smooth and with significant deterministic roughness. Analytical solutions and three-dimensional (azimuthally symmetric)

boundary integral equation method solutions are used to benchmark the facet-model time series. Very good agreement between the facet model and the exact solutions are found, even for situations where the validity of the Kirchoff approximation is questionable.

9:00

**5aPAa5. A modal analysis of the interaction of acoustical signals from submerged fluid-filled elastic shells.** M. F. Werby (NRL Code 7181, Stennis Space Center, MS 39529)

To most easily identify vibrational components in target-sound scatter, an acoustical background is desirable. Here a new background for fluid-filled elastic structures is subtracted in a coherent manner which leads to a very interesting picture of what happens in such interactions. What is shown is that the results of such scatter yield the coherent effect of the elastic vibrations, the rescatter from within the shell containing the water, but surprisingly there is a proliferation of water-borne waves distributed among all modes as demonstrated in this analysis of the residual modal terms. These “water-borne” modes are well defined, fairly narrow, and evenly displaced. A time domain picture ought to be of interest in further analysis. [Work supported by the Office of Naval Research and Naval Research Laboratory.]

9:15

**5aPAa6. Method of superposition applied to axisymmetric structures.** Angie Sarkissian (Naval Res. Lab., Washington, DC 20375-5350, angie@aquanrl.navy.mil)

The method of superposition is applied to axisymmetric structures to solve for the scattered field produced by a general, non-axisymmetric incident field. The scattered field is written as a sum over the fields produced by rings of source distributions placed on a surface inside the scatterer. The source distribution on each ring is next expanded in circumferential modes. Similarly, the field produced by the source distribution is expanded in circumferential modes. The source distribution is solved for by requiring the scattered field to satisfy the proper boundary conditions on the structure surface. This reduces the computational load of the method previously applied of placing point sources on a surface inside the structure since it reduces the size of the matrices to be inverted, thus extending the applicability of the method to higher frequencies. Results are shown for a finite cylindrical structure with hemispherical end-caps satisfying soft boundary conditions. [Work supported by ONR.]

9:30

**5aPAa7. Spectral properties of Christoffel equation for Rayleigh waves.** Sergey Kuznetsov and Alexander Kaptsov (Inst. for Problems in Mech., Prospect Vernadskogo, 101a, Moscow, 117526 Russia)

Spectral properties of the matrix Christoffel equation are needed for construction of the exponentially decaying surface (Rayleigh) waves. The following analysis covers theorems on root characterization, the structure of eigenvalues, and eigenspaces of the Christoffel equation. The main results are obtained by three-dimensional complex formalism, which goes back to Rayleigh. For the case of anisotropic media this approach was exploited by Farnell in numerical analysis of Rayleigh waves propagating in anisotropic crystals and in analytical study by Stoneley and later by Dieulesaint and Royer. Sextic formalism for Rayleigh wave analysis was proposed by Stroh. This is based on the similarity of solutions for line dislocations in unbounded medium and the propagation of Rayleigh waves. This approach was developed by Barnett and Lothe and later by Chadwick *et al.*, so theorems of uniqueness and existence for Rayleigh waves were proved. Substitution of a complex root in Christoffel’s equation produces an equation for the eigenvector (amplitude vector) determination. Propositions which flow out directly from the analysis of the structure of this equation describe spectral properties and characterization of eigenspaces. One of the most interesting results asserts that algebraic multiplicity of the phase speed equals its geometric multiplicity. [Work was supported by INTAS 96-2003.]

**5aPAa8. Analytical/numerical description of sound field in irregular waveguides.** Victor T. Grinchenko (Inst. of Hydromechanics NAS of Ukraine, 8/4 Zhelabov st., Kiev 252057, Ukraine, vin-igm@gu.kiev.ua)

The problems of the harmonic wave propagation in waveguides with varying cross sections and sound radiation from open ends of ones are considered. The main attention gives to development method of complete analytical description of the sound fields in both cases. This gives a basis to get accurate quantitative and qualitative description of the fields near the junctions of the waveguide cross sections and open ends. Numerical implementation of the general analytical formulas requires special proce-

dures to be adequate the local singularities in vicinities of corner points. The approach is based on the introduction of the idea of the general solution of the acoustical boundary problem. The idea gives a way to construct analytical solutions for the problems under consideration and can be extended to a broad spectrum of radiation and scattering problems. The well known solutions for regular waveguides in the form of normal waves are used as important part of the general solutions. Numerical data for concrete waveguides demonstrate the accuracy and convergence rate of the method. Analysis of the data shows interesting physical properties of propagating waves, flows of energy, and local structure of sound field at the juncture of waveguide parts.

FRIDAY MORNING, 5 NOVEMBER 1999

GARFIELD ROOM, 10:30 A.M. TO 12:30 P.M.

### Session 5aPAb

## Physical Acoustics: Sonoluminescence, Bubbles and Drops

Robin O. Cleveland, Chair

*Aerospace and Mechanical Engineering, Boston University, 110 Cummings Street, Boston, Massachusetts 02215*

### Contributed Papers

10:30

**5aPAb1. Using phase space diagrams to interpret multiple frequency drive sonoluminescence.** Jeffrey A. Ketterling and Robert E. Apfel (Dept. of Mech. Eng., Yale Univ., New Haven, CT 06520)

The recent experimental results of J. Holzfuss, M. Rüggeberg, and R. Mettin [Phys. Rev. Lett. **81**, 1961 (1998)] in which a second harmonic drive system was used to generate sonoluminescence (SL) have been analyzed in the context of the dissociation hypothesis (DH) of D. Lohse and S. Hilgenfeldt [J. Chem. Phys. **107**, 6986 (1997)]. The second harmonic introduces two more variables that are under experimental control: a phase and an additional pressure term to the acoustic drive pressure. Diffusive equilibrium curves for a fixed gas concentration were calculated as was the Mach criterion. A phase space diagram was constructed to permit the prediction of regions of stable SL, unstable SL, stable non-SL, and unstable non-SL. These were compared to Holzfuss' experimental observations, and excellent quantitative agreement was seen. The results provide further evidence that the underlying assumptions of DH are sound. They also indicate the utility of DH for determining appropriate experimental conditions to achieve SL and for optimizing an experimental system.

10:45

**5aPAb2. All good oscillations must come to an end.** William C. Moss,<sup>a)</sup> Joanne L. Levatin<sup>a)</sup> (Lawrence Livermore Natl. Lab., L-200, 7000 East Ave., Livermore, CA 94550), and Andrew Szeri<sup>b)</sup> (Univ. of California Berkeley, Berkeley, CA 92697)

The Rayleigh–Plesset equation (RPE) describes the growth, collapse, and subsequent rebounds of a sonoluminescing bubble, as long as unphysical values of the surface tension and viscosity are used. If realistic values of the surface tension and viscosity are used, the RPE produces insufficiently damped bubble dynamics, which disagree with experimental data. Every derivation of the RPE assumes that the compressibility of the liquid gives rise to the damping. We show that the compressibility of the liquid is at most a second-order effect. The damping comes from the most compressible component of the system, which is the gas, not the liquid. This simple physical reasoning generates a new RPE that reproduces the extreme damping that is observed experimentally, using tabulated values of surface tension and viscosity; a result that is unobtainable with all previous versions of the RPE. Our results affect the calculated limits of

stability of sonoluminescing bubbles as well as the calculated magnitude of acoustic emissions from collapsing bubbles.<sup>a)</sup>This work was performed under the auspices of the U.S. Department of Energy under Contract No. W-7405-Eng-48. <sup>b)</sup>Supported by NSF.

11:00

**5aPAb3. The impact of mass diffusion on SBSL dynamics in a variable-gravity environment.** Sean C. Wyatt, Charles R. Thomas, Jason C. Raymond, Ronald A. Roy, and R. Glynn Holt (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummings St., Boston, MA 02215)

Recent experiments performed on a KC-135 parabolic flight aircraft [Matula *et al.*, J. Acoust. Soc. Am. **102**, 3185 (1997); Young *et al.*, *ibid.* **103**, 3077 (1998)] indicate that the optical emission intensity of SBSL (air bubble in water) increases with decreasing gravitational acceleration. Wyatt *et al.* [J. Acoust. Soc. Am. **104**, 1770 (1998); **105**, 960 (1999)] argue that this effect is primarily due to small changes in the equilibrium radius for diffusively stable pulsations. This concept is further explored in a series of variable-acceleration experiments employing a water-filled acoustic resonator in which individual air bubbles are levitated. The critical role of mass diffusion is assessed by varying the dissolved gas concentration. Results are presented and compared with predictions from numerical simulations. [Work supported by NASA.]

11:15

**5aPAb4. Dynamics of laser-induced cavitation bubbles.** Iskander Akhatov, Nailya Vakhitova, Kamil Zakirov (Inst. of Mech., Ufa Branch of Russian Acad. of Sci., 6 K. Marx Str., Ufa, 450000 Russia), Robert Mettin, Olger Lindau, and Werner Lauterborn (Universität Göttingen, Bürgerstraße 42-44, D-37073, Göttingen, Germany)

Bubbles in liquids can be formed in a variety of ways. A convenient method to produce a single bubble in a liquid is to focus a short pulse of laser light into the liquid. A bright light-emitting plasma of obviously high pressure expands forming a gas-vapor bubble. After some time bubble expansion stops at some maximum radius. From thereon the bubble starts to shrink driven by the ambient pressure and leading to a violent collapse possibly followed by rebounds. A mathematical model that describes the

evolution of such a laser-induced cavitation bubble is presented. The model takes into account heat conduction of the liquid and gas phases, evaporation and condensation processes of the liquid phase, gas diffusion from the liquid into the bubble, and compressibility of the liquid. A reasonable comparison of theoretical results with experimental data is obtained. [Work partially supported by Internationales Büro des BMBF (Contract RUS-133-1997) and European Commission (INCO-Copernicus, Contract ERBIC15CT980141).]

11:30

**5aPAb5. Measurement of the complex impedance of a bubbly fluid in the vicinity of the bubble resonance frequency.** Preston S. Wilson, Ronald A. Roy, and William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

The acoustic impedance of bubbly fluids has been investigated and results for mixtures above and below resonance are common in the literature [e.g., A. B. Wood, *A Textbook of Sound* (Macmillan, New York, 1941)]. Because of the high attenuation encountered at frequencies near individual bubble resonance, traditional standing-wave and pulse propagation methods fail and little information has appeared in the literature for this regime. A new technique was developed to overcome this difficulty. Laboratory measurements of the complex plane-wave impedance of mixtures containing monodispersed air bubbles and water have been obtained for frequencies spanning the individual bubble resonance frequency. The experiments were conducted in a thick-walled water-filled tube, piston driven at one end and terminated with a bubbly fluid at the other end. This arrangement allows for operation near the bubble resonance frequency, since high attenuation is not encountered inside the waveguide itself. In addition, the standard transfer function method [J. Y. Chung and D. A. Blaser, *J. Acoust. Soc. Am.* **68**, 907–921 (1980)] can be utilized, which results in broadband complex-valued impedance measurements. Results over a range of frequencies (low kHz) and void fractions with known size distributions will be presented and compared to existing theory. [Work supported by ONR.]

11:45

**5aPAb6. Oscillation of bubbles released from needles in natural and artificial sediments.** Doru Velea, Craig Hickey, James M. Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Coliseum Dr., University, MS 38677), and Ali Kolaini (Univ. of California, Santa Barbara, Santa Barbara, CA 93106)

The acoustical characteristics of bubbles differ significantly in sediments as compared to those in sediment-free water. To examine this effect, an experiment has been designed by releasing bubbles from a needle/hydrophone assembly inside saturated natural and artificial sediments. Observations show that the resonance frequency of these bubbles released from needles in the sediment decreases and their total damping coefficient increases significantly compared to bubbles released from needles in

sediment-free water. For example, a bubble with a radius of 2.5 mm in water has a resonance frequency of approximately 1.3 kHz and a quality factor of approximately 30. These parameters reduced to approximately 800 Hz and 2 for a similar bubble in a natural sediment with a mean particle size of 30  $\mu\text{m}$ . Theoretically the total damping coefficient of an oscillating bubble in water consists of the thermal, acoustic and viscous contributions. Another dissipation factor, the Darcy damping, is found to be dominant in acoustic emissions of bubbles in sediments. This term accounts for the interaction forces between the liquid and the porous skeleton. The theoretical predictions of the total damping coefficient of bubbles in sediments (in the rigid matrix approximation) are discussed and the results are compared with experimental observations.

12:00

**5aPAb7. Modeling the response of acoustically levitated foam drops.** R. Glynn Holt, J. Gregory McDaniel (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215), and Iskander Sh. Akhatov (Inst. of Mech., Ufa Branch of Russian Acad. of Sci., Ufa, 450000 Russia)

Small samples of aqueous foam ("foam drops") can be acoustically levitated and manipulated for purposes of investigating their rheological properties [Holt *et al.*, *J. Acoust. Soc. Am.* **105**, 1146 (1999)]. We report here advances in the effort to model the response of foam drops to static and time-varying modulation of the acoustic field to extract rheological properties: the yield stress, the bulk viscoelastic shear and dilatational moduli (complex), and the effective bulk shear and dilatational viscosity numbers. We have begun to develop a bubble-based model of foam mechanics which incorporates the effects of the acoustic field and individual bubble oscillations. Such a model represents a significant advance over traditional models of foam rheology. [Work supported by NASA and NRC.]

12:15

**5aPAb8. Theoretical analysis of free-shape oscillations of surfactant-bearing drops.** Xiaohui Chen and Robert E. Apfel (Dept. of Mech. Eng., Yale Univ., 9 Hillhouse Ave., New Haven, CT 06520-8286)

A theoretical analysis is presented for the free-shape oscillations of surfactant-bearing drops suspended in air using the method of spherical harmonic expansions. Two extreme types of surfactant mass transfer between the bulk and the surface of drops are considered: (1) no surfactant transport as the drop oscillates; and (2) the surfactant transport is so fast that the surface surfactant concentration remains constant and homogeneous. It is found that the shear viscosity of bulk phase, surface elasticity, surface shear viscosity and surface dilatational viscosity have the same order of influence on the free-damping constant when the surface shear viscosity is smaller than a certain value. Otherwise, the free-damping constant is determined by the surface shear viscosity, and the influence of the surface dilatational viscosity, the surface elasticity, and the shear viscosity of the bulk phase on the drop oscillation is negligible.

## Session 5aSA

## Structural Acoustics and Vibration: Vibration and Radiation—Complex Structures

Rudolph Martinez, Chair

Cambridge Acoustical Associates, 84 Sherman Street, Cambridge, Massachusetts 02140

## Contributed Papers

8:00

**5aSA1. Comparing the vibrational behavior of wood and aluminum baseball bats.** Daniel A. Russell (Sci. and Math Dept., Kettering Univ., Flint, MI 48504, drussell@kettering.edu)

Experimental modal analysis was performed on a variety of wood and aluminum baseball bats to study the differences and similarities in vibrational behavior. Wood bats were found to exhibit split pairs of vibrational modes whose frequencies depend on whether the plane of vibration is along or across the grain. The frequency separation for these duplicate modes varies considerably from bat to bat and appears to be dependent on the quality of wood. Data will be presented for bats made from ash, maple, an ash-hickory-maple combination, and 6-ply laminated ash. Aluminum bats exhibit a single set of bending modes, with similar shapes but wider frequency spacing than wood bats. However, aluminum bats also vibrate with several cylindrical bell-type modes, the lowest of which is found to be responsible for the loud "ping" produced when a metal bat strikes a ball. Aluminum bats also allow greater modifications for improved bat performance. An example of an aluminum bat with a dynamic absorber will be compared to other metal and wood bats.

8:15

**5aSA2. Sports acoustics: A simple acoustic method for player performance evaluation during volleyball games.** David G. Browning (Dept. of Phys., East Hall, Univ. of Rhode Island, Kingston, RI 02881)

When struck, a volleyball resonates resulting in an acoustic signal with a frequency of approximately 200 Hz [D. G. Browning *et al.*, *J. Acoust. Soc. Am.* **97**, 3399 (1995)]. By modifying a small, battery-powered, handheld, sound level meter to receive only in the 125–250 Hz band (now called a SMACKMETER), this sound could easily be monitored from the stands or sidelines. No significant crowd or other noise was encountered in this frequency band during games. It was found that on the university level of womens' competition, signals varied approximately 10 dB from a soft serve to a hard spike; for high school games the range was less in most cases. It appears to offer a coach or fan a small, economical device to evaluate one aspect of player performance, specifically how hard they strike the ball, during games and warm-ups.

8:30

**5aSA3. Vibro-acoustic effects of a soft compliant layer on submerged finite cylindrical shells.** Peter C. Herdic, Brian H. Houston, Douglas M. Photiadis, and Earl G. Williams (Naval Res. Lab, Code 7130, Washington, DC 20375)

The structural acoustics associated with a compliant material applied to the outer surface of a framed fluid-loaded cylinder is examined. Spatially dense laser Doppler vibrometer (LDV) and near field acoustic holography (NAH) measurements are employed to capture the detailed response under radial point force excitation. In the mid- and high-frequency regimes, the soft, low impedance layer is associated with significant reduction in the normal velocity at all wavenumbers. Essentially, there is very little motion of the outer compliant layer due to the impedance mis-

match with the fluid. Further, these data indicate that Bloch wave components are modified somewhat due to reduced fluid-loading effects. However, at low frequencies the normal velocities at the fluid interface are greatly accentuated due to simple resonances associated with the impedance of the acoustic medium, the soft spring-like compliant layer, and the structure. Although these resonances are typically characterized as subsonic, enhanced levels are observed in the farfield. Also, it is found experimentally that the input drive impedance changes with the addition of the compliant layer, thus, modifying the energy delivered to the system by the shaker. The data will be reported along with simple impedance models that illustrate the physics of the behavior.

8:45

**5aSA4. Circumferential waves on an immersed, fluid filled elastic cylindrical shell.** X. L. Bao, P. K. Raju (Ross 201, ME Dept., Auburn Univ., Auburn, AL 36849, pkraju@eng.auburn.edu), and H. Uberall (Catholic Univ. of America, Washington, DC 20064)

The existence of various types of circumferential waves, both predominantly shell or fluid borne, and the repulsion of their dispersion curves is discussed here for an infinite, thin elastic, circular-cylindrical shell immersed in a fluid and filled with another fluid. The study is based on analytic calculation of the partial-wave resonances in the acoustic scattering amplitude of a normally incident plane wave. A large number of cases of repulsion are found in the phase-velocity dispersion curves of the various types of circumferential waves due to the shell-fluid coupling.

9:00

**5aSA5. A highly efficient system for wave number frequency analysis using an sly in an underwater housing.** Walter H. Boober (Naval Undersea Warfare Ctr., Code 8211, Newport, RI 02841) and Phillip Abbott (Ocean Acoust. Services and Instrumentation Systems, Inc., Lexington, MA 02173)

The development of a water tight housing with an optical quality acrylic window allows for greater range of adaptation to imposed constraints for vibration analysis underwater. Most systems for such analysis utilize tanks with windows, or fiberoptic probes requiring carefully controlled rigging and positioning methods. Use of the underwater housing gains a range of freedom from many of the barriers which can inhibit or seriously limit the aforementioned methods for underwater vibration analysis. Utilizing this unique system it was possible to instrument, rig, position, test, and do an onsite cursory evaluation of wave number frequency on a 4.0 m long  $\times$  533 m diameter cylinder. A 4.5-kg maximum force shaker was installed axially on an internal bulkhead. Approximately 2.43 m of the shell was free-flooded and 1.26 m filled with lime. The shaker was situated in the cavity separating the two sections. The cylinder was suspended vertically from a rotator and centered in 10 m of fresh water. The sly was positioned at approximately 9 m from the shell allowing for a 3.3-m axial scan of the shell.



9:15

**5aSA6. Acoustic signature measurements of a small submersible.** Joseph M. Cuschieri and David Vendittis (Ctr. for Acoust. and Vib., Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, FL 33431)

In an effort to characterize the acoustic signature of a small autonomous submersible, the FAU Ocean Explorer (OEX) class AUV (autonomous underwater vehicle), measurements of the acoustic source level were performed in a reverberant tank, with the submersible held stationary, with and without the propeller. For comparison, source level measurements were also performed off-shore at the South Florida Test Facility (NSWC-CD), in Ft. Lauderdale, FL and at the Acoustic Research Detachment (NSWC-CD), Lake Pend Oreille in Bayview, ID. The results of these measurements show that because of the relatively slow rotational speed of the propeller, the acoustic signature is low, although distinct features in the signature are discrete tones, which seem to be coming from the propeller. In fact measurements of the acoustic signature under power were difficult to obtain even in low background noise environments. For this reason, the reverberant tank measurements present a very effective approach provided broadband information is sufficient and directivity information is not necessary. Results from these three sets of measurements will be presented, together with an analysis of what are potentially the significant sources contributing to the acoustic signature. [Work supported by ONR.]

9:30–9:45 Break

9:45

**5aSA7. Generator of fluid reactances and resistances for blades vibrating at high frequencies.** R. Martinez and K. Martini (Cambridge Acoustical Assoc., 84 Sherman St., 3rd Level, Cambridge, MA 02140)

Blades and two-sided bladelike objects vibrating at high frequencies experience a spatial distribution of radiation loading with the following two features: (1) away from the edges: a primarily resistive recoil pressure, oscillating about a  $2\rho c$  mean (that mean being the familiar “1-D” solution); and (2) near and at the edges: a complicated mixture of resistive and reactive fields that account for diffraction effects. This paper discusses a novel analytical/numerical tool for predicting (1) and (2) inexpensively and in a unified manner. The theory achieves that by considering an unconventional product of Fresnel functions, each representing an edge. It succeeds in generating a distribution of finite-fluid impedances even for sonic patterns of vibration, i.e., even for prescribed wave numbers at “coincidence,” for which an additive correction scheme using the same Fresnel integrals would fail. The new computational tool is unexpectedly accurate all the way down to the acoustically compact range, failing to recover by only 27% the well-known entrained mass of a rigidly vibrating strip at low frequencies. The presentation will close with a brief discussion of how the presence of a mean flow affects the above conclusions. [Work supported by NSWC/CD’s 6.2 program.]

10:00

**5aSA8. Visualization of acoustic pressure fields radiated from a complex vibrating structure.** Nassif Rayess and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

This paper presents visualization of acoustic pressures radiated from a complex vibrating structure using the Helmholtz equation least-squares (HELs) method [Z. Wang and S. F. Wu, *J. Acoust. Soc. Am.* **102**, 2020–2032 (1997); **104**, 2054–2060, (1998)]. The structure under consideration has the same shape and dimension as those of a full-size vehicle front end. Random noises are generated by an engine noise simulator that consists of many loudspeakers of different sizes. The radiated acoustic pressures are measured a certain distance away from the vehicle surface, which are then taken as the input to a computer model based on HELs formulations.

Experiments are conducted in a 12 ft by 12 ft by 6.5 ft fully anechoic chamber at the Acoustics, Vibration, and Noise Control Laboratory of Wayne State University. The reconstructed acoustic pressures on vehicle surfaces and in the field are compared with the measured data. Also shown are comparisons of the reconstructed and measured acoustic pressure spectra at various locations on the vehicle back surface. Results show that good agreement can be obtained with relatively few expansion terms over the low- to midfrequency range. The accuracy of reconstruction, however, decreases with the increase of frequency as expected. [Work supported by NSF.]

10:15

**5aSA9. Power variances in a reverberant elastic system.** Richard L. Weaver, Oleg Lobkis, and Igor Rozhkov (Theoret. and Appl. Mech., Univ. of Illinois, 104 S. Wright St., Urbana, IL 61801)

Frequency-dependent fluctuations in the power transmissibility of a reverberant elastic body are measured and compared with theoretical estimates. Wideband (0.1–2 MHz) transient sources are applied to aluminum blocks of dimensions of the order of 3 cm (greater than typical wavelengths) and found to generate signals with durations of the order of 100 ms. At low frequencies the power spectra of these waveforms, found by FFT, show the usual distinct resonance peaks that allow unambiguous mode counting. At higher frequencies, where modal overlap is significant, mode counting is more difficult. Nevertheless, modal density is, in principle, recoverable by analysis of the variance of the power spectra. An improved theory for the relative variance is presented, which includes the effects of variations in modal widths and includes the effects of spectral rigidity. The theory assumes that modal amplitudes are Gaussian random numbers, as is predicted by random matrix theory. The method consistently overestimates modal densities, by about 50%. After elimination of other causes, the discrepancy is attributed to the assumption on modal amplitude statistics.

10:30

**5aSA10. Numerical comparison of acoustic infinite elements.** Jeffrey Cipolla (HKS, Inc., 1080 Main St., Pawtucket, RI 02860)

Numerical studies of various infinite element formulations are conducted: conjugated types, unconjugated, decay (unconjugated), and DAA1 elements of the Olsen/Bathe type. In addition, an “element” implementing the Bayliss *et al.* first-order condition is studied. All elements except the decay are implemented in separable coordinates. The element formulations are studied analytically, using axisymmetric modal impedances of a sphere. These studies reveal limitations, advantages, and appropriate methods of usage for the various formulations. Subsequently, the various elements are implemented in production-type software, and numerical comparisons are conducted. These studies reveal the relative accuracy of the formulations as well, but also permit discussion of numerical performance. Fully coupled, three-dimensional structural acoustic meshes are studied. Comparisons are made against a BEM method, and analytic solutions where appropriate. When used properly, infinite element methods can deliver accuracy comparable to the BEM for three-dimensional acoustic problems, at considerably reduced cost. However, this study reveals the superiority of some infinite element formulations.

10:45

**5aSA11. Acoustic propagation in strong surface ocean ducts. Scattering from elastic shells that excite the lowest-order symmetric Lamb mode and amplitude-modulated signals.** M. F. Werby (NRL Code 7181, Stennis Space Center, MS 39529)

The concept of group velocity is routinely used in terms of a mathematical construct but frequently confused with the actual physical process of energy transport. The origin comes either from using the stationary phase method in constructing the time domain from the frequency domain, from calculating sound signals at differing frequencies and identifying the

arrivals with an envelop function, or by using wave-packet theory (usually in quantum mechanics) with something like Gaussian pulses centered about some carrier frequency. In many cases this definition of the derivative of angular frequency with respect to wave number is adequate and does characterize the arrival times of energy packets. There are cases for which such a definition is meaningless and occurs when the stationary phase method to first order breaks down. Usually when this happens the related pulse signal is amplitude modulated. Two examples are demonstrated. One has to do with wave propagation in which there is strong ducting at the ocean surface. The other occurs when scattering from an elastic shell, the first symmetric Lamb mode, is excited. In each case the results are atypical, surprising, and delightful. [Work sponsored by the Office of Naval Research and the Naval Research Laboratory.]

11:00

**5aSA12. Vibroacoustical preventive method for non-destructive testing and evaluation of rotative bladed systems.** Leonid M. Gelman and Lyudmila N. Udovenko (Dept. of Nondestructive Testing, Natl. Tech. Univ. of Ukraine, 37, Peremogy pr., Kiev, 252056 Ukraine)

Method uses spectral component parameters of vibroacoustical noise of systems as features. Estimation of mentioned parameters allows evaluation of the levels of mechanical and aerodynamic loadings of rotative bladed systems. Experimental results with aircraft gas turbine engine are presented.

FRIDAY MORNING, 5 NOVEMBER 1999

KNOX ROOM, 8:30 A.M. TO 12:35 P.M.

## Session 5aSC

### Speech Communication: From Acoustics to Words

Mark A. Pitt, Cochair

*Department of Psychology, The Ohio State University, 1885 Neil Avenue Mall, Columbus, Ohio 43210*

Keith A. Johnson, Cochair

*Department of Linguistics, The Ohio State University, 1712 Neil Avenue, Columbus, Ohio 43210*

**Chair's Introduction—8:30**

#### *Invited Papers*

8:35

**5aSC1. The size and generality of units in speech perception.** Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada)

In English, the perception of syllables can be largely predicted from the perception of phonemes. Experiments reviewed by Allen [IEEE Trans. Speech and Aud. Proc. 2, 567–577] have consistently shown that the correct identification of nonsense CVC syllables in noise is extremely well predicted from the marginal identification rates of their constituent phonemes. Simulations suggest that this result can be easily achieved when syllable patterns can be 'factored' into phoneme parts. Representations allowing even slightly idiosyncratic relationships between stimuli and syllables apparently cannot produce such results. Evidence for phoneme factorability is also provided by the results of certain parametric speech perception experiments. In such experiments, listeners hear synthetic syllables that span two or more categories of two or more segments (e.g., "bad," "bet," "bat," "bet"). Analysis of several such experiments shows that syllable identification can be accurately modeled using independent phoneme-sized units that are sensitive to stimulus properties, plus certain highly constrained influences of higher level symbolic elements (e.g., words). These higher level influences are limited to bias effects that do not interact with stimulus properties. These results suggest that small, highly general elements play the key role in the transduction of signals to symbol in speech. [Work supported by SSHRC.]

8:55

**5aSC2. The prelexical code hypothesis.** Emmanuel Dupoux and Christophe Pallier (Laboratoire de Sci. Cognit. et Psycholinguistique EHESS/CNRS, 54 boulevard Raspail, Paris, 75006 France, dupoux@iscp.ehess.fr)

The prelexical code hypothesis states that there exists a level intermediate between acoustic/phonetic representation and lexical entries. This level encodes syllable-sized structural aspects of the speech stream in a language-specific fashion. We review evidence in favor of this hypothesis in recent studies using phoneme and fragment detection, adaptation to distorted speech, discrimination, ERPs, long-term priming, as well as studies of speech perception in the young infant. We present evidence that such a prelexical code is compiled during the first years of life and remains unaffected by later linguistic experience. We then discuss the competing hypotheses as to the detailed format of the prelexical code (syllables, half-syllables, autosegmental representations, etc.).

9:15–9:40

Commentary by Peter Ladefoged

9:40

**5aSC3. The merge model: Speech perception is bottom-up.** Dennis Norris (MRC Cognition and Brain Sci. Unit, 15 Chaucer Rd., Cambridge, Cambs, UK)

Does feedback from the lexicon influence the processing of phonemic or phonetic information in normal speech perception? Arguments are presented that top-down feedback from the lexicon is no value in spoken word recognition. In fact, top-down feedback is as likely to be harmful as helpful. Consistent with this analysis, recent studies of phoneme identification and phonetic categorisation have generated data that are incompatible with interactive models like TRACE [McClelland and Elman (1986)] which employ top-down feedback. However, other data rule out bottom-up models like the Race model of Cutler and Norris (1979). A new, bottom-up model is presented, the Merge model. Merge can simulate these problematic results by merging information from both lexical and sublexical analyses while maintaining a strictly feed-forward data path for spoken word recognition. Merge builds on the principles of lexical competition and bottom-up processing incorporated in the Shortlist model [Norris (1994)].

10:00

**5aSC4. The data are what the data are: Top-down processing in spoken word recognition is necessary.** Arthur G. Samuel (Dept. of Psych., SUNY Stony Brook, Stony Brook, NY 11794, asamuel@psych1.psy.sunysb.edu)

For the last two decades there has been an ongoing debate about the architecture of the word recognition process: Is this process accomplished in a strictly autonomous, bottom-up fashion, or does information at higher levels affect the perception of lower-level representations? I will review results from a number of experimental settings that support the necessity of positing top-down, lexical-to-phonemic effects. I will argue that models that are presented as strictly autonomous must therefore either be at odds with the data, or must not in fact be fully autonomous. The literature indicates that perception of spoken words is (naturally) based primarily on the perceptual input, but that lexical knowledge also plays a significant role: Top-down processing is a necessary component of an accurate model.

10:20–10:45

Commentary by James Sawusch

10:45–11:00 Break

11:00

**5aSC5. Perceptual adaptation to talker-specific characteristics of speech.** Lynne C. Nygaard (Dept. of Psych., 532 N. Kilgo Circle, Emory Univ., Atlanta, GA 30322, lnygaar@emory.edu)

Recent research suggests that familiarity with a talker's voice increases the intelligibility of their speech. Listeners who have learned the specific vocal characteristics of a talker are better able to perceive the linguistic content of that talker's speech than listeners who have not. Studies will be presented that used a voice learning paradigm to investigate the nature of this perceptual adaptation to talker's voice. In each study, listeners were familiarized with a set of talker's voices and then tested on their ability to extract the linguistic content of speech. Two general questions were addressed. First, what type of learning is necessary to become familiar with a talker's voice? Second, what kind of knowledge is acquired when listeners learn to recognize voices? In general, the results suggest that perceptual adaptation to a talker's voice is extremely detailed and context-dependent. These findings suggest that any proposed mechanism of perceptual compensation in speech must include long-term memory for specific talker characteristics.

11:20

**5aSC6. How infants' word segmentation abilities are affected by talker variability.** Peter W. Jusczyk (Depts. of Psych. and Cognit. Sci., Ames Hall, Johns Hopkins Univ., Baltimore, MD 21218, jusczyk@jhu.edu) and Derek M. Houston (Johns Hopkins Univ., Baltimore, MD)

To begin building a lexicon, infants must extract, encode and store the sound patterns of words from fluent speech. Recent investigations suggest that infants begin this process at about 7.5 months of age [Jusczyk and Aslin (1995)]. Of course, to be useful in word recognition, lexical representations must endure in memory and generalize to tokens produced by different talkers. Investigations of infants word segmentation abilities have generally focused on items produced by a single talker. We have begun to investigate the extent to which infants are able to generalize their representations of words across different talkers. Our findings suggest that the ability of 7.5-month-olds to generalize across talkers are initially limited, but improve considerably over the next three months. Moreover, in contrast to the widely held view that representations of the sound structure of words are abstract prototypes, our results suggest that infants may encode information about talker characteristics into their representations. We discuss the development of these capacities, the conditions that promote generalizability across talkers, and the implications of these findings for understanding the growth of the lexicon. [Work supported by NICHD, NIDCD, and NIMH.]

Commentary by Richard Pastore

*Contributed Papers*

11:55

**5aSC7. Lexical activation produces impotent phonemic percepts.**

James M. McQueen, Anne Cutler (Max-Planck-Inst. for Psycholinguist., Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, James.McQueen@mpi.nl), and Dennis Norris (MRC Cognition and Brain Sci. Unit, Cambridge CB2 2EF, UK)

Dutch listeners categorized a /ta/–/da/ continuum before adaptation, and again during an adaptation phase in which they heard multiple repetitions of five different words. Dutch words, like English words [A. G. Samuel, *Cog. Psych.* **32**, 97–127 (1997)], produced reliable adaptation: there were more /t/ decisions when each adapting word contained a /d/ (and no /t/'s) than when each word contained a /t/ (and no /d/'s). But when these phonemes were replaced with signal-correlated noise, such that the phonemic restoration illusion was induced, there was no adaptation, contrary to the English results. Thus, although listeners heard the replaced phonemes in the noise, these phonemic percepts did not cause adaptation. These results are consistent with the Merge model [Norris *et al.*, *Behav. Brain Sci.* (in press)], in which the lexicon can influence phonemic decisions (and hence the restoration illusion), but, because there is no feedback from lexical to prelexical levels, cannot induce acoustic-phonetic adaptation. Samuel (1997) claimed on the basis of his results that the lexicon can produce potent phonemic percepts, that is, percepts which can modulate prelexical processing; the present results challenge this claim.

12:10

**5aSC8. Probabilistic phonotactics and levels of processing and representation in spoken word recognition: When is a word not a word?** Michael Vitevitch (Speech Res. Lab., Dept. of Psych., Indiana Univ., Bloomington, IN 47405)

Vitevitch and Luce (1998, 1999) hypothesize that two levels of processing—lexical and sub-lexical—are involved in spoken word recognition. Competition is observed at the lexical level: Words and nonwords with low phonotactic probability/sparse neighborhoods are responded to more quickly than stimuli with high-phonotactic probability/dense neighborhoods. Facilitation is observed at the sub-lexical level: Nonwords with high-phonotactic probability/dense neighborhoods are responded to more quickly than nonwords with low-phonotactic probability/sparse neighborhoods. Nonwords can be processed at either level, whereas words predominantly compete with each other at the lexical level. However, the present experiment demonstrates that words can be processed at a sub-lexical level. Target words varying in phonotactic probability/neighborhood density were presented in a same–different task with varying proportions of word and nonword pairs. When the stimulus set consisted mostly of word pairs, competition was observed: Words with low-phonotactic probability/sparse neighborhoods were responded to more quickly than words with high-probability/dense neighborhoods. When the stimulus set consisted mostly of nonword pairs, facilitation was observed: Words with high-phonotactic probability/dense neighborhoods were responded to more quickly than words with low-probability/sparse neighborhoods. The implications of these results for models of spoken word recognition are discussed. [Work supported by NIH-NIDCD Training Grant DC-00012.]

12:25–12:35

**Discussion**

FRIDAY MORNING, 5 NOVEMBER 1999

FAIRFIELD ROOM, 8:10 TO 11:55 A.M.

**Session 5aUW****Underwater Acoustics: Processing, Transduction and Noise**

William M. Carey, Chair

*Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02115**Contributed Papers*

8:10

**5aUW1. Sparse array technology for 3-D sonar imaging systems.**

John M. Impagliazzo (Naval Undersea Warfare Ctr., Code 8212, Bldg. 679, Newport, RI 02841, impagliazzojm@npt.nuwc.navy.mil), Steven M. Kay (Univ. of Rhode Island, Kingston, RI 02881), Alice M. Chiang, and Steven R. Broadstone (Teratech Corp., Burlington, MA 01803-3308)

A three-dimensional sonar imaging system is under development for use by Navy divers for mine-field reconnaissance and mine-hunting systems. Divers require a small, low-power, lightweight acoustic imaging

system with high resolution for examining and disposing of underwater ordnance. Sparse array technology is investigated to develop a sonar requiring the minimal number of signal processing channels while maintaining the beamwidth and sidelobe structure required for high-resolution imaging. Beamwidth and sidelobe structure are examined as a function of transducer element count, channel bandwidth, and element location when optimized using various cost functions. Locally optimum solutions are computed for uniformly weighted sparse arrays. Cost functions evaluated for the optimization process compute average sidelobe energy, peak sidelobe energy and a weighted average sidelobe energy. Images will be



formed using multiple CCD/CMOS time-delay beamforming circuits to compute the large number of beams required to provide a real-time high-resolution image.

8:25

**5aUW2. Scaling of time-reversing array performance in the presence of ambient noise.** Sunny R. Khosla and David R. Dowling (Dept. of Mech. Eng. and Appl. Mech., Univ. of Michigan, Ann Arbor, MI 48109, drd@engin.umich.edu)

Time-reversing arrays (TRA) have the capability to exploit unknown multipath environments to retrofocus acoustic waves, in both time and space, back to their original source location. Although TRA retrofocusing may be robust, it is degraded by ambient noise. The noise received by the array and the noise existing locally in the focal region cause a reduction in the focal-point signal-to-noise ratio and simultaneous waste of the useful power output of the array. A vertical TRA operating in simple environments containing an isotropic noise field is considered. A focal probability criteria is developed to quantify the retrofocusing performance of the TRA. This approach leads to a simple scaling law that incorporates the ambient noise level, the original source level, the array element level, the array element number, and the acoustic environment parameters. Results for free space and a shallow-water sound channel will be presented. Operating regimes strongly influenced by original source level and TRA element power level are clearly defined by this effort. [Work sponsored by the Office of Naval Research, Ocean Acoustics.]

8:40

**5aUW3. Wind-generated, low-frequency, ambient-noise source levels revisited.** William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215) and David G. Browning (Univ. of Rhode Island, Kingston, RI 02881)

A central issue concerning high-resolution array processing is the limit due to natural sources of ambient noise and how can it be calculated. This question has been addressed by several investigators [W. Carey *et al.*, *J. Acoust. Soc. Am.* **84**, 201 (1988); D. Kewley *et al.*, *J. Acoust. Soc. Am.* **88**, 1894–1902 (1990); F. Jensen *et al.*, *Computational Ocean Acoustics* (AIP, Woodbury, NY, 1994), pp. 517–547] and is still important to the calculation of absolute ambient noise levels. This paper revisits our previous discussion [W. Carey *et al.*, *J. Acoust. Soc. Am.* **84**, 201 (1988)] and reviews recent developments in the determination of wind-driven source levels. Source-level conventions used by several investigators are reviewed and source-level measurements for shallow- and deep-water are summarized with respect to frequency, wind speed, and measurement technique. The presentation of wind-driven noise sources in normal mode, fast field, and parabolic equation computer codes are considered with respect to the frequency-dependent noise source. The order (monopole, dipole, quadrupole) of the source of sound and the use of spatial correlation functions shall be discussed.

8:55

**5aUW4. Broadband super-resolution in a reverberant environment using “pairwise” correlations.** Marc P. Olivieri (GORCA Technologies, Inc., 300 W. Rte. 38, Moorestown, NJ 08057, olivieri@gorca.com)

In recent years, the underwater acoustic community has been focusing its attention on the problem of shallow water as a harsh acoustic environment, since traditional narrow-band systems have failed to achieve their expected deep-water performance. At present, new broadband techniques are sought to solve this problem. In previous work, the Ambient Noise Sonar [M. P. Olivieri and S. A. L. Glegg, *J. Acoust. Soc. Am.* **99**, 2454 (1996)] was developed to locate broadband sources in shallow water. In the case of multipaths or correlated sources, this broadband “pairwise” cross-correlation technique introduces ambiguities. To resolve them, a set of adaptive and eigenvector beamformers is proposed for broadband coherent processing. Because the method is based on pairwise correlations, it trades off degrees of freedom in space (case of an equi-space line array)

for degrees of freedom in frequency (bandwidth). Adaptive beamformers for the cross-frequency covariance matrix are derived using “frequency smoothing” analogous to “spatial smoothing,” and super-resolution is achieved in the case of multiple correlated sources and low signal-to-noise ratios. Results from simulations and measured data will be presented, and the sensitivity of the adaptive beamformers with respect to the bandwidth will be discussed.

9:10

**5aUW5. Pressure-particle acceleration underwater acoustic intensity sensor.** Kang Kim, Gerald C. Lauchle, and Thomas B. Gabrielson (Grad. Prog. in Acoust. and Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804)

An underwater acoustic intensity sensor designed to measure one component of the acoustic intensity vector is discussed and evaluated experimentally. This sensor consists of a pressure transducer in the form of a piezoceramic hollow cylinder, and an accelerometer mounted inside of the cylinder. This is a pressure-acoustic particle acceleration type of intensity probe, and is henceforth denoted as a “*p-u-dot* intensity probe.” The probe body has syntactic foam endcaps that are adjusted in thickness to allow the entire assembly to be neutrally buoyant. It is coated with polyurethane for waterproofing while maintaining acoustical transparency in water. Each of the two transducers making up the intensity sensor is calibrated individually in both air and water. The integrated intensity probe is calibrated in a water-filled plane-wave tube. It is shown that this neutrally buoyant underwater acoustic *p-u-dot* probe measures one component of sound intensity that is in very close agreement to the predicted value using slow-waveguide acoustic theory. Some limitations in the experiments and future works are discussed. [Work supported by ONR, Code 321 MURI in Acoustic Transduction: Materials and Devices.]

9:25

**5aUW6. An acoustic interferometer for measuring sound-speed variations in a turbulent fluid.** Michael J. Daley (Appl. Res. Lab. and the Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804), Kenneth E. Gilbert (Univ. of Mississippi, University, MS 38677), and Ralph R. Goodman (Penn State Univ., State College, PA 16804)

This paper reports laboratory measurements of the effect of thermal fluctuations on sonar ping returns from two discrete scatterers. It is the first part of an ongoing investigation of acoustic backscatter from targets and surfaces immersed in a turbulent fluid containing thermal inhomogeneities. A simple theory was developed to relate sound-speed variations and the echo level from a pair of discrete targets. The theory treats the return level from the targets as an interferometer, sensitive to the spatially averaged sound speed between the targets. Measurements were made of 200-kHz echoes from a pair of targets for various temperature conditions. As an initial calibration test, the water column was heated as a unit and allowed to cool over time. In a second set of measurements, Kolmogorov thermal turbulence was created in the water column, and the fluctuating echo level was measured. Using the amplitude of the echo, the average sound speed between the two targets was calculated for each ping. The acoustically determined sound-speed variations were compared to those inferred from thermal fluctuations and found to be in approximate agreement. Sources of error and approaches for improving the measurements will be discussed briefly. [Work supported by ONR and the Applied Research Laboratory, Pennsylvania State University.]

9:40

**5aUW7. Modeling the unsteady forces on a finite-length circular cylinder in cross-flow.** Dean E. Capone (Penn State Univ., Appl. Res. Lab., P.O. Box 30, State College, PA 16804) and Gerald C. Lauchle (Penn State Univ., State College, PA 16804)

Semi-empirical models, for unsteady lift, drag, and axial forces, are developed to predict the spectral features of the unsteady forces on a finite-length, right circular cylinder in cross-flow. In general, the models

consist of two parts; the spatial variation of rms wall pressure on the cylinder, and the correlation lengths, or areas, which describe the spatial extent of the correlation of the unsteady pressures. Experiments were conducted in a low-noise wind tunnel to measure the statistics of the unsteady wall pressures on a model cylinder. The results from the measurements are incorporated into the theoretical models, and predictions are made for the spectral characteristics of the theoretical lift, drag, and axial forces. The unsteady lift and drag predictions using the models developed in this work agree well with previously measured unsteady force data measured on gradient hydrophones exposed to flow. Below  $St=0.8$  the unsteady lift and drag are found to be the dominant force on the cylinder, while above  $St=0.8$  the unsteady axial force dominates.

9:55

**5aUW8. Laser interrogation of the air–water interface for in-water sound detection: Initial feasibility tests.** Lynn T. Antonelli, Kenneth M. Walsh, and Andrew Alberg (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841, antoneli@lego.npt.nuwc.navy.mil)

Proof of concept experimentation recently demonstrated a new laser-based acoustic sonar technique to measure the velocity of the air–water interface using laser Doppler velocimeter technology. The laser acoustic sonar concept is unique since it directs laser light from the air onto the water surface. The light scattered from the interface contains Doppler information from which the boundary velocity is obtained. Data storage and signal processing can then be performed on the detected signal and the acoustic pressure in the water calculated from the measured velocity. The laser-based system presents an alternative means for broadband sonar reception that does not interfere with the water environment. Acoustic pressure signals as low as 119 dB relative to  $1 \mu\text{Pa}$ , between 2 and 50 kHz have been detected in the laboratory using the laser velocity sensor on a static water surface. The methods of implementing the laser acoustic velocity sensor to measure pressure fluctuations on a hydrodynamic interface were also explored in a high-velocity water tunnel [sponsored by the Office of Naval Research]. The results from the initial proof of concept experimentation from both the hydrostatic and hydrodynamic air–water interfaces is presented, demonstrating the feasibility of this sensor concept.

10:10–10:25 Break

10:25

**5aUW9. Updated technology baseline in underwater acoustic modeling, simulation, and analysis.** Paul C. Etter (Northrop Grumman Corp., Oceanic and Naval Systems, P.O. Box 1488, M.S. 9115, Annapolis, MD 21404, paul\_c\_etter@md.northgrum.com)

A comprehensive review of international developments in underwater acoustic modeling is used to construct an updated technology baseline containing 106 propagation models, 16 noise models, 17 reverberation models, and 25 sonar performance models. This updated technology baseline represents a 30% increase over the previous baseline published in 1996 [P. C. Etter, *Underwater Acoustic Modeling: Principles, Techniques and Applications*, 2nd ed. (E & FN Spon, London, 1996)]. When executed in higher-level simulations, these models can generate predictive and diagnostic outputs that are useful to acoustical oceanographers or sonar technologists in the analysis of complex systems operating in the undersea environment. Recent modeling developments described in the technical literature suggest two major areas of application: low-frequency, inverse acoustics in deep water; and high-frequency, bottom-interacting acoustics in coastal regions. Rapid changes in global geopolitics have opened new avenues for collaboration, thereby facilitating the transfer of modeling and simulation technologies among members of the international community. This accelerated technology transfer has created new imperatives for international standards in modeling and simulation architectures. National and international activities to promote interoperability among modeling and simulation efforts in government, industry, and academia will be reviewed. The implications of this updated baseline for technology planners will also be discussed.

10:40

**5aUW10. Transducer-positioning system scanning around an aquatic for measurements of target strength.** Ken Ishii, Kouichi Sawada, and Yoichi Miyanoohana (Natl. Res. Inst. of Fisheries Eng., Ebidai Hasaki, Kashima, Ibaraki, 314-0421 Japan)

The transducer-positioning system has been designed in order to measure acoustic target strength (TS) of fragile and unstable in shape aquatic targets such as a squid. It is characteristic that directivity of TS is measured while making a transducer scan a fixed target suspended in the water. This system is loaded in the electric flatcar on the top of a tank. The distance between a transducer and a target is 3.65 m. A transducer moved from  $-45$  to  $+45$  deg in rotation around a target. The minimum unit of set value for a destination is 1 mm. Furthermore, a Laser pointer has been installed so that it becomes easy to observe a point where an axis of ultrasonic beam passes through a target when a target is prepared for suspension. The Laser pointer is set up in exchange for a transducer and is adjustable three-dimensionally. The situation between a transducer and a target, precision of positioning, and specification required for acoustic measurement system are described in this paper.

10:55

**5aUW11. The transient conical wave field emitted by a supersonic source in a two media configuration.** Bert Jan Kooij (Lab. of Electromagnetic Res., Delft Univ. of Technol., Delft, The Netherlands, b.j.kooij@its.tudelft.nl) and Cornelis Kooij (formerly Eindhoven Univ. of Technol., Eindhoven, The Netherlands)

The acoustic conical wave field due to a supersonic motion of a rigid object over a half-space is investigated. The research presented here is of great importance for the modeling of the transmitted and reflected conical waves that are excited by supersonic airplanes that fly over sea or land. The media involved are acoustically modeled as homogeneous, isotropic fluids and the moving object is represented as an acoustic source of the dipole type. The linearized acoustic field equations are applied to obtain representations for the fields in a combined Laplace–Fourier transform domain. To these representations, which are mapped into the proper form, we apply the Cagniard–de Hoop technique in order to find closed-form time-domain solutions for the reflected and transmitted acoustic fields. In the Cagniard–de Hoop technique we use a new transformation in order to get separate expressions for the reflected and transmitted conical waves as well as the conical headwave that can occur under certain circumstances. These special wave phenomena come forward from the procedure with precursive fronts and clearly defined trailing edges. Numerical results for the reflected as well as the transmitted conical wave are presented in a configuration where conical headwaves can occur.

11:10

**5aUW12. Underwater acoustic vision.** K. Keller, S. Mack, J. Lucas, and J. Smith (Robotics and Computer Electron. Group, Dept. of Elec. Eng., Univ. of Liverpool, Brownlow Hill, Liverpool L69 3GJ, UK)

An underwater acoustic camera has been developed to enable imaging of objects and scenes in environments, where water clarity makes optical methods redundant. The camera consisted of multiple transmitters, acoustic lens, and a 256-element receiver array. Continuous waves techniques were used, to provide range and intensity information which was then processed to produce three-dimensional images. The ranging technique uses a 1-MHz base frequency transmitted signal which was swept linearly according to a sawtooth pattern. The frequency difference between the transmitted and received signal frequencies, caused by the finite time of flight of the wave, determined the range of the target. Broadband transducers using PZT piezoelectric material were designed to provide acoustic illumination of the scene. The acoustic lens (either liquid or solid) focused the reflected rays onto the PZT/polymer receiver array. Images have been produced consisting of both single- and photomosaic frames. These showed the ability of the camera to produce images of objects of varying shapes and orientations, in environments where conventional optical sys-

tems were redundant (due to low visibility). A number of applications have been suggested for development based on the results produced, including subseabed mine detection and the docking of an underwater vehicle.

11:25

**5aUW13. Investigation of generation of low-frequency fields produced by parametric acoustic array, located in the waveguides with rough surfaces.** Irina Yu. Gryaznova, Sergey N. Gurbatov, Sergey A. Egorychev, Vasily V. Kurin (Nizhny Novgorod Univ., 23 Gagarin Ave., Nizhny Novgorod, 603600 Russia), Leonid M. Kustov, and Nikolay V. Pronchatov-Rubtsov (Nizhny Novgorod Univ., Nizhny Novgorod 603600)

Under natural conditions the parametric acoustical arrays (PAA) work as a rule near the bottom or near the sea surface. If there are free surfaces in the area of interacting of the high-frequency waves (HFW) then the multi-lobe directivity pattern of the PAA is formed in the horizontal plane. However, these effects appear only in waveguides which are bordered smooth surfaces, because these boundaries do not violate conditions of coherence for interacting waves. If there is a rough surface in the area of interacting of HFW, it leads to limit the length of PAA. In this case a single-lobe directivity pattern on the low-frequency (LF) is formed [Gurbatov *et al.*, *Radiophys. Quantum Electron.* **38**, 82–86 (1995)]. In this paper the results of laboratory investigations of the LF fields angular distribution in waveguides with the rough surfaces are presented. It is shown that the rough boundaries leads to increasing of angular distributions width in composing with similar characteristics in boundless space, and the quality kind of field distribution in the horizontal plane depends not only on its

mode structure, but on characteristics of rough waveguide boundaries. [Work supported by RFBR Grants No. 99-02-18354, No. 98-02-16402, INTAS Grant No. 11134.]

11:40

**5aUW14. Field of acoustic parametric array generated by a point source in a cylindrical waveguide with rough boundaries.** Michael Georgi (Taganrog State Univ. of Radio Eng., Dept. of Hydroacoustic and Medical Technique, Taganrog, 347909 Russia, 2 dacha, d.50, kv.4)

Consider the possibility of using obtained theoretical solving for the diagnostics and checking a structure of the nonlinear acoustic parametric array (APA) field in inwardly closed or half-closed volumes of diagnosing structures via the determination of field structure over distinctive parameters: pressures, displacing, oscillatory velocities, etc., with allowance for a physical nonlinearity of the media, hydrodynamic, boundary and other conditions of ultrasonic field forming. A theoretically designed method of solving an interior problem on the finding of ultrasonic field in a cylindrical waveguide with rough surface at irregularity description functions  $h(a,z)$ , where the method of consequent approximations is used for obtaining and representing boundary conditions under small magnitudes of irregularities, commensurable with magnitudes of acoustic field wavelengths. Herewith roughness (irregularities) is considered to be, for example, a border-reflecting surface, beside which is a geometric with a small change in the scale of working wavelength, in our case, field of primary pumping waves (PPW). The physical problem model represents itself as consideration of an acoustic field of the monochromatic point source of an inwardly circular cylinder with radius  $a$ , the physical characteristics of which are assumed as ideal reflecting, and the irregularity (roughness) of its surface is given in general type by the function of roughness' description  $h(a,z)$ .

FRIDAY AFTERNOON, 5 NOVEMBER 1999

HAYES ROOM, 1:00 TO 2:05 P.M.

### Session 5pAAa

## Architectural Acoustics: Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

Leo L. Beranek, Chair

975 Memorial Drive, Suite 804, Cambridge, Massachusetts 02138-5755

Chair's Introduction—1:00

### Invited Paper

1:05

**5pAAa1. The Vern Knudsen Lecture: The compleat concert hall: A century of acoustics development and its lessons for the design of concert halls.** R. Lawrence Kirkegaard (Kirkegaard & Assoc., 4910 Main St., Downers Grove, IL 60515)

From the opening of Carnegie Hall just before the turn of the (20th) century, the art and science of concert hall design has advanced significantly . . . but that advance has been over many bumpy roads and with a number of detours along the way. As a profession we need to do all we can to assure that the halls we build in the future will serve as well in a hundred years as those that were built a century ago. In some instances that is a very tall challenge indeed. This provocative paper seeks to be visionary as well as retrospective, laudatory as well as critical. It is illustrated by experience and observation and delivered with respectfulness, but with enough challenge and irreverence to stimulate thoughtful and challenging discussion.

**Session 5pAAb****Architectural Acoustics: Architectural Renovation and Restoration: Acoustic Issues**

Scott D. Pfeiffer, Chair

*Kirkegaard and Associates, 4910 Main Street, Downers Grove, Illinois 60515***Chair's Introduction—2:15*****Invited Papers*****2:20****5pAAb1. Auditorium and shell renovation at DeVos Hall, Grand Rapids, MI.** Joseph W. A. Myers (Kirkegaard & Assoc., 4910 Main St., Downers Grove, IL 60515, jmyers@kirkegaard.com)

DeVos Hall (Grand Rapids, MI) is a good multi-purpose hall that had some acoustic shortcomings, particularly in its function as home to the Grand Rapids Symphony, and was renovated to improve acoustic performance for orchestral use. The renovation combined modest modifications to room surfaces with a complete rethinking and reworking of the orchestra shell. Significant improvement was achieved at modest expense and without compromising the hall's use for nonorchestral functions.

**2:40****5pAAb2. Renovating a landmark: London's Royal Festival Hall.** Carl P. Giegold and Paul T. Calamia (Kirkegaard & Assoc., 4910 Main St., Downers Grove, IL 60515)

Since its opening in 1951, London's Royal Festival Hall has been plagued with a number of acoustic shortcomings. Specifically, the hall suffers from a short reverberation time despite its size, a relatively weak bass response, a sense of acoustic "detachment" for listeners in the under-balcony, and difficult playing conditions on stage. Plans for a major renovation of the hall are currently being prepared for submission to the National Lottery. The authors will discuss the potential remedies for the issues noted above, with particular attention paid to implementation and measurement in a 1:20 scale model.

**3:00****5pAAb3. Key acoustical issues associated with the renovation and restoration of the Southern Theatre, Columbus, Ohio.** David W. Kahn and Ronald T. Eligator (Acoustic Dimensions, 2 East Ave., Larchmont, NY 10538)

For an acoustics designer of performing arts facilities, the renovation and restoration of a landmarked facility can be frustrating, as there are rarely opportunities to alter the original architecture, and therefore alter the original acoustics. The Southern Theatre renovation and restoration was somewhat unique with respect to the extent to which the theater interior was altered to address code issues, and to incorporate HVAC systems. The Southern Theatre originally opened in 1896. The stagehouse, main floor, and second balcony were rebuilt as part of the renovation. The second balcony, which originally had a separate entrance from the street, was redeveloped to incorporate contemporary theater seating, as well as some additional technical support spaces including control rooms and new lighting positions. With the wide scope of the renovation came the opportunity to introduce some important acoustic improvements. Among the key issues discussed are the acoustical design of a unique silent air supply system, and design of an unusual enclosure for music ensembles that develops the lower volume of the stagehouse as a partially coupled reverberation chamber.

**3:20****5pAAb4. Great halls: Great renovations. Renovation of Europe's two most famous concert halls—the Grosser Musikvereinssaal in Vienna and the Concertgebouw in Amsterdam.** Pamela Clements (Jaffe Holden Scarbrough Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, pclements@jhsacoustics.com)

It is not widely known that Europe's two leading concert halls were renovated in the past, renovation that perhaps significantly affected the acoustical quality of the spaces. In 1911 the Grosser Musikvereinssaal underwent major renovation to improve fire safety. The hall was almost gutted, the ceiling and stage rebuilt, the light fixtures on the balcony rails were removed, and the caryatids and entablature supporting the balcony were relocated from the front of the parterre to the side walls. At the Concertgebouw the seating was changed (again for fire safety), and in 1899 the stage was rebuilt to reduce the height of the very steep orchestral risers. This was done to improve the sound quality and playing conditions for the orchestra. In the case of the Musikverein, this author has found little contemporary comment on what may have been remarkably changed acoustical conditions. In the case of the Concertgebouw, the renovation was hailed as a great improvement—yet this stage is still much steeper than would be recommended by most acousticians today. This paper is based on plans, photographs, and other historical evidence about the acoustic conditions in the two halls before and after the renovations.



3:40

**5pAAb5. Noise isolation problems in the renovation of old urban factory buildings for residential use.** Angelo Campanella (Campanella Assoc., 3201 Ridgewood Dr., Columbus, OH 43026)

Special noise isolation problems arise when renovating old urban industrial buildings into apartments, condominiums, and hotels under the historic preservation rules. The living environment created should experience — as result of outdoor noise — not more than a maximum of 55 dBA in sleeping and living rooms. In addition, common walls and floor/ceilings should have sound isolation values of STC50 (apartments) to STC55 (condominiums) if at all possible. The ceilings of many rooms and loft spaces are often the exposed slatted wood or bottom of the floor above. This must be preserved in appearance, disallowing ceiling surface noise sound isolation treatment. Floor/ceilings should have impact sound isolation values of IIC50 (apartments) to IIC55 (condominiums) if at all possible. Window sound isolation is primarily selected for measured maximum noise levels caused by local truck traffic and emergency vehicle warning sirens on the basis of octave band noise and transmission loss values. The thicker window required must be compatible with existing metal or wood window frames. Test results and design concepts for some improved and preserved existing old commercial masonry urban structures will be presented.

4:00

**Southern Theatre Walking Tour**  
See section 35 in the Meeting Information for details.

FRIDAY AFTERNOON, 5 NOVEMBER 1999

MCKINLEY ROOM, 1:45 TO 3:30 P.M.

### Session 5pMU

## Musical Acoustics: Free Reed Instruments II

James P. Cottingham, Chair  
*Department of Physics, Coe College, Cedar Rapids, Iowa 52402*

### *Invited Papers*

1:45

**5pMU1. Concertinas: nuts, bolts, and music.** Allan W. Atlas (Ph.D.—DMA Programs in Music, Grad. Ctr./City Univ. of New York, 365 Fifth Ave., New York, NY 10016-4309, aatlas@GC.cuny.edu)

Developed ca. 1830 by the physicist Sir Charles Wheatstone, the concertina has had a checkered career. This paper will discuss the various types of concertinas, outline the history of the instrument, discuss its construction, provide information on “how it works,” and intersperse live examples. Briefly, there are two national traditions: British and German, with various types of concertinas in each. In England, on which we shall concentrate, the concertina “fell” from its mid-19th-century success in both the concert hall and the drawing rooms of England’s socio-economic elite to become associated with folk music, the music hall, and the Salvation Army by the end of the century. After falling into almost total oblivion between the two world wars, the instrument has enjoyed an impressive comeback in the realm of folk music and has even aroused some interest (small, to be sure) in art–music circles. The instrument’s construction will be considered, as will the manner in which one concertina in particular works: the so-called “English,” on which live examples will be presented. The paper will conclude with a few words about the German bandoneon, its emigration to Argentina, and its evolution into the “signature” timbre of the tango orchestra.

2:15

**5pMU2. The construction and operation of the khaen.** Terry E. Miller (Ctr. for the Study of World Music, Kent State Univ., Kent, OH 44242, tmiller1@kent.edu )

The khaen is perhaps the most important musical instrument of the Lao people of northeastern Thailand and Laos. It is constructed with free reeds mounted in bamboo pipe walls inside a carved wooden wind chamber. Each reed of the khaen sounds for both directions of air flow (inhaling and exhaling). The reed vibration is strongly coupled to the pipe resonance, and the reed will sound only if a small hole near the reed is closed, causing the resonant frequency of the pipe to be near that of the reed. The construction of the khaen is discussed and illustrated with video recordings made in northeastern Thailand. Principles of operation of the instrument are discussed and demonstrated. Comparisons will be made with other Asian free reed instruments.

2:45–3:30

### Informal Demonstration/Performance

## Session 5pUW

## Underwater Acoustics: Shallow Water Acoustics

Brian J. Sperry, Chair

*Applied Ocean Physics and Engineering Department, Woods Hole Oceanographic Institution,  
Woods Hole, Massachusetts 02543*

## Contributed Papers

1:00

**5pUW1. Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth.** Gerald L. D'Spain and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

The waveguide invariant summarizes in a single scalar parameter the dispersive propagation characteristics for a given environment. An expression is derived for the invariant in shallow water environments where the bottom bathymetry varies with range and azimuth. It then is applied to the interpretation of broadband, single-element spectrograms collected during SWellEx-3, a shallow water acoustics experiment off the coast of San Diego. The unusual behavior of the interference patterns seen in the spectrograms from this experiment during the transmission of a 75–150 Hz pseudo-random noise sequence by a source towed over range-varying bottom bathymetry is well matched by the model predictions. [Work supported by ONR, Code 321US.]

1:15

**5pUW2. Measurement of shallow-water reverberation characteristics.** Peter G. Cable (BBN Technologies, Union Station, New London, CT 06320) and William M. Carey (Boston Univ., Boston, MA 02215)

The frequency dependence and level of shallow-water (SW) reverberation, key parameters in sonar performance, represent the combined effect of scattering from the bottom, volume, and sea surface, and are a function of the transmission loss. SW, however, has the attribute of a seasonal dependence to the sound-speed structure, which can emphasize one of the above. A downward refracting profile in a gently sloping environment emphasizes the bottom and when the transmission is measured provides for a means of correcting for this loss and estimating the angle of incidence. This paper describes such measurements performed in several SW–sandy-bottom experiments with horizontal arrays, vertical arrays, and specially designed and calibrated sources. The scattering results are shown to be homogeneous and isotropic and to exhibit a frequency-dependent characteristic below 1 kHz with a minimum value apparently determined by the sediment layering. Numerical modeling suggests that scattering arises from the top several meters of sediment. The values determined here are consistent with previous measurements and the angular dependence determined by the use of vertical arrays and analytical calculations are consistent with Lambert's Law. Limits on the values of this type of reverberation and performance differences with sonar operating frequency are discussed.

1:30

**5pUW3. Experimental evaluation of the performance of coherent digital communications algorithm in littoral ocean.** Azmi A. Al-Kurd and T. C. Yang (Naval Res. Lab., Washington, DC 20375)

Underwater acoustic communications data using phase modulated signals were collected during the Littoral Warfare Advanced Development 98-1 experiment, conducted in the Gulf of Mexico in Nov. 1998. CW pulse and quadrature phase shifted keying (QPSK) signals were projected from a towed source. Data of three baud rates were collected to study the temporal and spatial variation of the acoustic impulse response of the ocean and to evaluate the performance of a phase coherent digital com-

munication algorithm. Post-experiment analysis examined the Doppler shift, signal amplitude fluctuation, signal-to-noise ratio of the received signal, channel impulse response, temporal correlation function of the received signals, and signal fluctuation statistics during several segments of the experiment. The objective is to determine the characteristics of signal propagation in littoral environment, and to determine whether and how these characteristics affect the bit-error-rate of a coherent acoustic communication algorithm. The analysis results serve as guidance in the design of a reliable coherent digital communication system. [This work is supported by the Office of Naval Research.]

1:45

**5pUW4. Flow-controlled acoustic propagation channels on continental shelves.** Marshall H. Orr and Peter C. Mignerey (Code 7120, Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Continental shelves have steep changes in bathymetric gradients. These gradients are caused by a wide variety of geological processes and are often associated with the shelf/slope break, canyons, and banks created during pauses in transgressive/regressive sea-level cycles. Tidal flow over these bathymetric features causes temporal variability in the local sound-speed field. The temporal variability is repeatable over successive semidiurnal tidal cycles. The flow ranges from subcritical to supercritical. The variability in the flow over gradients in the bathymetry can result in the formation of a stratified sound channel, perturbation of the channel by linear and noninternal waves, flow stagnation with separation, and hydraulic jumps that can have associated mixing events. Each of these flow types will cause unique sound-speed fields that can impact the properties of an acoustic signal propagating through the region. Calculations that show the relative importance of some flow types on the temporal variability of the acoustic signals that have propagated through a flow-controlled channel are presented. [Work supported by the Office of Naval Research.]

2:00

**5pUW5. Stochastic analysis of the internal wave effect on the SWARM observations.** Yongke Mu and Mohsen Badiey (College of Marine Studies, Univ. of Delaware, Newark, DE 19716)

Wave propagation through random fluctuating media in time and space is a common theme in different fields. One such area, which provides a challenge, is shallow water acoustic wave propagation through internal waves. The effect of such propagation can range from an almost deterministic wave field to a completely stochastic one. The Shallow Water Acoustics in a Random Medium (SWARM) experiment has provided a comprehensive data set for studying the solitary internal wave effects on the broadband acoustic propagation. In order to understand the acoustic scattering observed in the SWARM data, we combined a nonlinear internal wave together with random variations of the environment, the correlation of which maps the internal wave regimes observed in the SWARM data, to the numerical simulations. The full wave parabolic equation (PE) model is adopted for simulating the broadband acoustic propagation. By varying the spatial and temporal statistics of the environment, the acoustic arrival signatures can be generated randomly with a given mean and standard

deviation. Through this exercise we quantify the environmental statistics and its relationship to the acoustic waves by comparing the numerical and experimental data.

2:15

**5pUW6. Studies of shallow-water acoustic coherence using the modal phase structure function.** Brian Sperry and James Lynch (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The modal phase structure function, defined by the cross correlation of the phase for any given mode at two different points in space, time, or frequency, can provide some useful insights into the physics of array coherence in shallow water. Using this indicator of the acoustic field coherence, as well as the usual  $\langle pp^* \rangle$  definition, we examine the coherence of acoustic signals in the 10-to-1000-Hz range scattered by both bottom and water column inhomogeneities. In previous work (Sperry, Ph.D. thesis, 1999), we emphasized large-scale water column inhomogeneities, and ignored some of the finer scale ocean phenomena that could cause significant mode coupling. In this work, we address both phenomena and the general issue of mode coupling. In particular, we address the soliton field in the water column and small-scale inhomogeneities in the bottom sediments. Water column environmental data from the PRIMER experiment and bottom environmental data from the STRATAFORM experiment are used to guide our analyses. Results are compared to acoustic observations for the PRIMER experiment, where both acoustic and environmental data were gathered simultaneously. [Work supported by ONR.]

2:30–2:45 Break

2:45

**5pUW7. Azimuthally dependent transmission loss and horizontal coherence in the presence of shelf internal waves.** Roger M. Oba and Steven Finette (Naval Res. Lab., Washington, DC 20375-5350)

A three-dimensional acoustic model, FOR3D, is used to propagate an acoustic field from a continuous wave point source to a horizontal array of receivers. The space-time variability of the continental shelf environment is modeled by shallow water of constant depth containing two types of internal waves: a spatially diffuse, randomly isotropic field, and a deterministically propagated, transversely invariant solitary wave packet. The internal wave packets cause azimuthal and range dependence in the properties of the acoustic signal as a function of (a) the bearing from the source, as referenced to the wave packet's propagation vector, and (b) the location of the packet in the propagation path. The simulations also demonstrate a degree of out-of-plane scattering. Ensemble averaging is used to simulate horizontal coherence in environments containing combinations of the internal waves, and to illustrate the comparative importance of the diffuse and wave packet contributions. The cases studied here at selected frequencies between 100 and 400 Hz suggest that the performance of horizontal array receivers may require full three-dimensional modeling of propagation and media. [Work supported by the Office of Naval Research.]

3:00

**5pUW8. Computed time-reversing array retrofocus azimuth size in a dynamic bottom-varying shallow-water sound channel.** Michael R. Dungan and David R. Dowling (Dept. of Mech. Eng. and Appl. Mech., Univ. of Michigan, Ann Arbor, MI 48109, drd@engin.umich.edu)

A time-reversing array (TRA) can retrofocus acoustic energy to the location of its original source even when the acoustic environment is complex and unknown. However, vertical TRA performance in the azimuth direction is entirely determined by variations in the water column and

bathymetry. Predictions of TRA retrofocusing performance in this study are made with a parabolic equation code (RAM by M. D. Collins). In addition to the sensitivity of TRA retrofocusing to time variation in the water column of a shallow-water sound channel, this study addresses the azimuth size of the retrofocus. Interestingly, the azimuthal retrofocus size decreases with increasing range when internal waves are present. However, azimuthal retrofocusing is degraded as the acoustic propagation from the TRA is influenced by internal-wave dynamics within the water column. The range at which the effect of bottom variability on the retrofocus azimuth size matches that of internal waves is characterized. The internal waves are based on measured oceanic statistics drawn from the SWARM '95 experiment. Retrofocusing results are presented for a variety of ranges and acoustic frequencies. [Work supported by the Office of Naval Research, Ocean Acoustics.]

3:15

**5pUW9. Results from a simulation study of mode coupling due to acoustic wave-soliton interactions.** Stanley A. Chin-Bing, David B. King, Alex Warn-Varnas, Robert A. Zingarelli (Naval Res. Lab., Stennis Space Center, MS 39529), and Joseph E. Murphy (Univ. of New Orleans, New Orleans, LA 70148)

There is a strong interest in understanding the interactions between ocean acoustic waves and shallow water solitons. Central to these interactions are the acoustic mode conversions caused by the solitons. Such mode conversions potentially occur each time the acoustic wave and soliton interact; however, at certain combinations of acoustic-oceanographic wavelengths, the interactions can produce large losses in acoustic signal. Mode coupling results will be discussed from simulation studies that examined the signal loss due to interactions with solitons in the Strait of Messina. Coupled mode formulations were avoided so that interpretations of results were not dependent upon any particular mode coupling scheme. Rather, the full-wave, two-way, finite-element model, FFRAME, was used to produce numerically accurate predictions of the complex pressure field as it passed through the soliton. This complex field was then decomposed into its component wave numbers via an FFT to determine the mode structure. This simulation included the entire acoustic field (forward propagation, backscatter, multiple scatter, etc.). The results were realistic and difficult to interpret. Results will be shown at various points along the way as the acoustic wave passed through the soliton. [Work supported by ONR/NRL and by a High Performance Computing DoD grant.]

3:30

**5pUW10. Studies of internal wave and bottom effects on broadband shallow-water propagation.** Timothy F. Duda, James C. Preisig, and George V. Frisk (Appl. Ocean Phys. and Eng. Dept., MS 11, Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Broadband numerical modeling of acoustic propagation through the continental-shelf environment is used to evaluate the impacts of various natural features on 50–1000 Hz signals. Previous studies have shown that thermocline depth variations due to internal waves, bores, or fronts cause mode coupling. Furthermore, if the coupling causes net flux of acoustic energy into efficiently propagating low-order modes, then more energy propagates to distant ranges, and vice versa. In addition, for moving internal waves the signal changes over time, with time scales of minutes. These effects are dependent on geoacoustic properties of the seafloor and on frequency. Details of the effects are quantified for various depths, bottom types, and frequencies. [Work supported by ONR Ocean Acoustics.]